

RESEARCH

Open Access



# PE/PPE mutations in the transmission of *Mycobacterium tuberculosis* in China revealed by whole genome sequencing

Wei-wei Fang<sup>1,2</sup>, Xiang-long Kong<sup>3</sup>, Jie-yu Yang<sup>1,2</sup>, Ning-ning Tao<sup>1</sup>, Ya-meng Li<sup>1,4</sup>, Ting-ting Wang<sup>1</sup>, Ying-Ying Li<sup>1,4</sup>, Qi-lin Han<sup>1,2</sup>, Yu-zhen Zhang<sup>1,2</sup>, Jin-jiang Hu<sup>1,2</sup>, Huai-chen Li<sup>1</sup> and Yao Liu<sup>1\*</sup>

## Abstract

**Objective** This study aims to examine the impact of PE/PPE gene mutations on the transmission of *Mycobacterium tuberculosis* (*M. tuberculosis*) in China.

**Methods** We collected the whole genome sequencing (WGS) data of 3202 *M. tuberculosis* isolates in China from 2007 to 2018 and investigated the clustering of strains from different lineages. To evaluate the potential role of PE/PPE gene mutations in the dissemination of the pathogen, we employed homoplastic analysis to detect homoplastic single nucleotide polymorphisms (SNPs) within these gene regions. Subsequently, logistic regression analysis was conducted to analyze the statistical association.

**Results** Based on nationwide *M. tuberculosis* WGS data, it has been observed that the majority of the *M. tuberculosis* burden in China is caused by lineage 2 strains, followed by lineage 4. Lineage 2 exhibited a higher number of transmission clusters, totaling 446 clusters, of which 77 were cross-regional clusters. Conversely, there were only 52 transmission clusters in lineage 4, of which 9 were cross-regional clusters. In the analysis of lineage 2 isolates, regression results showed that 4 specific gene mutations, PE4 (position 190,394; c.46G > A), PE\_PGRS10 (839,194; c.744 A > G), PE16 (1,607,005; c.620T > G) and PE\_PGRS44 (2,921,883; c.333 C > A), were significantly associated with the transmission of *M. tuberculosis*. Mutations of PE\_PGRS10 (839,334; c.884 A > G), PE\_PGRS11 (847,613; c.1455G > C), PE\_PGRS47 (3,054,724; c.811 A > G) and PPE66 (4,189,930; c.303G > C) exhibited significant associations with the cross-regional clusters. A total of 13 mutation positions showed a positive correlation with clustering size, indicating a positive association. For lineage 4 strains, no mutations were found to enhance transmission, but 2 mutation sites were identified as risk factors for cross-regional clusters. These included PE\_PGRS4 (338,100; c.974 A > G) and PPE13 (976,897; c.1307 A > C).

**Conclusion** Our results indicate that some PE/PPE gene mutations can increase the risk of *M. tuberculosis* transmission, which might provide a basis for controlling the spread of tuberculosis.

**Keywords** *Mycobacterium tuberculosis*, Whole genome sequence, PE/PPE, Transmission, China

\*Correspondence:

Yao Liu  
doctorliuyao@126.com

<sup>1</sup>Department of Respiratory and Critical Care Medicine, Shandong Provincial Hospital, Shandong First Medical University, Jinan, Shandong 250021, PR China

<sup>2</sup>Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan, Shandong, People's Republic of China

<sup>3</sup>Shandong Artificial Intelligence Institute, Qilu University of Technology & Shandong Academy of Sciences, Jinan, Shandong, PR China

<sup>4</sup>Shandong University of Traditional Chinese Medicine, Jinan, Shandong, PR China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Tuberculosis (TB) continues to be one of the most prevalent and deadly communicable disease that is a major global health challenge [1]. The Covid-19 pandemic has disrupted TB services which left many TB patients undiagnosed and untreated, leading to an increase in TB deaths and transmission [2, 3]. The evolution and spread of tuberculosis threaten to undermine the success of tuberculosis treatment and control programs [4]. In order to effectively combat TB, it is crucial to have a comprehensive understanding of TB's transmission mechanisms.

One hallmark of the *Mycobacterium tuberculosis* (*M. tuberculosis*) genome is the presence of the multigenic PE/PPE family of proteins, accounting for about 10% of the coding region of the genome [5]. The standard H37Rv has 99 PE genes and 69 PPE genes, which characterized by conserved N-terminal prolineglutamate (PE) and proline-proline-glutamate (PPE) motifs [6]. Based on the high polymorphism of C-terminal amino acid sequence, the PE family can be divided into PE\_PGRS (polymorphic GC-rich sequences) and PE (with no distinctive features) genes, and the PPE family includes PPE\_MPTR (major polymorphic tandem repeats), PPE\_SVP (with a GxxSVPxxW motif), PPE\_PPW (with a PxxPxxW motif) and PPE genes with no distinctive features [5, 7–9]. The unique sequences of PE/PPE proteins might underlie their specific physiological roles during *M. tuberculosis* infection.

Many PE/PPE proteins have been shown to play an important role in antigenicity, immune-modulation and virulence in *M. tuberculosis* [10–12]. For example, PPE68 and PE35 were identified as required for *M. tuberculosis* virulence [13, 14]. Cell necrosis is associated with the spread and virulence of *M. tuberculosis* because it leads to the release and dissemination of the tuberculin pathogen [15]. This function has been reported for PE\_PGRS33 [16] and PPE27 [17]. PPE39 [18], PE\_PGRS5 [19] and PE\_PGRS17 [20] were shown to play roles in host cell interaction and immune regulation. Alternatively, the highly cellular immune response suggests that some PE/PPE proteins may be better diagnostic and vaccine candidates [21].

Some genes were found to contribute to enhanced transmission of *M. tuberculosis*, such as mutation in the ESX-5 type VII secreted protein EsxW [22] and mutation in the lldD2 promoter [23]. The role of PE/PPE genes in transmission, albeit less studied, has shown that mutations in PPE54, for instance, contribute to the enhanced spread of the disease in Malawi [24], highlighting their potential significance in the epidemiological dynamics. Here, we have used whole genome sequence data from 3202 Chinese isolates and detected homoplastic single nucleotide polymorphisms (SNPs) in PE/PPE gene region to assess the impact of these homoplastic SNPs on the

spread of *M. tuberculosis*. The result may provide new insights into the impact of the PE/PPE gene mutations on the spread of *M. tuberculosis* in China.

## Method

### Sample collection

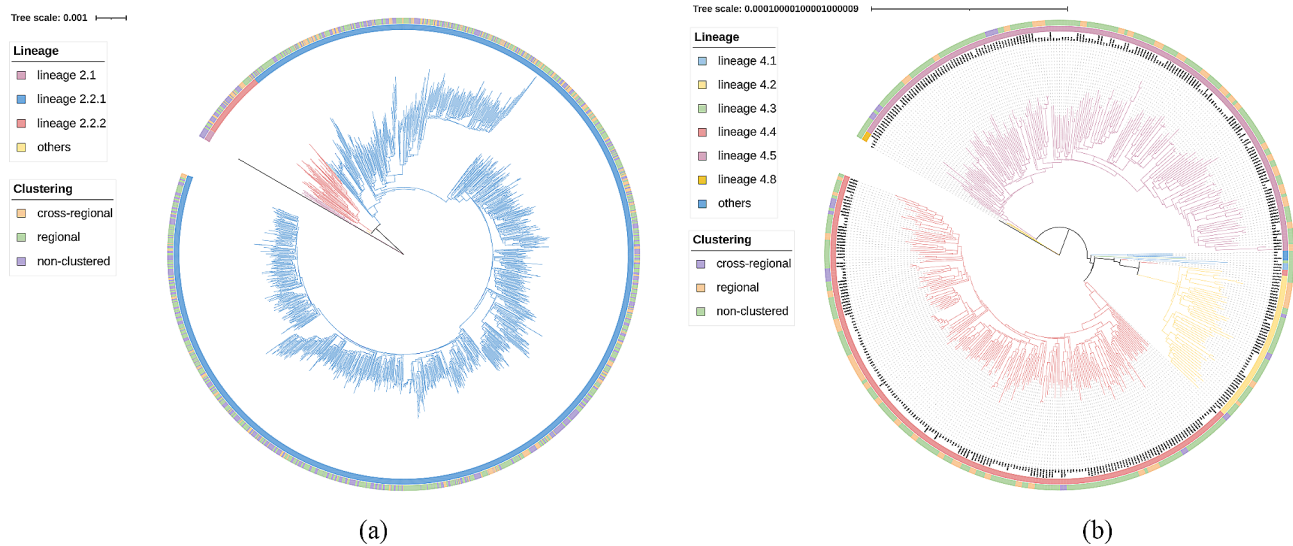
The *M. tuberculosis* strains analyzed in this study included two sets of samples (Supplementary Table 1). (1) Newly sequenced whole genome dataset. We included 1,550 culture-confirmed TB samples with drug susceptibility test (DST) results reported to Shandong Tuberculosis Surveillance System during 2013–2017, and the genome sequence data were deposited in the National Center for Biotechnology Information (NCBI) BioProject database (Accession number is PRJNA1002108). (2) Countrywide collection of publicly available clinical isolates of *M. tuberculosis*. We downloaded the WGS data of 1,755 isolates from the European Nucleotide Archive repository. The genomes were sampled from 2007 to 2018, and the geographic distribution covered 30 of the 34 provincial regions of China. Individual patient identifiers were removed before data analysis and reporting. An informed consent waiver and ethical approval were obtained from the Ethics Committee of Shandong Provincial Hospital affiliated to Shandong First Medical University.

### Whole-genome sequencing

The genome of the 1468 Shandong isolates was sequenced using Illumina HiSeq 4000. Quality assessment of all acquired reads was performed with FastQC v.0.11.9 (version 0.11.7), and 1447 samples passed quality control [25]. Low-quality raw reads from paired-end sequencing were discarded. Reads were then aligned to the H37Rv (NC\_000962) reference genome using BWA-MEM (version 0.7.17-r1188) [26]. Duplicate reads and clipped alignments were removed with Samtools markdup (version 1.15) and Samclip (version 0.4.0) [27, 28], and only samples with a coverage rate of 98% or higher and a minimum depth of at least 20 were included. The filtered vcf file was annotated with snpEFF (version 4.3t) to obtain the final sample single nucleotide polymorphisms (SNPs) [29].

### Homoplastic SNPs identification

We used snippy-core (version 4.6.0) to obtain SNPs from entire 3202 isolates. To assign lineages, we analyzed *M. tuberculosis* WGS data using TBProfiler (version 4.3.0) [30, 31]. The IQ-TREE (v1.6.12) model “JC+I+G4” used 1000 ultrafast bootstrap replicates and treeTime (v0.9.0) to construct and date maximum-likelihood (ML) phylogenetic tree [32]. The occurrence of homoplastic mutations can be attributed to convergent evolution driven by selective pressures. Homoplasmy analysis was performed



**Fig. 1** (a) Phylogenetic tree of 2745 Chinese *M. tuberculosis* strains in lineage 2. (b) Phylogenetic tree of 443 Chinese *M. tuberculosis* strains in lineage 4

employing the HomoplasyFinder software, in adherence to established protocols and methodologies. The SNPs were considered as homoplastic if they occurred independently within different transmission clades and did not form a monophyletic clade based on the provided phylogenetic tree and nucleotide alignment consistency index [33, 34]. Homoplastic SNPs located in the PE/PPE gene region, with a minor allele frequency (MAF) > 0.005 were included for further analysis.

### Transmissibility analysis

Clusters were defined as strains with a genetic distance of 10 SNPs or fewer, indicating recent transmission [22]. These clusters were categorized as cross-regional or regional, depending on whether they included strains from two or more of China's seven geographic regions (Northwest, Northern, Northeast, Central, Eastern, Southern, and Southwest China) [35]. Furthermore, transmission clusters were subdivided into small clusters (2 isolates), medium-sized clusters (3–6 isolates), and large clusters (more than 6 isolates) based on the number of isolates within each cluster [36, 37].

### Statistical analysis

To streamline the analysis, homoplastic SNPs with a MAF < 0.005 (approximately 15

isolates) in the PE/PPE gene region were excluded from the analysis. Between clustered and non-clustered, as well as cross-regional and regional clusters, we employed univariate regression analysis and included sites with  $P$  values < 0.2 in the subsequent multivariate regression analysis [38]. To analyze the effect of PE/PPE gene mutations on cluster size, we conducted a Spearman's rank correlation analysis. The statistical analysis was performed using IBM SPSS Statistics (version 26.0).

**Table 1** Lineage and demographic factors associated with transmission clusters ( $\leq 10$  SNP) of *M. tuberculosis* strains in China

Lineage	Clustered No (%)		Non-clustered No (%)	Total (n = 3202) No (%)
	Regional	Cross-regional		
1	0 (0.00)	0 (0.00)	4 (100.00)	4 (100.00)
2	1080 (39.34)	382 (13.92)	1283 (46.74)	2745 (100.00)
3	2 (20.00)	0 (0.00)	8 (80.00)	10 (100.00)
4	107 (24.16)	25 (5.64)	311 (70.20)	443 (100.00)

All reported statistical tests were two-sided, and  $P$  values < 0.05 were considered statistically significant.

## Results

### Sample structure

Of all domestic 3202 strains, 2745 isolates (85.7%) belonged to lineage 2 (94.4% belonged to sublineage 2.2.1), 443 (13.8%) isolates belonged to lineage 4, only 14 isolates belonged to other lineages (lineage 1 and lineage 3). We constructed maximum likelihood phylogenetic trees for lineage 2 and lineage 4 *M. tuberculosis* isolates, respectively (Fig. 1a and b). The results of strain clustering showed that 1462 strains in lineage 2 were grouped into 446 transmission groups, which were consisted of 2 to 107 isolates. In lineage 4, a total of 52 clusters contained 132 isolates, ranging in size from 2 to 9 isolates (Table 1). It should be noted that lineage 2 strains had a considerably larger proportion of strains in transmission clusters than lineage 4 strains (53.3% vs. 29.8%,  $P < 0.001$ ).

### The effect of PE/PPE gene mutations on transmission of L2 strains

A comprehensive analysis revealed 1,141 homoplastic SNPs in lineage 2 strains, as detailed in Supplementary

Table 2. After excluding SNPs with a MAF below 0.005, a total of 140 homoplasic SNPs from the PE/PPE gene region were selected for further analysis. Comparing clustered and non-clustered strains, 45 mutation sites with statistical significance ( $P < 0.05$ ) in the univariate regression analysis. To further investigate these associations, the 59 loci with  $P$  values less than 0.2 in the univariate analysis were included in a multivariate logistic regression analysis. The results indicated that 9 sites were identified as influencing factors ( $P < 0.05$ ), with PE4 (position 190,394; *c.46G>A*; OR, 2.183; 95% CI, 1.025–4.651), PE\_PGRS10 (839,194; *c.744 A>G*; OR, 1.668; 95% CI, 1.220–2.280), PE16 (1,607,005; *c.620T>G*; OR, 3.741; 95% CI, 2.039–6.864) and PE\_PGRS44 (2,921,883; *c.333 C>A*; OR, 12.664; 95% CI, 1.696–94.357) considered as risk factors for strain clustering (Table 2).

The 382 strains belonging to lineage 2 formed 77 cross-regional clusters, ranging from 2 to 6 geographic regions. Among the 7 geographic regions, Northern China (31.9%) and Central China (27.3%) exhibited the highest proportion of these cross-regional clusters, followed by and Southwest China (23.1%) and Northwest China (19.5%). In the univariate analysis, 57 SNPs exhibited statistically significant differences between cross-regional and regional clusters ( $P < 0.05$ ). Subsequent multivariate logistic regression analysis identified 10 mutations as influencing factors ( $P < 0.05$ ), with 4 mutation positions recognized as risk factors for cross-regional clusters, including PE\_PGRS10 (839,334; *c.884 A>G*; OR, 2.706; 95% CI, 1.081–6.774), PE\_PGRS11 (847,613; *c.1455G>C*; OR, 4.342; 95% CI, 1.636–11.525), PE\_PGRS47 (3,054,724; *c.811 A>G*; OR, 2.099; 95% CI, 1.211–3.637) and PPE66 (4,189,930; *c.303G>C*; OR, 6.511; 95% CI, 1.679–25.242) (Supplementary Table 3).

The correlation analysis between mutation sites and cluster size revealed that 19 mutation positions were significantly associated with cluster size ( $P < 0.05$ ), with 13 mutation positions positively correlated with clustering size ( $r_s > 0$ ), including PE\_PGRS1 (132,417), PE\_PGRS6 (623,472), PE\_PGRS9 (836,658), PE16 (1,607,005), PPE26 (2,027,484), PPE34 (2,165,286), PPE35 (2,167,926), PPE44 (3,079,877), PPE54 (3,736,628), PPE56 (3,762,013), PE\_PGRS58 (4,032,218), PE\_PGRS58 (4,032,760) and PPE69 (4,375,628). For further details refer to Fig. 2.

#### The effect of PE/PPE gene mutations on transmission of L4 strains

A total of 205 homoplasic SNPs were detected in lineage 4 strains, as presented in Supplementary Table 4 for reference. After excluding homoplasic SNPs with a MAF below 0.005, 74 SNPs in the PE/PPE gene region were selected for in-depth examination. A significant difference in 6 PE/PPE gene mutation positions was detected between clustered and non-clustered strains, as per the

single-factor analysis ( $P < 0.05$ ). To further investigate these associations, a multivariate logistic regression analysis was conducted, focusing on the 22 loci with  $P$ -values less than 0.2 from the initial univariate analysis. However, no mutations were observed in lineage 4 strains that seemed to facilitate transmission, as displayed in Table 3.

Furthermore, 25 lineage 4 strains grouped into 9 cross-regional clusters, with strains in each cluster spanning two different geographic regions. After conducting univariate analysis, 9 mutation sites were selected for a multivariate regression analysis, revealing that 4 positions were significantly associated with cross-regional clusters ( $P < 0.05$ ). PE\_PGRS4 (338,100; *c.974 A>G*; OR, 6.090; 95% CI, 1.702–21.793) and PPE13 (976,897; *c.1307 A>C*; OR, 3.505; 95% CI, 1.103–11.132) considered as risk factors for cross-regional transmission of strains (Supplementary Table 5). Due to the low prevalence of lineage 4 strains in China and the relatively small sample size, we did not further analyze the transmission cluster size of lineage 4 strains.

#### Discussion

We analyzed 3202 domestic isolates (including the 1447 Shandong isolates and 1755 publicly available isolates) in this study. The genomic analysis of *M. tuberculosis* across the country reveals that lineage 2 is the predominant strain, contributing significantly to the tuberculosis burden in China, followed by lineage 4. Moreover, lineage 2 strains had a considerably larger proportion of strains in transmission clusters than lineage 4 strains. In subsequent analyses, we identified 4 PE/PPE gene mutations associated with the spread of lineage 2 strains, but no mutations were found to enhance the transmission of lineage 4 strains.

Homoplasic mutations are mutations independently occurring in different clades of an organism. Homoplasic changes may be a result of convergence evolution due to selective pressures [39]. Previous reports have shown that homoplasic SNPs were present in all functional categories of genes, with PE/PPE gene family having the highest ratio of homoplasic SNPs compared to the total SNPs identified in the same functional category [34], but the relationship between these SNPs within the PE/PPE gene region and strain transmission has not been described. In our study, the results of homoplasy analysis of strains showed that 1,141 homoplasic SNPs were identified in strains of lineage 2 and 78 homoplasic SNPs were confirmed in strains of lineage 4, respectively.

The PE/PPE genes are especially abundant in pathogenic mycobacteria, suggesting that they play a major role in mycobacterial survival and pathogenesis, although the precise function of these proteins is largely unknown [6, 40]. Based on our findings, we observed a missense

**Table 2** Analysis of the PE/PPE gene mutations in clustering and non-clustering of lineage 2

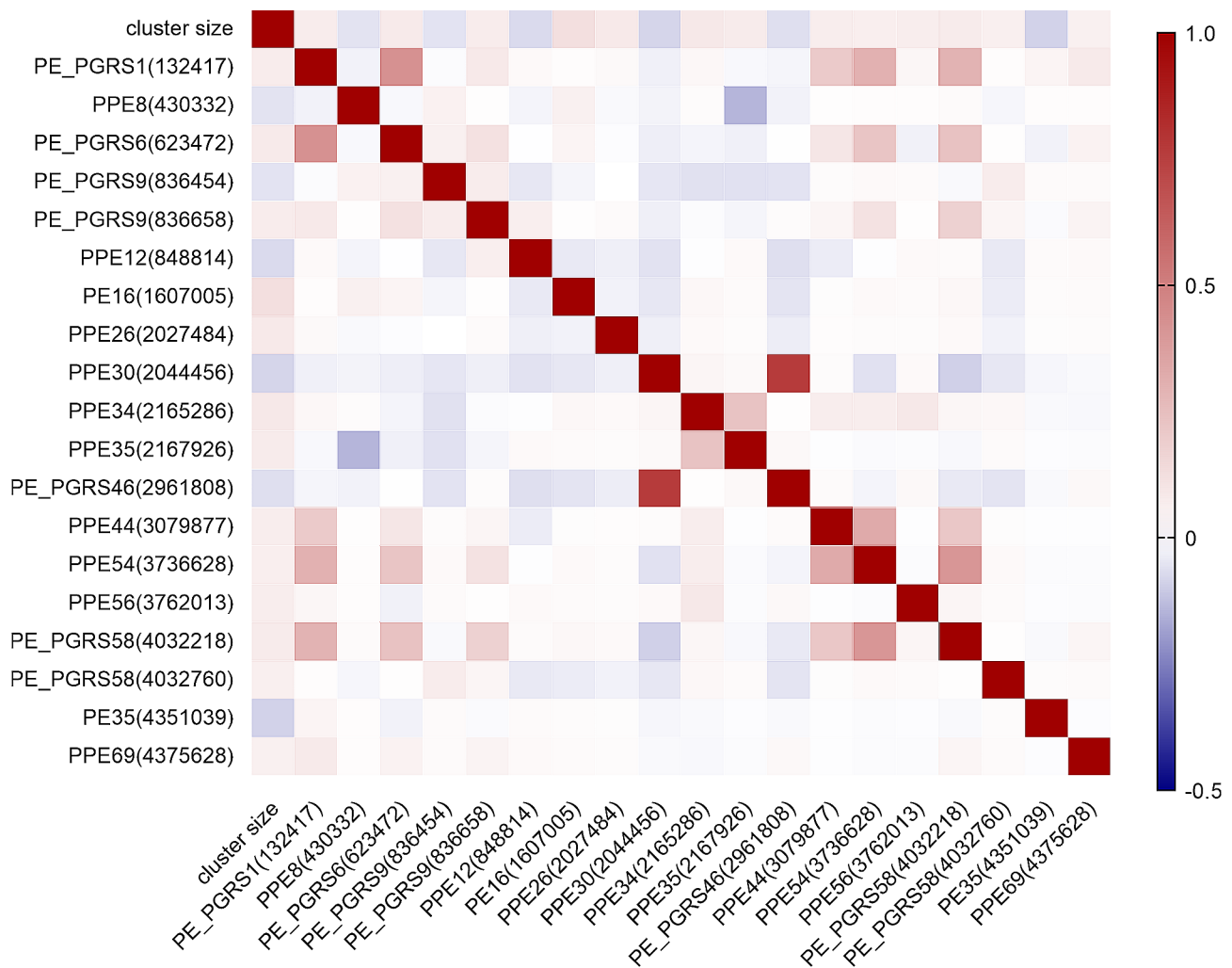
Position	Gene	SNP	Effect	MAF	OR(95%CI)	P-value
188,800	PE3	c.40A > G	p.Thr14Ala	0.928	1.389 (0.254–7.587)	0.704
190,394	PE4	c.46G > A	p.Ala16Thr	0.009	<b>2.183 (1.025–4.651)</b>	<b>0.043</b>
333,892	PE_PGRS3	c.2419C > G	p.Arg807Gly	0.375	1.277 (0.952–1.712)	0.103
337,820	PE_PGRS4	c.1254C > T	p.Gly418Gly	0.085	-	0.998
338,020	PE_PGRS4	c.1054T > G	p.Cys352Gly	0.059	-	0.998
338,100	PE_PGRS4	c.974A > G	p.Asn325Ser	0.072	-	0.998
340,132	PPE3	c.769G > A	p.Glu257Lys	0.823	-	-
340,372	PPE3	c.1009T > C	p.Ser337Pro	0.886	0.656 (0.383–1.124)	0.125
340,885	PPE3	c.1522G > A	p.Gly508Ser	0.011	0.669 (0.334–1.342)	0.258
372,913	PPE6	c.2799T > G	p.Gly933Gly	0.813	<b>0.311 (0.116–0.834)</b>	<b>0.020</b>
430,332	PPE8	c.4348T > C	p.Phe1450Leu	0.013	0.478 (0.046–4.991)	0.537
531,901	PPE10	c.314C > A	p.Ala105Glu	0.008	0.653 (0.154–2.759)	0.562
673,564	PE_PGRS7	c.2353G > A	p.Ala785Thr	0.190	1.251 (0.945–1.656)	0.118
836,087	PE_PGRS9	c.387C > T	p.Gly129Gly	0.007	2.273 (0.429–12.045)	0.335
836,426	PE_PGRS9	c.726A > C	p.Leu242Leu	0.109	1.412 (0.673–2.962)	0.362
836,454	PE_PGRS9	c.754A > G	p.Thr252Ala	0.111	0.53 (0.244–1.151)	0.109
836,538	PE_PGRS9	c.838A > G	p.Asn280Asp	0.508	1.099 (0.89–1.358)	0.381
838,990	PE_PGRS10	c.540C > G	p.Ala180Ala	0.288	<b>0.477 (0.356–0.638)</b>	<b>&lt;0.001</b>
839,194	PE_PGRS10	c.744A > G	p.Thr248Thr	0.302	<b>1.668 (1.220–2.280)</b>	<b>0.001</b>
839,684	PE_PGRS10	c.1234G > A	p.Ala412Thr	0.012	2.199 (0.196–24.727)	0.523
847,613	PE_PGRS11	c.1455G > C	p.Glu485Asp	0.017	<b>0.462 (0.271–0.789)</b>	<b>0.005</b>
1,212,432	PE_PGRS21	c.873C > A	p.Gly291Gly	0.307	0.863 (0.630–1.182)	0.360
1,218,896	PE_PGRS22	c.2428G > A	p.Gly810Ser	0.399	0.960 (0.770–1.197)	0.720
1,299,305	PPE17	c.500C > T	p.Pro167Leu	0.784	-	0.997
1,561,939	PPE20	c.171G > C	p.Glu57Asp	0.743	4.322 (0.917–20.359)	0.064
1,572,825	PE_PGRS25	c.1033G > C	p.Ala345Pro	0.005	-	0.999
1,606,673	PE16	c.288G > T	p.Ala96Ala	0.800	0.498 (0.13–1.91)	0.309
1,607,005	PE16	c.620T > G	p.Leu207Arg	0.020	<b>3.741 (2.039–6.864)</b>	<b>&lt;0.001</b>
1,655,943	PE_PGRS29	c.779T > A	p.Phe260Tyr	0.025	1.657 (0.519–5.291)	0.394
1,862,424	PE_PGRS30	c.2959G > A	p.Gly987Ser	0.006	12.414 (0.727–211.884)	0.082
1,931,718	PPE22	c.937C > G	p.Leu313Va	0.794	0.338 (0.109–1.05)	0.061
2,027,030	PPE26	c.241G > A	p.Ala81Thr	0.037	3.019 (0.58–15.721)	0.189
2,027,484	PPE26	c.695G > A	p.Gly232Asp	0.010	2.375 (0.992–5.688)	0.052
2,041,230	PPE28	c.1778A > G	p.Tyr593Cys	0.005	-	-
2,045,310	PE_PGRS32	c.1533T > C	p.Ile511Ile	0.918	0.556 (0.202–1.531)	0.256
2,168,604	PPE35	c.2009C > T	p.Pro670Leu	0.006	<b>0.202 (0.064–0.633)</b>	<b>0.006</b>
2,226,875	PE_PGRS35	c.632G > C	p.Gly211Ala	0.018	-	-
2,424,370	PE_PGRS38	c.469G > A	p.Ala157Thr	0.018	-	-
2,601,760	PE23	c.1030G > A	p.Ala344Thr	0.792	1.556 (0.203–11.901)	0.670
2,634,732	PPE39	c.861C > T	p.Ser287Ser	0.009	2.333 (0.162–33.604)	0.534
2,796,232	PE_PGRS42	c.1154C > T	p.Ala385Val	0.006	-	0.999
Position	Gene	SNP	Effect	MAF	OR(95%CI)	P-value
2,803,711	PE_PGRS43	c.2526C > G	p.Gly842Gly	0.005	-	0.999
2,921,883	PE_PGRS44	c.333C > A	p.Pro111Pro	0.007	<b>12.664 (1.696–94.357)</b>	<b>0.013</b>
2,943,839	PE_PGRS45	c.1147A > C	p.Ser383Arg	0.011	0.620 (0.130–2.955)	0.549
2,961,808	PE_PGRS46	c.634G > A	p.Gly212Ser	0.051	1.254 (0.908–1.731)	0.170
3,054,321	PE_PGRS47	c.408A > G	p.Gly136Gly	0.917	<b>0.197 (0.044–0.893)</b>	<b>0.035</b>
3,163,655	PE_PGRS48	c.1388G > C	p.Gly463Ala	0.009	1.17 (0/197–6.943)	0.863
3,201,578	PPE45	c.443C > T	p.Ala148Val	0.009	0.279 (0.047–1.650)	0.159
3,510,429	PPE52	c.889C > G	p.Pro297Ala	0.005	6.052 (0.365–100.387)	0.209
3,729,986	PPE54	c.6950A > C	p.Glu2317Ala	0.005	-	-
3,742,483	PE_PGRS50	c.292C > T	p.Pro98Ser	0.006	0.377 (0.144–1.20)	0.055
3,779,978	PE_PGRS51	c.1411G > C	p.Gly471Arg	0.006	0.761 (0.134–4.322)	0.758

**Table 2** (continued)

3,895,585	PPE60	<i>c.1160 C&gt;T</i>	<i>p.Pro387Leu</i>	0.784	1.264 (0.156–10.218)	0.826
3,929,151	PE_PGRS53	<i>c.2583 C&gt;T</i>	<i>p.Gly861Gly</i>	0.007	0.624 (0.130–2.992)	0.555
4,032,218	PE_PGRS58	<i>c.941 C&gt;T</i>	<i>p.Ala314Val</i>	0.760	1.513 (0.851–2.691)	0.159
4,032,625	PE_PGRS58	<i>c.534 C&gt;A</i>	<i>p.Gly178Gly</i>	0.354	1.163 (0.871–1.553)	0.307
4,032,760	PE_PGRS58	<i>c.399G&gt;T</i>	<i>p.Gly133Gly</i>	0.022	1.336 (0.395–4.518)	0.641
4,190,456	PPE67	<i>c.62G&gt;T</i>	<i>p.Gly21Val</i>	0.010	1.96 (0.908–4.231)	0.086
4,351,039	PE35	<i>c.295G&gt;T</i>	<i>p.Glu99*</i>	0.944	0.321 (0.068–1.514)	0.151

SNP, single nucleotide polymorphisms; MAF, minor allele frequency; OR, Odds ratio; CI, confidence interval;

-means there is no result in statistical software or the result was too large and nonsense

**Fig. 2** Correlation analysis of PE/PPE gene mutation positions and cluster size

mutation (*c.46G>A*, *p.Ala16Thr*) at position 190,394 of PE4.

(Rv0160c), and another missense mutation (*c.620T>G*, *p.Leu207Arg*) at position 1,607,005 of PE16 (Rv1430), which have been associated with increased risk of transmission of *M. tuberculosis* within lineage 2. Despite the lingering uncertainties surrounding the exact function of PE4, recent study has shed light on its role in enhancing mycobacterial survival within macrophages [41].

PE16, a member of the serine hydrolase superfamily with esterase activity, is particularly notable for its ability to hydrolyze short- to medium-chain fatty acid esters [42]. Both PE4 and PE16 play pivotal roles in the progression of mycobacterial infections. To fully comprehend the mechanisms through which these mutations facilitate transmission, further in-depth research is imperative. A synonymous mutation at position 839,194 of PE\_PGRS10 (Rv0747) (*c.744 A>G*, *p.Thr248Thr*) has been found to

**Table 3** Analysis of the PE/PPE gene mutations in clustering and non-clustering of lineage 4

Position	Gene	SNP	Effect	MAF	OR (95% CI)	P-value
337,820	PE_PGRS4	c.1254 C>T	p.Gly418Gly	0.085	1.037 (0.565–1.903)	0.906
337,959	PE_PGRS4	c.1115T>G	p.Ile372Ser	0.045	0.469 (0.175–1.258)	0.133
338,020	PE_PGRS4	c.1054T>G	p.Cys352Gly	0.059	0.357 (0.122–1.041)	0.059
340,153	PPE3	c.790G>A	p.Glu264Lys	0.006	2.369 (0.878–6.388)	0.088
340,372	PPE3	c.1009T>C	p.Ser337Pro	0.886	1.048 (0.399–2.750)	0.925
623,508	PE_PGRS6	c.716 C>G	p.Ala239Gly	0.830	0.749 (0.385–1.461)	0.397
674,416	PE_PGRS7	c.1501G>A	p.Gly501Ser	0.047	0.949 (0.584–1.544)	0.834
836,426	PE_PGRS9	c.726 A>C	p.Leu242Leu	0.109	1.014 (0.394–2.613)	0.976
836,538	PE_PGRS9	c.838 A>G	p.Asn280Asp	0.508	0.982 (0.510–1.890)	0.957
836,658	PE_PGRS9	c.958 A>G	p.Thr320Ala	0.769	0.943 (0.438–2.030)	0.881
837,033	PE_PGRS9	c.1333 A>G	p.Thr445Ala	0.910	0.702 (0.172–2.865)	0.622
838,990	PE_PGRS10	c.540 C>G	p.Ala180Ala	0.288	1.027 (0.329–3.209)	0.963
839,334	PE_PGRS10	c.884 A>G	p.Lys295Arg	0.369	0.465 (0.099–2.187)	0.332
839,348	PE_PGRS10	c.898 A>G	p.Ser300Gly	0.348	2.013 (0.431–9.394)	0.374
927,110	PE_PGRS13	c.1750 A>G	p.Ser584Gly	0.297	1.839 (0.642–5.268)	0.256
927,385	PE_PGRS13	c.2025 A>G	p.Gly675Gly	0.330	1.791 (0.521–6.162)	0.355
2,026,619	PE18	c.143 C>G	p.Ser48Trp	0.010	1.865 (0.826–4.211)	0.134
2,165,286	PPE34	c.2026T>G	p.Ser676Ala	0.902	1.891 (0.599–5.966)	0.277
2,387,733	PE_PGRS37	c.240 A>G	p.Glu80Glu	0.893	0.853 (0.361–2.020)	0.718
3,054,724	PE_PGRS47	c.811 A>G	p.Ser271Gly	0.244	0.779 (0.294–2.069)	0.617
3,736,628	PPE54	c.308 A>C	p.Glu103Ala	0.927	0.169 (0.016–1.785)	0.139
3,940,802	PE_PGRS55	c.1186 A>G	p.Asn396Asp	0.296	1.493 (0.668–3.336)	0.328

SNP, single nucleotide polymorphisms; MAF, minor allele frequency; OR, Odds ratio; CI, confidence interval;

affect the transmission of lineage 2 isolates. Mazandu and Mulder [43] have predicted that PE\_PGRS10 is involved in lipid metabolism, although this has yet to be experimentally verified, an important feature of mycobacterial pathogenicity. This synonym mutation can affect the spread of *M. tuberculosis*, indicating that the synonym mutations of PE/PPE genes are not all neutral mutations, which is consistent with the previous research results that the synonym mutations of yeast genes are mostly strong non-neutral mutations [44].

The protein functions of PE\_PGRS are unique to mycobacteria and are secreted or cell surface associated [45]. This suggests that they could be involved in mediating the interaction between the macrophages and the bacteria [46–48]. PE\_PGRS proteins were translocated through the plasmatic membrane in an ESX5-dependent mechanism [49, 50], once in the outer membrane PE\_PGRS proteins may “float” on the micromembrane outer leaflet and may possibly be released to exert their activity. The cell wall-anchored/secretory protein PE\_PGRS 11 (Rv0754) exhibits a significant interaction with TLR2, driving the maturation and activation of human dendritic cells (DCs), thereby enhancing their capacity to stimulate CD4+T cells [20]. Analysis of the effects of deletion or over-expression of PE\_PGRS47 (Rv2741) implicated this protein in the inhibition of autophagy in infected host phagocytes. As a functionally significant and non-redundant bacterial component, PE\_PGRS47 contributes to the modulation of both innate and adaptive

immune responses. This finding implicates PE\_PGRS47 as a potential target for enhancing antigen presentation and fostering protective immunity during vaccination or infection [51]. A missense mutation (c.1455G>C, p.Glu485Asp) at position 847,613 of PE\_PGRS11, and another missense mutation (c.811 A>G, p.Ser271Gly) at position 3,054,724 of PE\_PGRS47 have been found to promote the cross-regional spread of lineage 2 isolates.

A previous study of the influence of genomic variants on *M. tuberculosis* transmission in Malawi showed significant convergent evolution in a mutation in PPE54, which associated with transmission [24]. Notably, 68.7% of the strains in this study belonged to lineage 4, and only 3.8% belonged to lineage 2. However, in our study, we did not find that the mutation of PE54 could promote the spread of lineage 4 strains in China. One possible explanation for this could be the bacteriological diversity of the lineage 4 sublineages, which show complex population structure with 21 sublineages and 15 internal groups [35, 52]. Another possible reason for this difference is the limited sample size of lineage 4, which might have been inadequate to identify less common lineage 4 strains or associated mutations. A larger dataset might potentially yield more precise findings, suggesting that the current sample size might have introduced a degree of inaccuracy.

Our study is the first to explore the association of PE/PPE gene mutations with *M. tuberculosis* transmission in China. Nevertheless, we could illustrate some limitations. The distribution of *M. tuberculosis* isolates in the current

dataset may not accurately reflect the prevalence of TB in some areas, but our data have a large sample size and broad geographical distribution which allowed us to analyze how *M. tuberculosis* spread across the whole country. Additionally, due to the low prevalence of lineage 4 in China and the relatively small sample size, we did not subdivide the transmission cluster sizes of lineage 4 strains. The higher sensitivity of lineage specific analysis may require a larger sample size to be achieved.

## Conclusion

In summary, our work highlights two main *M. tuberculosis* lineages of transmission in China, as well as some PE/PPE gene mutations can increase the risk of MTB transmission in lineages 2 and 4, respectively, providing valuable insights for the treatment of TB. We believe that the PE/PPE family will remain a highly active area of research with various exciting features yet to be discovered.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-024-03352-y>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6

## Author contributions

FWW collected the data and drafted the manuscript. XLK analyzed the data. YJY, HJJ, TNN, LYM, WTT, LYY, HQL and ZYZ commented and revised the manuscript. HCL and YL conceptualized, designed the study, and acquired funding for the present study. All authors approved publication of the manuscript.

## Funding

This work was supported by the Department of Science & Technology of Shandong Province (CN) (Nos. 2007GG30002033 and 2017GSF218052), Natural Science Foundation of Shandong Province (CN) (No. ZR2020KH013 and ZR2021MH006), and Jinan Science and Technology Bureau (CN) (No. 201704100). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Data availability

The newly sequenced whole genome dataset of 1,447 *M. tuberculosis* strains was deposited in the NCBI Bio Project (accession number is PRJNA1002108), and 1755 other isolates were downloaded from the European Nucleotide Archive repository. (accession numbers are provided in Supplementary Table 1). Any additional data are available from the corresponding authors upon reasonable request.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

## Competing interests

The authors declare no competing interests.

Received: 8 November 2023 / Accepted: 26 May 2024

Published online: 10 June 2024

## References

- Global Tuberculosis Report. 2021. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2021>.
- GLOBAL Tuberculosis Report. 2022. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>.
- Chakaya J, Petersen E, Nantanda R, et al. The WHO Global Tuberculosis 2021 Report - not so good news and turning the tide back to end TB. *Int J Infect Dis.* 2022;124(Suppl 1):S26–9. <https://doi.org/10.1016/j.ijid.2022.03.011>.
- Farhat MR, Shapiro BJ, Kieser KJ, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet.* 2013;45(10):1183–9. <https://doi.org/10.1038/ng.2747>.
- Cole ST, Brosch R, Parkhill J, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* 1998;393(6685):537–44. <https://doi.org/10.1038/31159>.
- 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(5):901–16. <https://doi.org/10.1111/mmi.12981>.
- Adindla S, Guruprasad L. Sequence analysis corresponding to the PPE and PE proteins in *Mycobacterium tuberculosis* and other genomes. *J Biosci.* 2003;28(2):169–79. <https://doi.org/10.1007/BF02706216>.
- Tundup S, Akhter Y, Thiagarajan D, Hasnain SE. Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. *FEBS Lett.* 2006;580(5):1285–93. <https://doi.org/10.1016/j.febslet.2006.01.042>.
- Brennan MJ, Delogu G. The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends Microbiol.* 2002;10(5):246–9. [https://doi.org/10.1016/S0966-842X\(02\)02335-1](https://doi.org/10.1016/S0966-842X(02)02335-1).
- Mohareer K, Tundup S, Hasnain SE. Transcriptional regulation of *Mycobacterium tuberculosis* PE/PPE genes: a molecular switch to virulence? *J Mol Microbiol Biotechnol.* 2011;21(3–4):97–109. <https://doi.org/10.1159/000329489>.
- Kohli S, Singh Y, Sharma K, Mittal A, Ehtesham NZ, Hasnain SE. Comparative genomic and proteomic analyses of PE/PPE multigene family of *Mycobacterium tuberculosis* H<sub>37</sub>Rv and H<sub>37</sub>Ra reveal novel and interesting differences with implications in virulence. *Nucleic Acids Res.* 2012;40(15):7113–22. <https://doi.org/10.1093/nar/gks465>.
- Mishra KC, de Chastellier C, Narayana Y, et al. Functional role of the PE domain and immunogenicity of the *Mycobacterium tuberculosis* triacylglycerol hydrolase LipY. *Infect Immun.* 2008;76(1):127–40. <https://doi.org/10.1128/IAI.00410-07>.
- Sasseti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A.* 2003;100(22):12989–94. <https://doi.org/10.1073/pnas.2134250100>.
- Jiang Y, Wei J, Liu H, et al. Polymorphisms in the PE35 and PPE68 antigens in *Mycobacterium tuberculosis* strains may affect strain virulence and reflect ongoing immune evasion. *Mol Med Rep.* 2016;13(1):947–54. <https://doi.org/10.3892/mmr.2015.4589>.
- Behar SM, Divangahi M, Remold HG. Evasion of innate immunity by *Mycobacterium tuberculosis*: is death an exit strategy? *Nat Rev Microbiol.* 2010;8(9):668–74. <https://doi.org/10.1038/nrmicro2387>.
- Dheenadhayalan V, Delogu G, Brennan MJ. Expression of the PE\_PGRS 33 protein in *Mycobacterium smegmatis* triggers necrosis in macrophages and enhanced mycobacterial survival. *Microbes Infect.* 2006;8(1):262–72. <https://doi.org/10.1016/j.micinf.2005.06.021>.
- Yang G, Luo T, Sun C, et al. PPE27 in *Mycobacterium smegmatis* enhances mycobacterial survival and manipulates cytokine secretion in mouse macrophages. *J Interferon Cytokine Res.* 2017;37(9):421–31. <https://doi.org/10.1089/jir.2016.0126>.
- Choi HH, Kwon KW, Han SJ, et al. PPE39 of the *Mycobacterium tuberculosis* strain Beijing/K induces Th1-cell polarization through dendritic cell maturation. *J Cell Sci.* 2019;132(17):jcs228700. <https://doi.org/10.1242/jcs.228700>.
- Grover S, Sharma T, Singh Y, et al. The PGRS Domain of *Mycobacterium tuberculosis* PE\_PGRS protein Rv0297 is involved in endoplasmic



- reticulum stress-mediated apoptosis through toll-like receptor 4. *mBio*. 2018;9(3):e01017–18. <https://doi.org/10.1128/mBio.01017-18>.
20. Bansal K, Elluru SR, Narayana Y, et al. PE\_PGRS antigens of *Mycobacterium tuberculosis* induce maturation and activation of human dendritic cells. *J Immunol*. 2010;184(7):3495–504. <https://doi.org/10.4049/jimmunol.0903299>.
  21. Vordermeier HM, Hewinson RG, Wilkinson RJ, et al. Conserved immune recognition hierarchy of mycobacterial PE/PPE proteins during infection in natural hosts. *PLoS ONE*. 2012;7(8):e40890. <https://doi.org/10.1371/journal.pone.0040890>.
  22. Holt KE, McAdam P, Thai PVK, et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat Genet*. 2018;50(6):849–56. <https://doi.org/10.1038/s41588-018-0117-9>.
  23. Brynildsrud OB, Pepperell CS, Suffys P, et al. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci Adv*. 2018;4(10):eaat5869. <https://doi.org/10.1126/sciadv.aat5869>.
  24. Sobkowiak B, Banda L, Mzembe T, Crampin AC, Glynn JR, Clark TG. Bayesian reconstruction of *Mycobacterium tuberculosis* transmission networks in a high incidence area over two decades in Malawi reveals associated risk factors and genomic variants. *Microb Genom*. 2020;6(4):e000361. <https://doi.org/10.1099/mgen.0.000361>.
  25. Andrews S. FastQC A quality control tool for high throughput sequence data. Cambridge (UK): Babraham Bioinformatics; 2010.
  26. Jung Y, Han D, BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics*. 2022;38(9):2404–13. <https://doi.org/10.1093/bioinformatics/btac137>.
  27. Li H, Handsaker B, Wysoker A, et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
  28. Danecek P, Bonfield JK, Liddle J et al. Twelve years of SAM tools and BCF tools. *Gigascience*. 2021, 10(2).
  29. Cingolani P, Platts A, Wang le. 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)*. 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695>.
  30. Phelan JE, O'Sullivan DM, Machado D, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med*. 2019;11(1):41. <https://doi.org/10.1186/s13073-019-0650-x>.
  31. Coll F, Mc Nerney R, Guerra-Assunção JA, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun*. 2014;5:4812. <https://doi.org/10.1038/ncomms5812>.
  32. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300>.
  33. Ruesen C, Chaidir L, van Laarhoven A, et al. Large-scale genomic analysis shows association between homoplastic genetic variation in *Mycobacterium tuberculosis* genes and meningeal or pulmonary tuberculosis. *BMC Genomics*. 2018;19(1):122. <https://doi.org/10.1186/s12864-018-4498-z>.
  34. Tantivitayakul P, Ruangchai W, Juthayothin T, et al. Homoplastic single nucleotide polymorphisms contributed to phenotypic diversity in *Mycobacterium tuberculosis*. *Sci Rep*. 2020;10(1):8024. <https://doi.org/10.1038/s41598-020-64895-4>.
  35. Li YF, Yang Y, Kong XL, et al. Transmission dynamics and phylogeography of *Mycobacterium tuberculosis* in China based on whole-genome phylogenetic analysis. *Int J Infect Dis*. 2024;140:124–31. <https://doi.org/10.1016/j.ijid.2023.10.015>.
  36. Chiner-Oms Á, Sánchez-Busó L, Corander J, et al. Genomic determinants of speciation and spread of the *Mycobacterium tuberculosis* complex. *Sci Adv*. 2019;5(6):eaaw3307. <https://doi.org/10.1126/sciadv.aaw3307>.
  37. Farhat MR, Freschi L, Calderon R, loerger T, Snyder M, Meehan CJ, et al. GWAS for quantitative resistance phenotypes in *Mycobacterium tuberculosis* reveals resistance genes and regulatory regions. *Nat Commun*. 2019;10:2128. <https://doi.org/10.1038/s41467-019-10110-6>.
  38. Schrag A, Siddiqui UF, Anastasiou Z, Weintraub D, Schott JM. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study. *Lancet Neurol*. 2017;16(1):66–75. [https://doi.org/10.1016/S1474-4422\(16\)30328-3](https://doi.org/10.1016/S1474-4422(16)30328-3).
  39. Grandjean L, Gilman RH, Iwamoto T, et al. Protective evolution and topologically disruptive polymorphisms among multidrug-resistant tuberculosis in Peru. *PLoS ONE*. 2017;12(12):e0189838. <https://doi.org/10.1371/journal.pone.0189838>.
  40. Daleke MH, Cascioferro A, de Punder K, et al. Conserved Pro-glu (PE) and Pro-pro-glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. *J Biol Chem*. 2011;286(21):19024–34. <https://doi.org/10.1074/jbc.M110.204966>.
  41. Singh SK, Tripathi DK, Singh PK, Sharma S, Srivastava KK. Protective and survival efficacies of Rv0160c protein in murine model of *Mycobacterium tuberculosis*. *Appl Microbiol Biotechnol*. 2013;97(13):5825–37. <https://doi.org/10.1007/s00253-012-4493-2>.
  42. Sultana R, Vemula MH, Banerjee S, Guruprasad L. The PE16 (Rv1430) of *Mycobacterium tuberculosis* is an esterase belonging to serine hydrolase superfamily of proteins. *PLoS ONE*. 2013;8(2):e55320. <https://doi.org/10.1371/journal.pone.0055320>.
  43. Mazandu GK, Mulder NJ. Function prediction and analysis of mycobacterium tuberculosis hypothetical proteins. *Int J Mol Sci*. 2012;13(6):2783–302. <https://doi.org/10.3390/ijms13067283>.
  44. Shen X, Song S, Li C, Zhang J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature*. 2022;606(7915):725–31. <https://doi.org/10.1038/s41586-022-04823-w>.
  45. Tian C, Jian-Ping X. Roles of PE\_PGRS family in *Mycobacterium tuberculosis* pathogenesis and novel measures against tuberculosis. *Microb Pathog*. 2010;49(6):311–4. <https://doi.org/10.1016/j.micpath.2010.07.004>.
  46. Banu S, Honoré N, Saint-Joanis B, Philpott D, Prévost MC, Cole ST. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol*. 2002;44(1):9–19. <https://doi.org/10.1046/j.1365-2958.2002.02813.x>.
  47. Espitia C, Lacllette JP, Mondragón-Palomino M, et al. The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? *Microbiol (Reading)*. 1999;145(Pt 12):3487–95. <https://doi.org/10.1099/00221287-145-12-3487>.
  48. Brennan MJ, Delogu G, Chen Y, et al. Evidence that mycobacterial PE\_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun*. 2001;69(12):7326–33. <https://doi.org/10.1128/IAI.69.12.7326-7333.2001>.
  49. Bottai D, Di Luca M, Majlessi L, et al. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. *Mol Microbiol*. 2012;83(6):1195–209. <https://doi.org/10.1111/j.1365-2958.2012.08001.x>.
  50. Burggraaf MJ, Speer A, Meijers AS, et al. Type VII secretion substrates of pathogenic mycobacteria are processed by a surface protease. *mBio*. 2019;10(5):e01951–19. <https://doi.org/10.1128/mBio.01951-19>.
  51. Saini NK, Baena A, Ng TW, et al. Suppression of autophagy and antigen presentation by *Mycobacterium tuberculosis* PE\_PGRS47. *Nat Microbiol*. 2016;1(9):16133. <https://doi.org/10.1038/nmicrobiol.2016.133>.
  52. Freschi L, Vargas R Jr, Husain A, et al. Population structure, biogeography and transmissibility of *Mycobacterium tuberculosis*. *Nat Commun*. 2021;12(1):6099. <https://doi.org/10.1038/s41467-021-26248-1>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.