CrossMark

ORIGINAL PAPER

Whole genome sequence analysis of multidrug-resistant *Mycobacterium tuberculosis* Beijing isolates from an outbreak in Thailand

Sanjib Mani Regmi[1] · Angkana Chaiprasert[1] · Supasak Kulawonganunchai[2] ·
Sissades Tongsima[2] · Olabisi Oluwabukola Coker[1] · Therdsak Prammananan[2] ·
Wasna Viratyosin[2] · Iyarit Thaipisuttikul[1]

Abstract The *Mycobacterium tuberculosis* Beijing family is often associated with multidrug resistance and large outbreaks. Conventional genotyping study of a community outbreak of multidrug-resistant tuberculosis (MDR-TB) that occurred in Kanchanaburi Province, Thailand was carried out. The study revealed that the outbreak was clonal and the strain was identified as a member of Beijing family. Although, the outbreak isolates showed identical spoligotyping and mycobacterial interspersed repetitive units-variable number tandem repeats patterns, a discrepancy regarding ethambutol resistance was observed. In-depth characterization of the isolates through whole genome sequencing of the first and the last three isolates from our culture collection showed them to belong to principal genetic group 1, single nucleotide polymorphism (SNP) cluster group 2, sequence type 10. Compared with the *M. tuberculosis* H37Rv reference genome, 1242 SNPs were commonly found in all isolates. The genomes of these isolates were shown to be clonal and highly stable over a 5-year period and two or three unique SNPs were identified in each of the last three isolates. Genes known to be associated with drug resistance and their promoter regions, where applicable, were analyzed. The presence of low or no fitness cost mutations for drug resistance and an additional L731P SNP in the *rpoB* gene was observed in all isolates. These findings might account for the successful transmission of this MDR-TB strain.

Keywords *Mycobacterium tuberculosis* · Multidrug resistance · Outbreak · Whole genome sequencing · Thailand

Introduction

In spite of immense efforts to overcome it, tuberculosis (TB) infection remains a major public health issue with an estimated 9 million new cases and 1.5 million deaths from TB worldwide in 2013 (The WHO 2014). The emergence and rapid spread of multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB pose additional global threats to the global TB control programs (Gandhi et al. 2010). The World Health Organization (WHO) ranked Thailand 18th of the 22 highest TB burdened countries in the world. With a population of around 67 million, Thailand was reported to have approximately 62,000 new and 110,000 prevalent TB cases in 2013. These figures were further complicated by the growing threat of MDR and XDR-TB. According to the WHO report, approximately 2 % of new TB cases and 19 % of previously treated cases were MDR-TB in Thailand (The WHO 2014). However, despite this major public health problem, little is known about the genetic characteristics of the isolates (Viratyosin et al. 2013; Coker et al. 2014). Therefore, information gained from whole genome sequencing (WGS) of MDR

Communicated by S. Hohmann.

✉ Angkana Chaiprasert
angkana.cha@mahidol.ac.th

[1] Department of Microbiology, Faculty of Medicine
Siriraj Hospital, Mahidol University, 2 Wanglang Road,
Bangkok-Noi, Bangkok 10700, Thailand

[2] National Center for Genetic Engineering and Biotechnology,
National Science and Technology Development Agency,
Pathum Thani 12120, Thailand

M. tuberculosis outbreak isolates could improve our understanding about the epidemics in this region.

With a prevalence of about 50 % of all Asian TB cases (van Soolingen et al. 1995; Parwati et al. 2010) and a worldwide dissemination (Bifani et al. 2002; Glynn et al. 2002; European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis 2006), the *M. tuberculosis* Beijing family has gained considerable attention. It is considered virulent (Parwati et al. 2010) and is often associated with drug resistance (Drobniewski et al. 2005; Niemann et al. 2010; Casali et al. 2012) and large outbreaks (Bifani et al. 1999; Toungoussova et al. 2002, 2003; Ioerger et al. 2010; Golesi et al. 2013). A previous epidemiological study of TB in Thailand revealed a large outbreak of MDR-TB in Kanchanaburi Province during 2002–2010 (Jiraphongsa et al. 2011). The genotyping of 64 isolates from 2003 to 2008 using spoligotyping and 24-loci mycobacterial interspersed repetitive units-variable number of tandem repeats (MIRU-VNTR) resulted in the clustering of 54 isolates indicating a clonal outbreak in the region. The strain was identified as a member of the Beijing family (Srilohasin 2013). Although these isolates had isogenic genotyping patterns, a discrepancy regarding ethambutol resistance was observed. Moreover, the extensive spread of this strain in the community indicated that it might harbor some genetic determinants to aid efficient transmission.

A recent study in a TB high-burden setting where the *M. tuberculosis* Beijing family was prevalent suggested a threshold of ≤5 single nucleotide polymorphisms (SNPs) to define strains in a transmission chain (Luo et al. 2014). However, the mutation rates of the *M. tuberculosis* Beijing family remain controversial (Werngren and Hoffner 2003; de Steenwinkel et al. 2012; Ford et al. 2013). For example, an in vivo experimental infection study on non-human primates under antibiotics therapy determined that the Beijing family had a higher mutation rate than other lineages (Ford et al. 2013) and a high degree of genetic diversity was reported among the serial isolates of the *M. tuberculosis* Beijing family obtained from the patients during the course of anti-TB therapy (Sun et al. 2012). However, fluctuation analysis, did not find any difference in mutation rate between Beijing and non-Beijing isolates (Werngren and Hoffner 2003). Additionally, next generation sequencing (NGS) analysis of Beijing family serial isolates recently determined that 8–9 SNPs were acquired over a period of 3 years suggesting low mutation rate in human host (Merker et al. 2013). Considering these controversial findings, it would be beneficial to carry out a direct investigation of the microevolution of clinical isolates in each setting.

In this study, therefore, WGS was performed on the first isolate (isolated in 2003) and the last three isolates (isolated in 2008) of the defined cluster from our culture collection. WGS analysis of the isolates also suggested a clonal spread of the strain, and that the outbreak might be attributed to the presence of drug resistance conferring mutations associated with low or no fitness costs in the strain. We also determined the acquired SNPs in the isolates accumulated over a 5-year period and found that they were genetically very stable with a maximum number of three unique SNPs compared with the first isolate. To the best of our knowledge, this study represents the first comprehensive analysis of MDR-TB isolates involved in a large community outbreak in Thailand.

Materials and methods

Selection of isolates

A retrospective cohort study by an epidemiological team identified 148 MDR-TB cases and confirmed a community outbreak of MDR-TB during 2002–2010 in Kanchanaburi Province, Thailand (Jiraphongsa et al. 2011). As a reference laboratory, we obtained the specimens from 2003 to 2008 for drug susceptibility testing (DST). Spoligotyping and 24-loci MIRU-VNTR of all 64 isolates were performed (Kamerbeek et al. 1997; Supply et al. 2006; Srilohasin 2013), and clustering of 54 isolates was observed suggesting a clonal outbreak of MDR-TB in this cohort. Spoligotyping and MIRU-VNTR typing of the isolates corresponded to SIT-1, MIT-17, and VIT-70 of the modern Beijing sub-lineage in the SITVITWEB database (Demay et al. 2012; Srilohasin 2013). The first isolate (DS-5538; isolated in August, 2003) and the last three isolates of the cluster (DS-17355, DS-17471, and DS-17472; isolated in March 2008) were selected from our culture collection. Isolates DS-5538, DS-17355, and DS-17471 showed identical DNA fingerprints and phenotypic DST results, and were selected to observe the cumulative genetic changes over the 5-year period. Among the 54 clustered isolates, 23 were reported to be susceptible to ethambutol. Isolate DS-17472, reportedly susceptible to ethambutol, was selected to observe whether it belongs to the same clone. All isolates were obtained from sputum specimens at the Drug-Resistant TB Laboratory Department of Microbiology, Faculty of Medicine Siriraj Hospital, Thailand.

Phenotypic drug susceptibility testing

Testing of the first- and the second-line anti-TB drugs was performed by the standard agar proportion method (WHO Geneva/IUATLD Paris 1998). Drug concentrations of 0.2 mg/l isoniazid, 1.0 mg/l rifampicin and linezolid, 5.0 mg/l ethambutol and ethionamide, 6.0 mg/l amikacin

Table 1 Gross statistics of whole genome sequencing and mapping of the reads

Strain ID	Total reads	Reference H37Rv length	Gross fold coverage	Mapped reads with reference (%)	Reference genome coverage (%)
DS-5538	46,498,574	4,411,532	1064.56	98.95	99.31
DS-17355	42,813,596	4,411,532	980.2	99.06	99.32
DS-17471	23,765,326	4,411,532	544.1	98.97	99.29
DS-17472	47,119,718	4,411,532	1078.78	99.0	99.3
Average	40,049,303	4,411,532	916.91	98.99	99.3

and kanamycin, and 2.0 mg/l streptomycin, para-aminosalicylic acid, ofloxacin, ciprofloxacin, levofloxacin, moxifloxacin and gatifloxacin were used for drug susceptibility testing.

DNA isolation and whole genome sequencing

Selected isolates were subcultured on Lowenstein–Jensen medium for 4 weeks at 37 °C. DNA extraction and purification were carried out using cetyltrimethylammonium bromide lysozyme method (Larsen et al. 2007). Sequencing of the isolates was carried out at Macrogen Inc. (Seoul, South Korea). Genomic libraries were prepared according to the recommendations of the TrueSeq DNA sample preparation kit (Illumina, San Diego, CA). The library pools were subjected to paired-end sequencing on a HiSeq 2000 platform (Illumina) generating 100-bp read lengths.

Reads mapping, SNP calling and confirmation

Paired-end raw reads of each isolate were independently aligned to the *M. tuberculosis* H37Rv reference genome (GenBank accession number: NC_000962.2) using Bowtie 2 version 2.2.0 (Langmead and Salzberg 2012). Bedtools version 2.20.1 (https://github.com/arq5x/bedtools2) was used to determine the reads coverage over the reference genome. Aligned reads of each isolate were sorted, indexed and combined into an mpileup file using SAMtools (Li et al. 2009). Single nucleotide variant (SNV) identification was performed using VarScan 2.2.11 (Koboldt et al. 2012). To ensure the quality of the SNVs, bases with Phred quality score of ≤20 and SNVs with coverage of fewer than 10 reads were discarded. Additionally, heterozygous SNVs with allele frequencies of <75 % that were commonly present in all four isolates were discarded, as they likely originated from mapping errors. The remaining variants were annotated using H37Rv annotations (GenBank accession number: NC_000962.2) and classified as synonymous, nonsynonymous, or intergenic.

When analyzing short reads, sequences of repetitive regions and paralogous gene families are known to be problematic because the short reads can be mapped to multiple loci. Thus, variants determined in PE, PPE, PE_PGRS, integrase, transposase and phage-related genes were discarded (Comas et al. 2010). Overall filtration processes and annotation were performed using in-house software written in Ruby code. All candidate SNV positions were then carefully observed in Integrative Genomics Viewer (Thorvaldsdottir et al. 2013). Some heterozygous variants with allele frequencies just above the borderline (between 75 and 80 %) in some isolates and <75 % in other isolates were curated manually because they were also likely to have been created from mapping errors. SNV filtration and curation parameters were chosen on the basis of previous work carried out in *M. tuberculosis* (Merker et al. 2013; Perez-Lago et al. 2014). All unique SNVs identified in each isolate were further validated by PCR amplification and Sanger sequencing.

Ethical approval

Ethical approval was obtained from the Institutional Review Board Committee of the Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand (Protocol No. 811/2556; EC2). This article does not contain any studies with human participants or animals performed by any of the authors.

Results

Whole genome sequencing

All selected isolates were successfully sequenced and the data were analyzed as described in "Materials and methods". The average number of reads yielded per isolate was 40,049,303 with a mean sequencing depth of 916X, when aligned to the H37Rv reference genome. An average of 98.99 % of the reads was successfully mapped to the reference. More than 99 % of the reference genome was shown to be covered by at least one aligned read. The gross statistics of alignment and coverage is summarized in Table 1. Remaining SNVs after filtration were included for subsequent analyses.

The resulting Fastq files from the four isolates were deposited in the NCBI short read archive (SRA) under accession numbers SRX691156, SRX691468, SRX691500, and SRX691501 for isolates DS-5538, DS-17355, DS-17471, and DS-17472, respectively.

Genetic variations

Compared with the H37Rv reference, 1242 common SNPs were identified in all isolates (Online Resource 1, Fig. 1). Of these, 162 were intergenic and 1080 were found in protein coding regions, of which 406 were synonymous and 674 were nonsynonymous. Comparative SNP analysis identified three, three and two unique SNPs in DS-17355, DS-17471 and DS-17472, respectively (Fig. 1; Table 2). Nonsynonymous mutations were found in 113 of the previously defined 760 essential genes (Comas et al. 2010). Among the eight unique SNPs, only one nonsynonymous SNP was observed in the essential gene *pks12* (Table 2).

The genetic background of the isolates was determined by WGS data. To characterize the principle genetic group (PGG) (Sreevatsan et al. 1997), allelic determination at *katG* 463 and *gyrA* 95 positions was performed. The presence of *katG* R463L and *gyrA* S95T alleles in all isolates confirmed that they belonged to PGG-1. *In silico* analysis of alignments in Integrated Genomics Viewer to elucidate regions of difference (RDs) (Tsolaki et al. 2005; Dou et al. 2008) showed that the isolates belonged to RD-type 4. Furthermore, SNP analysis for the determination of sequence type (ST) (Filliol et al. 2006) revealed that all isolates belong to SNP cluster group (SCG)-2, ST-10. All findings, including spoligotyping and MIRU-VNTR typing, were in agreement and confirmed the isolates to be clonal and belong to a modern sublineage of the *M. tuberculosis* Beijing genotype.

Drug resistance mutations

The isolates were tested for phenotypic drug susceptibility to the first- and the second-line drugs (Online Resource 1). All isolates were resistant to rifampicin, isoniazid and streptomycin, and all except DS-17472 were reported to be resistant to ethambutol. Well-known polymorphisms that correlated with phenotypic drug resistance were observed in all isolates (Table 3).

Rifampicin resistance is caused by specific mutations in *rpoB* which encodes the beta subunit of RNA polymerase. The presence of the *rpoB* S450L mutation in all strains conferred rifampicin resistance, and a nonsynonymous *rpoB* L731P mutation was also detected in all strains. No mutations were identified in *rpoA* and *rpoC* in any of the isolates, although mutations in these genes
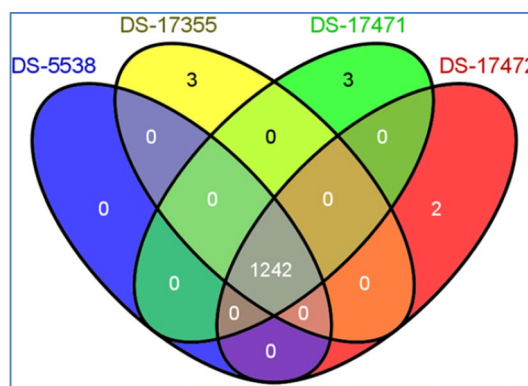

Fig. 1 Venn diagram showing overview of SNP distribution among the isolates. A total of 1242 SNPs were shared among the isolates. Three SNPs were unique to DS-17355 and DS-17471, and two SNPs to DS-17472

could have explained the compensatory role for fitness cost. Resistance to isoniazid is a complex process and has been explained by mutation in several genes including *katG, inhA, ahpC*, and *ndh*. The nonsynonymous mutation *katG* S315T was observed in all four isolates studied and conferred isoniazid resistance. Resistance to streptomycin can result from mutations in *rrs* or *rspL*, with the most common mutation being K43R in *rspL*. All four isolates carried the K43R mutation in *rspL*, conferring streptomycin resistance. Ethambutol resistance is most likely caused by the overexpression or structural variations in *embB*. Resistance to ethambutol in the isolates was conferred by the presence of the G406D *embB* mutation. However, the isolate DS-17472, which was reported susceptible to ethambutol, was also found to harbor the G406D mutation in *embB*. Those genes and their applicable promotor regions known to be associated with resistance to anti-TB drugs, and listed in TB drug resistance mutation database (Sandgren et al. 2009), were analyzed, but no other well defined or novel mutation that could be correlated to drug resistance was found.

It has been suggested that MDR in *M. tuberculosis* could be associated with the constitutive or inducible expression of efflux pump-related genes (Calgin et al. 2013; Black et al. 2014). However, the role of mutations in these genes in causing drug resistance has been poorly explored (Liu et al. 2014). We suspected that the mutations could have a cumulative effect on the formation of drug resistance in MDR-TB strains. We therefore investigated mutations in 41 known or putative drug efflux-related genes (Black et al. 2014) and identified a total of 10 nonsynonymous SNPs of which nine had previously been reported in the pansusceptible *M. tuberculosis* Beijing strain (Niemann et al. 2009). A mutation (H462N) in Rv1877 that encodes a conserved

Table 2 Strain-specific SNPs identified in the present study

Position	Base change	Amino acid change	Locus	Gene symbol	Annotations	Comments	Strain
306109	A>G	C81R	Rv0254c	cobU	Bifunctional cobalamin biosynthesis protein	Nonessential	DS-17355
2295057	G>A	S3977L	Rv2048c	pks12	Polyketide synthase Pks12	Essential	DS-17355
2374769	T>G	M508L	Rv2115c		ATPase	Nonessential	DS-17471
3265853	G>A	Syn	Rv2934	ggtB	Gamma-glutamyltranspeptidase precursor	Nonessential	DS-17471
3889221	T>A	R41[a]	Rv3471c		Conserved hypothetical protein	Nonessential	DS-17472
4007219	G>A	Syn	Rv3565	aspB	Aspartate aminotransferase	Nonessential	DS-17471
4080776	T>C	–	Intergenic	–	–	–	DS-17472
4343282	C>A	Y171[a]	Rv3867		Conserved hypothetical protein	Nonessential	DS-17355

Syn synonymous

[a] Stop codon

Table 3 DST phenotype and drug resistance-confirming mutations identified in studied isolates

Anti-TB drug	Drug concentrations (mg/l)	DST phenotype	Gene symbol	Synonym	Position	Base change	Amino acid change	References
Rifampicin	1.0	Resistance	rpoB	Rv0667	761155	C>T	S450L	(Telenti et al. 1993)
Isoniazid	0.2	Resistance	katG	Rv1908	2155168	C>G	S315T	(Zhang et al. 1992)
Streptromycin	2.0	Resistance	rspL	Rv0682	781687	A>G	K43R	(Nair et al. 1993)
Ethambutol	5.0	Resistance[a]	embB	Rv3795	4247730	G>A	G406D	(Telenti et al. 1997; Ramaswamy et al. 2000)

[a] Isolate DS-17472 was susceptible to ethambutol

membrane protein was observed to be novel in all isolates studied.

Discussion

In recent years, rapid WGS using NGS platforms has unraveled the genetic make-up of microorganisms at a high resolution. In this study, we compared the genetic variations in outbreak isolates with the H37Rv reference and also identified the genetic background of these outbreak isolates as PGG-1, SCG-2; ST-10 of modern Beijing family of *M. tuberculosis* (Sreevatsan et al. 1997; Filliol et al. 2006). ST-10 of Beijing family has previously been reported to be predominant in many countries (Chen et al. 2012; Iwamoto et al. 2012) including Thailand (Faksri et al. 2011), although further research is required to explore the molecular biology underlying the emergence and rapid spread of this genotype.

Predicted genotype susceptibility results were identical to phenotype DST results in the present study, except for DS-17472. Although this isolate was reported to be susceptible to ethambutol, we observed the presence of the G406D *embB* mutation which confers the drug resistance (Ramaswamy et al. 2000). Phenotypic DST repeated in triplicate

for this isolate revealed the same result. This could be explained by the fact that the G406D mutation has been reported to confer only low level resistance to ethambutol (3.3–7.6 mg/l) (Safi et al. 2013), and that previous studies also reported this mutation in ethambutol-susceptible isolates (Lee et al. 2004; Park et al. 2012). Moreover, a re-evaluation of the current breakpoint for ethambutol (5 mg/l for 7H10 medium) has been highly debated (Schon et al. 2009; Gumbo 2010), while phenotype DST has been suggested to underreport ethambutol resistance (Johnson et al. 2006; Ioerger et al. 2010). Overall, the presence of genetically identical drug resistance conferring SNPs among the isolates further illustrates their clonality.

To identify a novel mechanism that could be correlated with drug resistance, our analysis revealed the presence of classic mutations in all drug resistance related genes studied except for the L731P mutation in *rpoB*. This SNP was previously reported as unique in the MDR outbreak strain X122 from Western Cape, South Africa (Ioerger et al. 2010). Recently, de Vos et al. reported that strains harboring the S450L *rpoB* mutation with compensatory mutations in RNA polymerase genes were associated with ongoing transmission of MDR-TB in the community (de Vos et al. 2013). Hence, a functional study is required to confirm whether this mutation enables the strain to tolerate the

fitness cost associated with drug resistance, or enhance its transmissibility. The analysis of mutations in drug efflux pump-related genes revealed the novel mutation H462N in Rv1877 which encodes a conserved membrane protein. The *lfrA* gene, homolog to Rv1877, was previously found to induce resistance to erythromycin in *M. smegmatis* (Li et al. 2004). However, it is not clear how mutations in Rv1877 affect susceptibility to ethambutol, so further research is necessary to confirm its function in causing drug resistance.

Because this MDR-TB strain was successfully transmitted in the community, there was a need to unravel the genetic determinants responsible for this transmission. Acquired drug resistance in *M. tuberculosis* strains is more often associated with the reduced fitness that might affect growth, stability, or transmission (Andersson 2006). Unlike other lineages of *M. tuberculosis*, the Beijing family is thought to show more potential in adopting the fitness cost by a genetic-specific capabilities that acquire low or no fitness cost mutations for drug resistance (Borrell and Gagneux 2009; Gagneux 2009) or by the most favorable epistatic interactions between drug resistance and compensatory mutations (Comas et al. 2012; Muller et al. 2013). Several studies have shown that MDR-TB strains harboring low or no fitness costs were better transmitted in the community than those with other mutations (van Soolingen et al. 2000; Gagneux et al. 2006a; Strauss et al. 2008; Naidoo and Pillay 2014). Interestingly, the outbreak strain in this study was also found to harbor low or no fitness cost mutations in *rpoB* (S450L), *katG* (S315T) and *embB* (G406D) which confer drug resistance to rifampicin, isoniazid and ethambutol, respectively (Pym et al. 2002; Gagneux et al. 2006b; Safi et al. 2013). Overall, the successful transmission of the strain resulting in a large community outbreak may be attributed to the presence of drug resistance conferring mutations associated with low or no fitness costs, or the additional effect of the L731P *rpoB* mutation.

SNPs accumulated over a 5-year period in the selected isolates of the cluster (Srilohasin 2013) were determined. Although the mutation rate of the Beijing family remains controversial (Werngren and Hoffner 2003; de Steenwinkel et al. 2012; Sun et al. 2012; Ford et al. 2013; Merker et al. 2013), the outbreak strain was found to be genetically stable over 5 years, in line with previous findings (Werngren and Hoffner 2003; Merker et al. 2013), because only two to three SNPs were found to be acquired by each of the last three isolates. This is in agreement with a previous study by Schurch et al. in which NGS was applied for the WGS of three isolates obtained over 14-year period. A maximum of four SNPs were acquired compared with the first isolate (Schurch et al. 2010). Recently, a model similar to our study was used to determine the genomic variation in a *M.*

tuberculosis outbreak strain belonging to the T2 sublineage. Compared with the index case, the strain demonstrated genomic stability over 9 years with only four acquired SNPs and a small deletion (Sandegren et al. 2011). Interestingly, the number of SNPs identified in our study is similar to that of previous studies despite differences in the genetic background of the strain and TB burden settings.

WGS of *M. tuberculosis* strains among epidemiologically linked patients in both TB low and high-burden settings revealed that epidemiologically linked TB strains can be genetically linked by five or fewer SNPs (Kato-Maeda et al. 2013; Roetzer et al. 2013; Walker et al. 2013; Luo et al. 2014). In our study, two or three SNPs were identified among the isolates studied, which supports these previous findings and may be useful in establishing epidemiological links among TB patients in high-TB burden settings where the *M. tuberculosis* Beijing family is predominant.

High molecular weight genomic DNA used in this study was obtained from isolate subcultures derived from stock cultures, which might affect the exact genetic make-up of the bacteria. Although, previous studies did not report any substantial impact of this phenomenon (Merker et al. 2013; Roetzer et al. 2013; Walker et al. 2013), it could be considered a limitation of the present study. Additionally, SNVs in repetitive regions such as PE, PPE, PE_PGRS genes and paralogous gene families were excluded from the analysis. These genes account for approximately 10 % of the coding region of the H37Rv genome, so it is possible that the isolates studied might harbor SNPs in these regions, and that the overall variation might be higher than reported.

In conclusion, this study successfully determined the genetic polymorphisms in the outbreak isolates of *M. tuberculosis* Beijing, ST-10. The isolates were found to be clonally related despite the discrepancy in their DST phenotype, and the genome of the outbreak strain was shown to be genetically very stable over a 5-year period. We propose that performing drug susceptibility testing and carrying out treatment of MDR-TB or XDR-TB in TB high-burden settings may not be sufficient to achieve the goal of the global TB control program, but the identification and rapid screening of genetic determinants in highly transmissible strains could greatly contribute to their eradication. Our work also identified possible genetic determinants that might be responsible for the efficient transmission of the strain in our community. These findings might have important implication for confirming the epidemiological links among the TB patients in high-TB burden settings or for rapid screening of highly transmissible MDR-TB strains to prevent their successful spread in the community.

References

Andersson DI (2006) The biological cost of mutational antibiotic resistance: any practical conclusions? Curr Opin Microbiol 9:461–465

Bifani PJ, Mathema B, Liu Z, Moghazeh SL, Shopsin B, Tempalski B, Driscol J, Frothingham R, Musser JM, Alcabes P, Kreiswirth BN (1999) Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. JAMA 282:2321–2327

Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN (2002) Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol 10:45–52

Black PA, Warren RM, Louw GE, van Helden PD, Victor TC, Kana BD (2014) Energy metabolism and drug efflux in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 58:2491–2503

Borrell S, Gagneux S (2009) Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis*. Int J Tuberc Lung Dis 13:1456–1466

Calgin MK, Sahin F, Turegun B, Gerceker D, Atasever M, Koksal D, Karasartova D, Kiyan M (2013) Expression analysis of efflux pump genes among drug-susceptible and multidrug-resistant *Mycobacterium tuberculosis* clinical isolates and reference strains. Diagn Microbiol Infect Dis 76:291–297

Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, Bentley SD, Parkhill J, Nejentsev S, Hoffner SE, Horstmann RD, Brown T, Drobniewski F (2012) Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res 22:735–745

Chen YY, Chang JR, Huang WF, Kuo SC, Su IJ, Sun JR, Chiueh TS, Huang TS, Chen YS, Dou HY (2012) Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on SNP and VNTR typing profiles in Asian countries. PLoS One 7:e39792

Coker OO, Regmi SM, Suriyaphol P, Chininmanu K, Prammananan T, Chaiprasert A (2014) Whole-genome sequence of a multi-drug-resistant *Mycobacterium tuberculosis* Beijing sequence type 10 isolate from an outbreak in Thailand. Genome Announc 2(4):e00803–e00814

Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S (2010) Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet 42:498–503

Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S (2012) Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44:106–110

de Steenwinkel JE, ten Kate MT, de Knegt GJ, Kremer K, Aarnoutse RE, Boeree MJ, Verbrugh HA, van Soolingen D, Bakker-Woudenberg IA (2012) Drug susceptibility of *Mycobacterium tuberculosis* Beijing genotype and association with MDR TB. Emerg Infect Dis 18:660–663

de Vos M, Muller B, Borrell S, Black PA, van Helden PD, Warren RM, Gagneux S, Victor TC (2013) Putative compensatory mutations in the *rpoC* gene of rifampin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission. Antimicrob Agents Chemother 57:827–832

Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N (2012) SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infection, Genetics and Evolution 12:755–766

Dou HY, Tseng FC, Lin CW, Chang JR, Sun JR, Tsai WS, Lee SY, Su IJ, Lu JJ (2008) Molecular epidemiology and evolutionary genetics of *Mycobacterium tuberculosis* in Taipei. BMC Infect Dis 8:170

Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetzov S, Zakharova S, Melentyev A, Fedorin I (2005) Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. JAMA 293:2726–2731

European Concerted Action on New Generation Genetic Markers and Techniques for the Epidemiology and Control of Tuberculosis (2006) Beijing/W genotype Mycobacterium tuberculosis and drug resistance. Emerg Infect Dis 12:736–743

Faksri K, Drobniewski F, Nikolayevskyy V, Brown T, Prammananan T, Palittapongarnpim P, Prayoonwiwat N, Chaiprasert A (2011) Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on IS*6110*, SNP, LSP and VNTR profiles from Thailand. Infect Genet Evol 11:1142–1149

Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, Bobadilla del Valle M, Fyfe J, Garcia-Garcia L, Rastogi N, Sola C, Zozio T, Guerrero MI, Leon CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendon A, Sifuentes-Osornio J, Ponce de Leon A, Cave MD, Fleischmann R, Whittam TS, Alland D (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188:759–772

Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM (2013) *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat Genet 45:784–790

Gagneux S (2009) Fitness cost of drug resistance in *Mycobacterium tuberculosis*. Clin Microbiol Infect 15(Suppl 1):66–68

Gagneux S, Burgos MV, DeRiemer K, Encisco A, Munoz S, Hopewell PC, Small PM, Pym AS (2006a) Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. PLoS Pathog 2:e61

Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJ (2006b) The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science 312:1944–1946

Gandhi NR, Nunn P, Dheda K, Schaaf HS, Zignol M, van Soolingen D, Jensen P, Bayona J (2010) Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. Lancet 375:1830–1843

Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D (2002) Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg Infect Dis 8:843–849

Golesi F, Brignatz J, Bellenfant M, Raoult D, Drancourt M (2013) *Mycobacterium tuberculosis* Beijing outbreak in a school in Marseille, France, 2012. Euro Surveill 18:20354

Gumbo T (2010) New susceptibility breakpoints for first-line antituberculosis drugs based on antimicrobial pharmacokinetic/pharmacodynamic science and population pharmacokinetic variability. Antimicrob Agents Chemother 54:1484–1491

Ioerger TR, Feng Y, Chen X, Dobos KM, Victor TC, Streicher EM, Warren RM, Gey van Pittius NC, Van Helden PD, Sacchettini JC (2010) The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa. BMC Genom 11:670

Iwamoto T, Grandjean L, Arikawa K, Nakanishi N, Caviedes L, Coronel J, Sheen P, Wada T, Taype CA, Shaw MA, Moore DA,

Gilman RH (2012) Genetic diversity and transmission characteristics of Beijing family strains of *Mycobacterium tuberculosis* in Peru. PLoS One 7:e49651

Jiraphongsa C, Wangteeraprasert T, Henpraserttae N, Sanguanwongse N, Panya L, Sukkasitvanichkul J, Pittayawonganon C (2011) Community outbreak of multidrug resistance tuberculosis, Kanchanaburi province, Thailand on 2002–June 2010. J Preventive Med Assoc Thail 1:261–271

Johnson R, Jordaan AM, Pretorius L, Engelke E, van der Spuy G, Kewley C, Bosman M, van Helden PD, Warren R, Victor TC (2006) Ethambutol resistance testing by mutation detection. Int J Tuberc Lung Dis 10:68–73

Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 35:907–914

Kato-Maeda M, Ho C, Passarelli B, Banaei N, Grinsdale J, Flores L, Anderson J, Murray M, Rose G, Kawamura LM, Pourmand N, Tariq MA, Gagneux S, Hopewell PC (2013) Use of whole genome sequencing to determine the microevolution of *Mycobacterium tuberculosis* during an outbreak. PLoS One 8:e58235

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Larsen MH, Biermann K, Tandberg S, Hsu T, Jacobs WRJ (2007) Genetic Manipulation of *Mycobacterium tuberculosis*. Curr Protoc Microbiol Chapter 10:Unit 10A 12

Lee AS, Othman SN, Ho YM, Wong SY (2004) Novel mutations within the embB gene in ethambutol-susceptible clinical isolates of *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 48:4447–4449

Li XZ, Zhang L, Nikaido H (2004) Efflux pump-mediated intrinsic drug resistance in *Mycobacterium smegmatis*. Antimicrob Agents Chemother 48:2415–2423

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAM tools. Bioinformatics 25:2078–2079

Liu F, Hu Y, Wang Q, Li HM, Gao GF, Liu CH, Zhu B (2014) Comparative genomic analysis of *Mycobacterium tuberculosis* clinical isolates. BMC Genom 15:469

Luo T, Yang C, Peng Y, Lu L, Sun G, Wu J, Jin X, Hong J, Li F, Mei J, DeRiemer K, Gao Q (2014) Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. Tuberculosis 94:434–440

Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rusch-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F, Niemann S (2013) Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. PLoS One 8:e82551

Muller B, Borrell S, Rose G, Gagneux S (2013) The heterogeneous evolution of multidrug-resistant *Mycobacterium tuberculosis*. Trends Genet 29:160–169

Naidoo CC, Pillay M (2014) Increased in vitro fitness of multi- and extensively drug-resistant F15/LAM4/KZN strains of *Mycobacterium tuberculosis*. Clin Microbiol Infect 20:O361–O369

Nair J, Rouse DA, Bai GH, Morris SL (1993) The rpsL gene and streptomycin resistance in single and multiple drug-resistant strains of *Mycobacterium tuberculosis*. Mol Microbiol 10:521–527

Niemann S, Koser CU, Gagneux S, Plinke C, Homolka S, Bignell H, Carter RJ, Cheetham RK, Cox A, Gormley NA, Kokko-Gonzales P, Murray LJ, Rigatti R, Smith VP, Arends FP, Cox HS, Smith G, Archer JA (2009) Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. PLoS One 4:e7407

Niemann S, Diel R, Khechinashvili G, Gegia M, Mdivani N, Tang YW (2010) *Mycobacterium tuberculosis* Beijing lineage favors the spread of multidrug-resistant tuberculosis in the Republic of Georgia. J Clin Microbiol 48:3544–3550

Park YK, Ryoo SW, Lee SH, Jnawali HN, Kim CK, Kim HJ, Kim SJ (2012) Correlation of the phenotypic ethambutol susceptibility of *Mycobacterium tuberculosis* with embB gene mutations in Korea. J Med Microbiol 61:529–534

Parwati I, van Crevel R, van Soolingen D (2010) Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. Lancet Infect Dis 10:103–111

Perez-Lago L, Comas I, Navarro Y, Gonzalez-Candelas F, Herranz M, Bouza E, Garcia-de-Viedma D (2014) Whole genome sequencing analysis of intrapatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. J Infect Dis 209:98–108

Pym AS, Saint-Joanis B, Cole ST (2002) Effect of katG mutations on the virulence of *Mycobacterium tuberculosis* and the implication for transmission in humans. Infect Immun 70:4955–4960

Ramaswamy SV, Amin AG, Goksel S, Stager CE, Dou SJ, El Sahly H, Moghazeh SL, Kreiswirth BN, Musser JM (2000) Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 44:326–336

Roetzer A, Diel R, Kohl TA, Ruckert C, Nubel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rusch-Gerdes S, Supply P, Kalinowski J, Niemann S (2013) Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med 10:e1001387

Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, McNeil M, Peterson SN, Chatterjee D, Fleischmann R, Alland D (2013) Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-beta-D-arabinose biosynthetic and utilization pathway genes. Nat Genet 45:1190–1197

Sandegren L, Groenheit R, Koivula T, Ghebremichael S, Advani A, Castro E, Pennhag A, Hoffner S, Mazurek J, Pawlowski A, Kan B, Bruchfeld J, Melefors O, Kallenius G (2011) Genomic stability over 9 years of an isoniazid resistant *Mycobacterium tuberculosis* outbreak strain in Sweden. PLoS One 6:e16647

Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB (2009) Tuberculosis drug resistance mutation database. PLoS Med 6:e2

Schon T, Jureen P, Giske CG, Chryssanthou E, Sturegard E, Werngren J, Kahlmeter G, Hoffner SE, Angeby KA (2009) Evaluation of wild-type MIC distributions as a tool for determination of clinical breakpoints for *Mycobacterium tuberculosis*. J Antimicrob Chemother 64:786–793

Schurch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, Siezen RJ, van Soolingen D (2010) High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. J Clin Microbiol 48:3403–3406

Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM (1997) Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci 94:9869–9874

Srilohasin P (2013) Development of a DNA chip and determination of genetic diversity of *Mycobacterium tuberculosis* in Thailand. Ph. D. Thesis, Mahidol University

Strauss OJ, Warren RM, Jordaan A, Streicher EM, Hanekom M, Falmer AA, Albert H, Trollip A, Hoosain E, van Helden PD, Victor TC (2008) Spread of a low-fitness drug-resistant *Mycobacterium tuberculosis* strain in a setting of high

human immunodeficiency virus prevalence. J Clin Microbiol 46:1514–1516

Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE 3rd, Mei J, Gao Q (2012) Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. J Infect Dis 206:1724–1733

Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44:4498–4510

Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, Matter L, Schopfer K, Bodmer T (1993) Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. Lancet 341:647–650

Telenti A, Philipp WJ, Sreevatsan S, Bernasconi C, Stockbauer KE, Wieles B, Musser JM, Jacobs WR Jr (1997) The *emb* operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. Nat Med 3:567–570

The WHO (2014) Global tuberculosis report 2014. World Health Organization, Geneva

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192

Toungoussova OS, Sandven P, Mariandyshev AO, Nizovtseva NI, Bjune G, Caugant DA (2002) Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia. J Clin Microbiol 40:1930–1937

Toungoussova OS, Mariandyshev A, Bjune G, Sandven P, Caugant DA (2003) Molecular epidemiology and drug resistance of *Mycobacterium tuberculosis* isolates in the Archangel prison in Russia: predominance of the W-Beijing clone family. Clin Infect Dis 37:665–672

Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, Kreiswirth BN, Van Soolingen D, Small PM (2005) Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J Clin Microbiol 43:3185–3191

van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, Qing HZ, Enkhsaikan D, Nymadawa P, van Embden JD (1995) Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol 33:3234–3238

van Soolingen D, de Haas PE, van Doorn HR, Kuijper E, Rinder H, Borgdorff MW (2000) Mutations at amino acid position 315 of the *katG* gene are associated with high-level resistance to isoniazid, other drug resistance, and successful transmission of *Mycobacterium tuberculosis* in the Netherlands. J Infect Dis 182:1788–1790

Viratyosin W, Kulawonganunchai S, Smittipat N, Juthayothin T, Penpassakarn P, Pasomsub E, Chantratita W, Chaiprasert A, Palittapongarnpim P (2013) Draft genome sequence of the *Mycobacterium tuberculosis* strain 43-16836, belonging to the Indo-Oceanic lineage, isolated from tuberculous meningitis in Thailand. Genome Announc 1

Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE (2013) Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis 13:137–146

Werngren J, Hoffner SE (2003) Drug-susceptible *Mycobacterium tuberculosis* Beijing genotype does not develop mutation-conferred resistance to rifampin at an elevated rate. J Clin Microbiol 41:1520–1524

WHO Geneva/IUATLD Paris (1998) Guidelines for surveillance of drug resistance in tuberculosis. WHO Geneva/IUATLD Paris. International Union Against Tuberculosis and Lung Disease. Int J Tuberc Lung Dis 2:72–89

Zhang Y, Heym B, Allen B, Young D, Cole S (1992) The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature 358:591–593