











Genomic evidence supporting the clonal expansion of extensively drug-resistant tuberculosis bacteria belonging to a rare proto-Beijing genotype

Prapaporn Srilohasin ^{a,b}, Therdsak Prammananan ^{b,c}, Kiaticchai Faksri ^{d,e}, Jody E. Phelan ^f, Prapat Suriyaphol ^{g,h}, Phalin Kamolwatⁱ, Saijai Smithtikarnⁱ, Areeya Disratthakitⁱ, Sanjib Mani Regmi ^j, Manoon Leechawengwongs^{b,k}, Rick Twee-Hee Ong ^l, Yik Ying Teo^l, Sissades Tongsimma ^{c,m}, Taane G. Clark ^{f,n} and Angkana Chairprasert ^{a,b}

^aOffice for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; ^bDrug Resistant Tuberculosis Research Fund, Siriraj Foundation, Bangkok, Thailand; ^cNational Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathum Thani, Thailand; ^dDepartment of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand; ^eResearch and Diagnostic Center for Emerging Infectious Diseases (RCEID), Khon Kaen University, Khon Kaen, Thailand; ^fFaculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK; ^gDivision of Bioinformatics and Data Management for Research, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; ^hResearch Group and Research Network Division, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; ⁱBureau of Tuberculosis, Department of Disease Control, Ministry of Public Health, Bangkok, Thailand; ^jDepartment of Microbiology, Gandaki Medical College Teaching Hospital, Pokhara, Nepal; ^kVichaiyut Hospital, Bangkok, Thailand; ^lSaw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore; ^mNational Biobank of Thailand, National Science and Technology Development Agency, Pathum Thani, Thailand; ⁿFaculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK

ABSTRACT

Tuberculosis disease (TB), caused by *Mycobacterium tuberculosis*, is a major public health issue in Thailand. The high prevalence of modern Beijing (Lineage 2.2.1) strains has been associated with multi- and extensively drug-resistant infections (MDR-, XDR-TB), complicating disease control. The impact of rarer proto-Beijing (L2.1) strains is less clear. In our study of thirty-seven L2.1 clinical isolates spanning thirteen years, we found a high prevalence of XDR-TB cases (32.4%). With ≤ 12 pairwise SNP distances, 43.2% of L2.1 patients belong to MDR-TB or XDR-TB transmission clusters suggesting a high level of clonal expansion across four Thai provinces. All XDR-TB (100%) were likely due to transmission rather than inadequate treatment. We found a 47 mutation signature and a partial deletion of the *fadD14* gene in the circulating XDR-TB cluster, which can be used for surveillance of this rare and resilient *M. tuberculosis* strain-type that is causing increasing health burden. We also detected three novel deletion positions, a deletion of 1285 bp within *desA3* (Rv3230c), large deletions in the *plcB*, *plcA*, and *ppe38* gene which may play a role in the virulence, pathogenesis or evolution of the L2.1 strain-type.

ARTICLE HISTORY Received 2 August 2020; Revised 27 October 2020; Accepted 15 November 2020

KEYWORDS: Clonal; emerging; MDR; *Mycobacterium tuberculosis*; pre-XDR; proto-Beijing; SNP; whole-genome sequencing; XDR


Article summary line Clonal expansion of extensively drug-resistant tuberculosis caused by *Mycobacterium tuberculosis* proto-Beijing genotype in Thailand

Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is one of the top 10 causes of mortality globally. The global problem of drug-resistant TB, especially multi (MDR-TB) and extensive (XDR-TB) forms, is complicating disease control [1]. TB in Thailand remains high with an estimated 153 per 100,000 populations while around 50,167 cases were bacteriologically confirmed in 2018. An estimated 4000 MDR/RR-TB, including 1312 confirmed, were among notified pulmonary TB cases [1]. Treatment for patients with drug-resistant TB is prolonged, expensive and outcomes are poor. Drug

resistance in *M. tuberculosis* is almost exclusively due to mutations including single nucleotide polymorphisms (SNPs) in genes coding for drug-targets or -converting enzymes [2,3]. Whole-genome sequencing (WGS) platforms are being used to understand the mutations underlying drug resistance and to characterize transmission, where *M. tuberculosis* isolates (sourced from different hosts) that have near-identical genomic variation are most likely to have arisen from a recent transmission event [4]. Using WGS, molecular epidemiological studies can investigate the cause of emerging drug resistance and focus on the contribution of

CONTACT Angkana Chairprasert  angkana.cha@mahidol.ac.th

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/22221751.2020.1852891>.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

potential genetic and environmental risk factors across a population to plan for TB control.

The *M. tuberculosis* complex has 8 main lineages (L1-8) [5], which vary in their geographic distribution and spread, with L1 (Indo-Oceanic) and L2 (East Asian) the predominant ones circulating in Thailand [6–10]. The modern Beijing sublineage (L2.2.1; AAF3) has been continuously reported to be associated with clonal expansion of MDR-TB [6,7,11]. In contrast, the “Manu-ancestor” or “proto-Beijing” genotype (spoligotype 523, lineage 2.1; [9]) has a lower prevalence [6,10], with few reports, including detection in the Chiba prefecture in Japan [12] and Guangxi province of Southern China [13]. Recently, we noticed a potential transmission event involving proto-Beijing strains causing XDR-TB [7]. To understand if there is further transmission of the L2.1 strain and any related XDR-TB clonal expansion, we have analysed the WGS from thirty-seven isolates collected over 13-year period from 14 provinces of Thailand.

Methods

L2.1 samples and sequencing

Clinical isolates sourced from 725 TB patients across 42 of 77 provinces from Thailand were classified as MDR-TB, pre-XDR, XDR-TB and pan-susceptible by applying phenotypic drug susceptibility testing (DST) assays using drug-containing Middlebrook 7H10 agar (M7H10). The 723 culturable isolates were from the retrievable stock cultures of 1415 MDR-, pre-XDR and XDR-TB from the Drug-Resistant Tuberculosis Research Fund Laboratory. These isolates were obtained from pulmonary TB patients diagnosed between 2005 and 2012 and covering 42 of 76 provinces in Thailand. All isolates were sub-cultured on Loewenstein-Jensen media and incubated at 37°C for four weeks. Multiple loopfulls of *M. tuberculosis* colonies were taken and DNA extraction was carried out using the cetyltrimethylammonium bromide-sodium chloride method [14]. Genomic DNAs were sequenced using Illumina platforms, across a number of sites, including Novogene AIT (Singapore), Genomic Institute Singapore (GIS), London School of Hygiene & Tropical Medicine, and Mahidol University. We applied the TBProfiler tool [15] to the WGS data to *in-silico* strain-type, and identified 35 (4.8%) isolates as L2.1 strains. To determine the genetic similarity between currently and recently circulating L2.1 causing XDR-TB, two recent L2.1 XDR-TB isolates (2015 and 2017) from Kanchanaburi province were blended into this study. Here L2.1 collection ($n = 37$) represents the largest number of L2.1 strains from Thailand. The raw sequencing data for the 37 samples are deposited in NCBI's Sequence Read Archive (SRA) database

(BioProject ID PRJNA63078). They were analysed together with other 19 L2.1 *M. tuberculosis* strains obtained from public data; 16 isolates from high prevalence area (China), one isolate from the neighbouring country (Laos), and two well-characterized L2.1 isolates from the Sanger Institute [16–19]. The accession numbers of all 56 genomes are listed in Table S1. The 37 L2.1 *M. tuberculosis* were isolated from Thai-patients, across 19 districts among 14 provinces, between 2005 and 2017 by the Drug-Resistant Tuberculosis Research Fund, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok (patient no.1–19 and 22–37) and the National Reference Tuberculosis Laboratory (patient no. 20 and 21). Epidemiological linkage between patients was not documented.

Ethics statement

Approval for the study was given by the Ethical and Scientific Committees of the Faculty of Medicine Siriraj Hospital, Mahidol University (EC No. 510/2561) and the Center for Ethics in Human Research, Khon Kaen University (HE601249).

Bioinformatics analysis

The L2.1 isolates ($n = 56$) consisted of 37 strains from our collection and 19 globally distributed. The quality of their sequence reads was checked using FastQC (v0.11.7) [20,21]. Sequencing reads were mapped to the *M. tuberculosis* H37Rv reference genome (NC_000962.3) using BWA-MEM software [22]. The sets of mapped paired-end reads were joined using the SAMtools suite. Variants were called using GATK software (v3.8) with HaplotypeCaller in -ERC GVCF mode for generating gVCF files. Multiple gVCF files were aggregated and combined into one GVCF file using GATK GenotypeGVCFs. Only the single nucleotide polymorphisms (SNPs) from VCF output were extracted using SelectVariants tool and filtered using VariantFiltration [23]. The snpEff software (v4.2) was used for variant annotation [24]. SNPs determined in the repetitive regions, paralogous gene families, and drug-resistant genes were discarded using VCFtools [25]. SNPs with allelic frequencies less than 75% or read-depth less than 10 reads were removed. The resulting set of high-confidence SNPs were used to construct a phylogenetic tree based on the maximum likelihood method and 1000 bootstrap replicates using RAxML (v8.0.0) software [26]. iTOL software was used for visualizing the phylogenetic trees [27]. Drug-resistance profiles (and lineages) were predicted *in-silico* using TBProfiler (v2.0) [15]. Spoligotypes were predicted using SpoTyping [28]. The *in-silico* analysis for 31 known regions of differences (RDs) was performed by RD-Analyzer [18].

DELLY software [29] was used to predict large structural variants (SVs) using minimum paired-end mapping quality 20 and standard deviation units as 3 times. High-quality SVs had to be supported by paired-end and split reads. All variants across the samples were merged into a single file using BCFtools software. All putative genetic markers were manually reviewed for accuracy using Integrated Genomic Viewer software [30], and, if considered as true variant they were included for delineation.

Results

Setting and expansion of *M. tuberculosis* L2.1 in Thailand

The 37 clinical *M. tuberculosis* strains used were isolated from pulmonary TB patients in 14 provinces from 2005 to 2017. In 2005, the L2.1 sub-lineage was found only in Kanchanaburi and Bangkok. Seven years later, L2.1 were found across 14 provinces and 5 were XDR-TB (Figure 1).

The overall rate of MDR-TB individuals with pre-XDR and XDR-TB are high. Twelve isolates (32.4%) are XDR-TB whereas 7 isolates (18.9%) are pre-XDR. Sixteen isolates are MDR-TB (43.2%) and the remaining 2 isolates are pan-susceptible. Demographic data of each patient is shown in (Table 1 and Table S1). The L2.1 incident rate ($n = 37$ or 5.1%) was high compared with studies in the Chiangrai province, Northern Thailand (1.0%) [10] and Guangxi province of Southern China (0.2%) [13]. Unfortunately, our collection had no isolates from Chiangrai. Within province the highest number of L2.1 strain was observed in Kanchanaburi ($n = 13$), followed by Nakhon Ratchasima ($n = 5$), Buriram and Bangkok ($n = 3$) province.

Genotyping analysis of the rare genotype L2.1 isolates in Thailand

The genotypic drug-resistance profile of each isolate is shown in Figure 2. *In silico* spoligotyping classified the 37 L2.1 strains into 12 genotypes, and the most common spoligotype (23/37; 62.2%) was SIT 523 (7777777777777771). Of these SIT 523 isolates, 12 (52.2%) were XDR-TB, and were sourced from 4 provinces; Chachoengsao ($n = 1$), Kanchanaburi ($n = 9$), Ratchaburi ($n = 1$), Suphanburi ($n = 1$) (Figures 1 and 2). All isolates harboured the extended RD105 deletion but had RD207, RD181 and *pks15/1* regions intact.

To investigate the within and between country spread of L2.1, an analysis of the genome variation of all 56 isolates (Thailand, China, Laos, and others) was performed. A total of 3685 high-confidence SNPs were identified across 56 *M. tuberculosis* isolates,

including 3206 (87.0%) in coding regions (1141 silent, 2028 missense, and 37 nonsense), 12 in non-coding and 467 in intergenic regions (Table S2). Across the 56 isolates, 2484 (67.4%) SNPs were identified in single isolates. A phylogenetic tree based on the 3685 SNPs identified revealed that Thai-isolates were mostly interleaved with other countries including China (CHA), Laos (LAO), and others (NA) (Figure 2). All 12 (100%) XDR-TB isolates were clustered together and were distinguished from the others by at least 47 SNPs. Twenty-seven of 47 SNPs had missense mutations. This large and long-lasting monophyletic clade with high bootstrap support (100%) included XDR-TB cases from Kanchanaburi ($n = 9$). The clade was also interleaved with XDR-TB isolates from Ratchaburi ($n = 1$), Chachoengsao ($n = 1$), and Supanburi ($n = 1$) (Figure 3(A)). The time range between the first and last isolate was 13 years (2005–2017). Moreover, the results revealed that 2 (28.6%) pre-XDR and 2 (12.5%) MDR-TB patients identified were limited to one geographic district and separated by at least 90 SNPs and 60 SNPs, respectively, from any other study isolates. This result implies that this particular strain was circulating locally. A genotypic clustering rate of 43.2% was found among patients who were infected with L2.1 strain, revealing the high level of transmission in this population.

Evidence of transmission

Potential transmission clusters were revealed by determining the pairwise SNP distance between the 37 Thai-isolates. The pairwise genetic distance varied between 0 and 381 SNPs (Table S3). Using a 12 SNP cut-off [31], 43.2% of patients belonged to 3 different possible clusters, including: (i) an XDR-TB cluster with a maximum size of 12 isolates (Figure 3(B)), (ii) 2 pre-XDR isolates, and (iii) 2 MDR-TB isolates. Mutation patterns associated with resistance were also unique for each cluster (Table S4), providing potential evidence of transmission. The MDR-TB cluster ($n = 2$; no.14 (2008) and 16 (2009)) was composed of patients from the same district and had identical SNP patterns, including mutations known to be responsible for drug resistance (*rpoB* Ser450Leu; *katG* Ser315Thr). Within the pre-XDR cluster ($n = 2$; no.4 (2008) and 6 (2012)), there was only 1 SNP difference (a synonymous mutation difference between isolates 4 and 6), and the established drug-resistance mutations were identical (*katG* Ser315Thr; *rpoB* Ser450Leu; *rpsL* Lys88Arg; *embB* Gly406Asp; *pncA* Tyr103His; *gyrA*: Asp94Asn; *thyA* deletion 3071595-3077080). For the XDR-TB cluster ($n = 12$), the pairwise genetic distance varied between 1 and 11 SNPs. Nine (75.0%) XDR-TB were isolated from Kanchanaburi province, including Thamaka ($n = 5$; no.9 (2005), 12 (2007), 15 (2008), 20 (2015), and 21

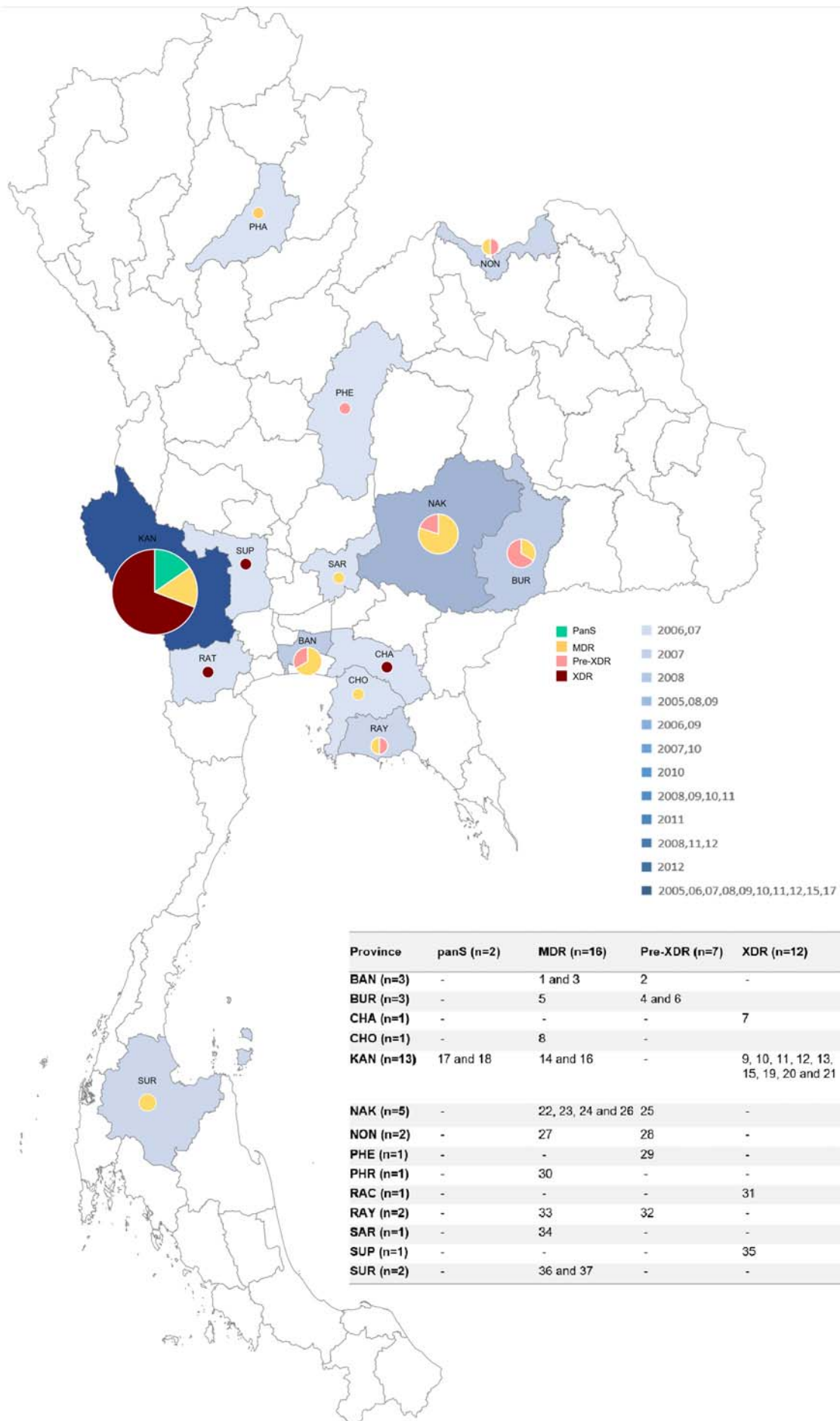


Figure 1. Geographic distribution of pan-susceptible (green), MDR-TB (yellow), pre-XDR (pink), and XDR-TB (dark red) caused by *M. tuberculosis* L2.1. Each province has been shaded according to the frequency of sample collection including Bangkok (BAN), Buriram (BUR), Chachoengsao (CHA), Chonburi (CHO), Kanchanaburi (KAN), Nakhonratchasima (NAK), Nongkhai (NON), Phetchabun (PHE), Phrae (PHR), Ratchaburi (RAT), Rayong (RAY), Saraburi (SAR), Suphanburi (SUP), and Suratthani (SUR).

Table 1. SVs identified among the L2.1 sub-population. The position is relative to the *M. tuberculosis* H37Rv reference genome (NC_000962.3).

Uniqueness	SV types	Start	Stop	Genes	Gene symbols and descriptions
MDR-TB cluster	Deletion	2900282	2900432	Rv2576c	membrane protein
	Deletion	2900903	2901086	Rv2577	hypothetical protein_metallophosphatase superfamily
Pre-XDR cluster	Deletion	1013430	1015391	Rv0909	Antitoxin
				Rv0910	toxin
	Deletion	2851350	2851718	Rv2527	<i>vapC17</i> : ribonuclease
	Deletion	3071595	3078945	Rv2761c	<i>hds</i> : type I restriction/modification system
				Rv2762c	specificity determinant
				Rv2763c	hypothetical protein
				Rv2764c	<i>dfrA</i> : dihydrofolate reductase
				Rv2765	<i>thyA</i> : thymidylate synthase
				Rv2766c	hydrolase
				Rv2767c	<i>fabG5</i> : short-chain type
				Rv2768c	membrane protein
				Rv2769c	<i>PPE43</i> : PPE family protein PPE43
					<i>pe27</i> : PE family protein PE27
					<i>fadD14</i> : fatty-acid-CoA ligase
XDR-TB cluster	Deletion	1180695	1180773	Rv1058	
All isolates	Deletion	2010609	2011436	Rv1777	<i>cyp144</i> : cytochrome P450
	Deletion	3606557	3607841	Rv3229c	<i>desA3</i> : stearyl-CoA 9-desaturase
				Rv3230c	stearyl-CoA 9-desaturase electron transfer protein
				Rv2350c	<i>plcB</i> : membrane-associated phospholipase B
				Rv2351c	<i>plcA</i> : membrane-associated phospholipase A
		Deletion	2629526	2634350	Rv2352c

(2017)), Muang ($n = 2$; no.11 (2007) and 13 (2008)), Bophloi ($n = 1$; no.10 (2006)) and Thamuang ($n = 1$; no.19 (2012)) districts. The other three isolates collected from patient in Chachoengsao (no.7 in 2008), Rachaburi (no.31 in 2012), and Suphaburi (no.35 in 2011). Ten (83.3%) XDR isolates had identical resistance mutation profiles (*katG* Ser315Thr; *rpoB* Ser450-Leu; *embB* Met306Ile; *pncA* Ile90Ser; *rrs*: 1401a > g; *gyrA* Asp94Gly; *folC* Glu40Gly). XDR-TB samples from Chachoengsao (no.7), and Suphanburi (no.35) were located 174 and 64.7 kilometres, respectively, away from earlier isolates. They had identical drug-resistant mutations to the majority of the XDR-TB cluster, except for the addition of the Leu527Val mutation in the (rifampicin compensatory) *rpoC* gene. The recent isolate (no. 20; year 2015) had an additional mutation (*thyX* -16C > T) related to para-aminosalicylic acid resistance. These results suggest that there was a long-term transmission of local clusters of XDR-TB, which carried common drug-resistance mutations, thereby explaining its success as a spreader [32–34].

During the study period, the first XDR-TB case (patient no. 9) was infected with an L2.1 XDR-TB clone from Kanchanaburi in 2005. Subsequent events have revealed that the XDR-TB clone has emerged continuously, of which 12 had at least one isolate over a span of 13 years, and been restricted to within 180 km across four provinces from the first case.

Structural diversity in clonal cluster population

Structure variations (SVs) were called relative to the *M. tuberculosis* H37Rv reference genome. Thirty-nine high-quality SVs were detected. Three common deletion regions across all combined dataset of L2.1 isolates were detected with sizes between 827 and

4824 bp (Table 1), all exclusive to the sub-lineage. The largest genome SV involved a deletion of *plcA* (Rv2351c) and *PPE38* (Rv2352c), and a partial deletion of *plcB* (Rv2350c). There were deletions unique to the clusters (MDR-TB 2; pre-XDR 2). Twelve isolates belonging to the XDR-TB cluster contained one unique deletion in the Rv1058 region, which was not found in any other isolates.

Discussion

The recent emergence of drug-resistant TB infections in Thailand has been attributed to strains in *M. tuberculosis* lineage 2, particularly the “modern Beijing” (Lineage 2.2.1) sub-lineage. *M. tuberculosis* from this sub-lineage are highly virulent causing disease outbreaks, with escape from the effects of BCG vaccination, they disseminate efficiently, and easily acquire antibiotic resistance [6,7,9,35]. The more rare “proto-Beijing” (L2.1) appears to be transmitting, and has a high propensity to be XDR-TB. In our study, the number of MDR-TB individuals with XDR-TB is ~6 times greater than the WHO global estimate (6.2%) [1], and all are clustered. Poor treatment adherence is not the only factor that contributes to drug resistance, but also the failure in controlling transmission of drug-resistant strains.

This study was a hospital-based retrospective analysis included the retrospective and passive nature of case-finding, based only on retrievable isolates from stock cultures during 2005–2012; therefore, they were a set of convenience isolates and may not be representative of all isolates from the community, and we cannot exclude some unintended bias. Additionally, TB database information of Thailand during 2005–2012 was not well developed, and the capacity of the laboratory to perform phenotypic

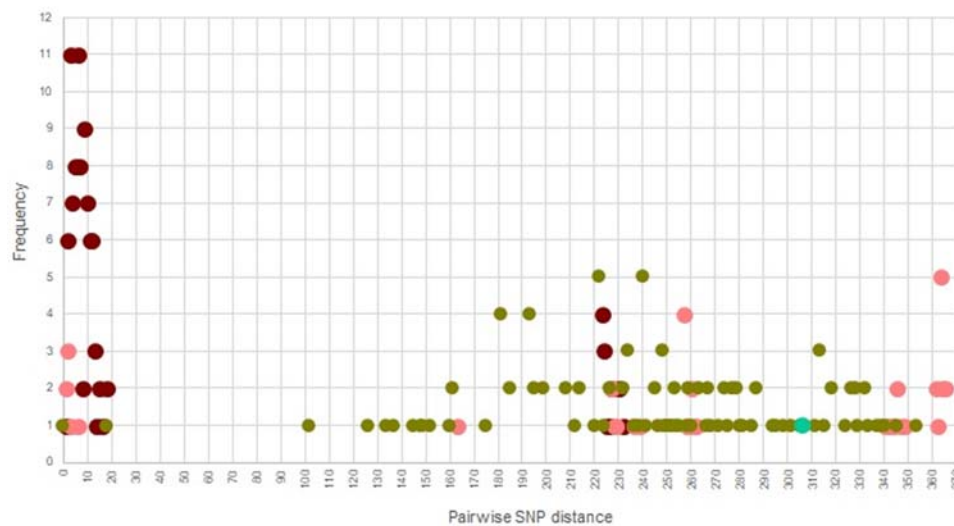


Figure 3 Continued

drug susceptibility testing was also limited. However, the number of L2.1 isolates retrieved is the largest to date from a single country, and the underlying sampling from 725 culturable *M. tuberculosis* covers 42 of 77 provinces [2,6,7,17]. Here we reported the presence of L2.1 isolates, including in Bangkok, Buriram, Chachoengsao, Chonburi, Kancharaburi, Nakhonratchasima, Nongkhai, Phetchabun, Phrae, Ratchaburi, Rayong, Saraburi, Suphanburi, and Suratthani, but not Chiangrai provinces [7,10].

WGS data were integrated with routinely clinical information to identify the putative TB cluster. Unfortunately, epidemiological information was not available in this study. The integration of social-network analysis with high-resolution bacterial genome sequencing would provide traceable information in evaluating TB transmission [36]. The 12-SNP distance was proposed by using studies in low incidence TB settings [31], and seems reasonable for our study, set in the contexts of low L2.1 population prevalence and the community transmission investigation. Using clustering threshold of 12 SNPs, we defined 3 distinct unique clonal clusters which carried particular drug-resistant mutation pattern for each clone (Table S4), implying diverse evolution histories for the individual population. A high clustering rate of 43.2% among L2.1 infected patients highlighted the high transmissibility of proto-Beijing strains. Two of the three clusters identified had one pair of strains and emerged in a particular region. Among the largest possible cluster groupings consisting of twelve XDR-TB transmission clusters, at least seven putative transmission events were postulated to underlie the largest cluster on the basis of a shared social setting and WGS genotyping, in which each individual isolate was supposed to be accumulated genetic diversity within-host [31]. Nine isolates were retrieved from patients in Kancharaburi province, of which four

district hospitals were involved in four putative transmission events; 5 (41.7%) isolates from Thamaaka, 2 isolates from Muang, 1 isolate from Thamuang, and 1 isolate from Bo phloi district. The other 3 events were possibly infected through social contact in Chachoengsao (patient no.7), Rachaburi (patient no.31), and Supanburi (patient no.35). However, the potential of clonal expansion reported here would switch attention towards a rare genotype causing XDR-TB in Kancharaburi province, not only in a particular hospital but through the high prevalence of MDR-TB, especially within L2.2.1; AAF3 [7,11,36,37]. The lack of transfer of these hyper-resistant strains to other geographical locations may be associated with adaptation to local-host population, as seen in L1 strains [9] or other factors which need to be explored in the future.

Forty-seven SNPs and a 78 bp deletion in *fadD14* (encoded fatty-acid-CoA ligase) were present in all XDR-TB clonal spreaders, but not other isolates studied. Notably, 27 (of the 47) SNPs resulted in a missense mutation and were potentially subject to natural selection. The genetic markers shared by the successful XDR-TB cluster could be related to virulence or used to screen for the existence of antibiotic-resistant bacteria. Specific resistance-conferring mutations and strain genetic background are often associated with a fitness cost [32,33]. The local clusters of XDR-TB carried common drug-resistance mutations, potentially explaining its successful spreading phenotype more readily than other genotypes. Unavoidably, host genetic background also influences the response to mycobacterial activity [6,17]. The successive appearance of XDR-TB cluster conferred overgrowth their ancestral drug-susceptible clone that become undetectable in this study. Mutations found within the XDR-TB cluster indicate a remarkable adaptation capacity during infection. Further, the factors influencing

transmission remain poorly understood, but as WGS becomes routine in a clinical setting, it will be possible to reconstruct the transmission chains and assess the host and *M. tuberculosis* genetic and non-genetic factors affecting transmissibility.

We detected three novel deletion positions which may play a role in the virulence, pathogenesis or evolution of the L2.1 strain-type. Poorer bacterial growth has been observed in presence of azole in *M. tuberculosis* strains with *CYP144* knockout [38]. Azole drugs may have activity against the L2.1 strain. Further, a deletion of 1285 bp within *desA3* and oxidoreductase Rv3230c may cause the L2.1 strain to be resistant to the second-line anti-TB drug isoxyl [39]. Large deletions in the *plcB*, *plcA*, and *ppe38* gene were also detected. The deletion of *ppe38* and loss of ESX-5 substrates secretion as found in modern Beijing sub-lineages may have increased their virulence and contributed to their global spread [40]. Whereas partial deletion of phospholipase C encoded by *plc* genes (*plcA*, *B*, *C* and *D*) may have a role in assisting the bacteria to escape from phagosomal containment, as found in *Listeria monocytogenes* or *Clostridium perfringens* [41]. This may cause lower virulence than a modern Beijing sub-lineage.

Overall, our study has revealed that *M. tuberculosis* sub-lineage L2.1 is geographically restricted but has a great propensity to be XDR-TB and transmit. We anticipate that revealing the XDR-TB burden of sub-lineage L2.1 will lead to more widespread surveillance and much-needed epidemiological studies. The strong recommendation from the Thailand national TB control program to implement N95 or equivalent mask wearing for active TB patients especially MDR-, pre-XDR and XDR-TB patients will be an additional measure to prevent TB transmission during intensive treatment.

Acknowledgments

This study is funded by the Newton UK-Thailand Research and Innovation Partnership Fund, grant number P-18-50228. PS is supported by the Postdoctoral fellowship award from Mahidol University. KF is supported by the National Research Council of Thailand and Health System Research Institute (HSRI) grant number 60-057 and 62-003. TGC is funded by the Medical Research Council UK (grant number MR/M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1) and BBSRC UK (grant number BB/R013063/1). RTHO and YYT are supported by National University of Singapore Yong Loo Lin School of Medicine Aspiration Fund (NUHSRO/2014/069/AF-New Idea/04). The work is dedicated to late HRH Princess Galyani Vadhana Krom Luang Naradhiwas Rajanagarindra, the patronage of Drug-Resistant Tuberculosis Fund, Siriraj Foundation, Mahidol University. The authors of this paper are grateful to the individuals who participated in this study and the local hospital staff for comprehensive

strengthening the quality of TB care consecutively. The authors would also like to acknowledge the strong collaboration and support from Bureau of Tuberculosis, Department of Disease Control, Ministry of Public Health, Thailand, without which the study would not have been as successful.

Disclosure statement

No potential conflict of interest was reported by the author (s).

Funding

This study is funded by the Newton UK-Thailand Research and Innovation Partnership Fund [grant number P-18-50228]. P. S. is supported by the Postdoctoral fellowship award from Mahidol University. K. F. is supported by the National Research Council of Thailand and Health System Research Institute (HSRI) [grant number 60-057 and 62-003]. T. G. C. is funded by the Medical Research Council UK [grant numbers MR/M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1] and Biotechnology and Biological Sciences Research Council (BBSRC UK) [grant number BB/R013063/1]. R. T. H. O. and Y. Y. T. are supported by National University of Singapore Yong Loo Lin School of Medicine Aspiration Fund [NUHSRO/2014/069/AF-New Idea/04].

ORCID

Prapaporn Srilohasin  <http://orcid.org/0000-0003-4882-516X>

Therdsak Prammananan  <http://orcid.org/0000-0002-5132-1440>

Kiatichai Faksri  <http://orcid.org/0000-0001-5022-4182>

Jody E. Phelan  <http://orcid.org/0000-0001-8323-7019>

Prapat Suriyaphol  <http://orcid.org/0000-0001-9709-4389>

Sanjib Mani Regmi  <http://orcid.org/0000-0003-0589-3605>

Rick Twee-Hee Ong  <http://orcid.org/0000-0002-0197-7820>

Sissades Tongshima  <http://orcid.org/0000-0002-1491-1839>

Taane G. Clark  <http://orcid.org/0000-0001-8985-9265>

Angkana Chairprasert  <http://orcid.org/0000-0002-6333-947X>

References

- [1] World Health Organization. Global tuberculosis report 2019. Report No.: 978-92-4-156571-4.
- [2] Silva MS, Senna SG, Ribeiro MO, et al. Mutations in *katG*, *inhA*, and *ahpC* genes of Brazilian isoniazid-resistant isolates of *Mycobacterium tuberculosis*. *J Clin Microbiol.* 2003;41(9):4471–4474.
- [3] Jonmalung J, Prammananan T, Leechawengwongs M, et al. Surveillance of pyrazinamide susceptibility among multidrug-resistant *Mycobacterium tuberculosis* isolates from Siriraj Hospital, Thailand. *BMC Microbiol.* 2010;10:223.
- [4] Faksri K, Kaewprasert O, Ong RT, et al. Comparisons of whole-genome sequencing and phenotypic drug susceptibility testing for *Mycobacterium tuberculosis*

- causing MDR-TB and XDR-TB in Thailand. *Int J Antimicrob Agents*. 2019;54(2):109–116.
- [5] Ngabonziza JCS, Loiseau C, Marceau M, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nat Commun*. 2020;11(1):2917.
- [6] Srilohasin P, Chairprasert A, Tokunaga K, et al. Genetic diversity and dynamic distribution of *Mycobacterium tuberculosis* isolates causing pulmonary and extrapulmonary tuberculosis in Thailand. *J Clin Microbiol*. 2014;52(12):4267–4274.
- [7] Disratthakit A, Meada S, Prammananan T, et al. Genotypic diversity of multidrug-, quinolone- and extensively drug-resistant *Mycobacterium tuberculosis* isolates in Thailand. *Infect Genet Evol*. 2015;32:432–439.
- [8] Faksri K, Drobniewski F, Nikolayevskyy V, et al. Epidemiological trends and clinical comparisons of *Mycobacterium tuberculosis* lineages in Thai TB meningitis. *Tuberculosis (Edinb)*. 2011;91(6):594–600.
- [9] Palittapongarnpim P, Ajawatanawong P, Viratyosin W, et al. Evidence for host-bacterial co-evolution via genome sequence analysis of 480 Thai *Mycobacterium tuberculosis* lineage 1 isolates. *Sci Rep*. 2018;8(1):11597.
- [10] Ajawatanawong P, Yanai H, Smittipat N, et al. A novel ancestral Beijing sublineage of *Mycobacterium tuberculosis* suggests the transition site to modern Beijing sublineages. *Sci Rep*. 2019;9(1):13718.
- [11] Regmi SM, Chairprasert A, Kulawonganchai S, et al. Whole genome sequence analysis of multidrug-resistant *Mycobacterium tuberculosis* Beijing isolates from an outbreak in Thailand. *Mol Genet Genomics*. 2015;290(5):1933–1941.
- [12] Yokoyama E, Hachisu Y, Hashimoto R, et al. Concordance of variable-number tandem repeat (VNTR) and large sequence polymorphism (LSP) analyses of *Mycobacterium tuberculosis* strains. *Infect Genet Evol*. 2010;10(7):913–918.
- [13] Liu Q, Ma A, Wei L, et al. China's tuberculosis epidemic stems from historical expansion of four strains of *Mycobacterium tuberculosis*. *Nat Ecol Evol*. 2018;2(12):1982–1992.
- [14] Larsen MH, Biermann K, Tandberg S, et al. Genetic Manipulation of *Mycobacterium tuberculosis*. *Curr Protoc Microbiol*. 2007;Chapter 10:Unit 10A.2.
- [15] 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.
- [16] Perdigao J, Silva C, Maltez F, et al. Emergence of multidrug-resistant *Mycobacterium tuberculosis* of the Beijing lineage in Portugal and Guinea-Bissau: a snapshot of moving clones by whole-genome sequencing. *Emerg Microbes Infect*. 2020;9(1):1342–1353.
- [17] Luo T, Comas I, Luo D, et al. Southern East Asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with Han Chinese. *Proc Natl Acad Sci U S A*. 2015;112(26):8136–8141.
- [18] Faksri K, Xia E, Tan JH, et al. In silico region of difference (RD) analysis of *Mycobacterium tuberculosis* complex from sequence reads using RD-Analyzer. *BMC Genomics*. 2016;17(1):847.
- [19] Coll F, McNerney R, Guerra-Assuncao JA, et al. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun*. 2014;5:4812.
- [20] Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- [21] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
- [22] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–595.
- [23] Disratthakit A, Toyo-Oka L, Thawong P, et al. An optimized genomic VCF workflow for precise identification of *Mycobacterium tuberculosis* cluster from cross-platform whole genome sequencing data. *Infect Genet Evol*. 2019;79:104152.
- [24] Cingolani P, Platts A, Wang le L, et al. 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.
- [25] Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–2158.
- [26] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–1313.
- [27] Letunic I, Bork P. Interactive tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47(W1):W256–W2W9.
- [28] Xia E, Teo YY, Ong RT. Spotyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. *Genome Med*. 2016;8(1):19.
- [29] Rausch T, Zichner T, Schlattl A, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28(18):i333–i3i9.
- [30] Thorvaldsdóttir H, Robinson J, Mesirov J. Integrative Genomics Viewer (Igv): high-Performance Genomics data Visualization and Exploration. *Brief Bioinformatics*. 2012;14.
- [31] Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis*. 2013;13(2):137–146.
- [32] Nguyen QH, Contamin L, Nguyen TVA, et al. Insights into the processes that drive the evolution of drug resistance in *Mycobacterium tuberculosis*. *Evol Appl*. 2018;11(9):1498–1511.
- [33] Gagneux S, Long CD, Small PM, et al. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science*. 2006;312(5782):1944–1946.
- [34] Shitikov E, Kolchenko S, Mokrousov I, et al. Evolutionary pathway analysis and unified classification of East Asian lineage of *Mycobacterium tuberculosis*. *Sci Rep*. 2017;7(1):9227.
- [35] Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis*. 2010;10(2):103–111.
- [36] Boonthanapat N, Soontornmon K, Pungrassami P, et al. Use of network analysis multidrug-resistant tuberculosis contact investigation in Kanchanaburi, Thailand. *Trop Med Int Health*. 2019;24(3):320–327.
- [37] Coker OO, Regmi SM, Suriyaphol P, et al. Whole-Genome sequence of a multidrug-resistant *Mycobacterium tuberculosis* Beijing sequence type 10

- isolate from an outbreak in Thailand. *Genome Announc.* **2014**;2(4).
- [38] Driscoll MD, McLean KJ, Cheesman MR, et al. Expression and characterization of *Mycobacterium tuberculosis* CYP144: common themes and lessons learned in the M. tuberculosis P450 enzyme family. *Biochim Biophys Acta.* **2011**;1814(1):76–87.
- [39] Phetsuksiri B, Jackson M, Scherman H, et al. Unique mechanism of action of the thiourea drug isoxyl on *Mycobacterium tuberculosis*. *J Biol Chem.* **2003**;278(52):53123–53130.
- [40] Ates LS, Dippenaar A, Ummels R, et al. Mutations in ppe38 block PE_PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nat Microbiol.* **2018**;3(2):181–188.
- [41] Le Chevalier F, Cascioferro A, Frigui W, et al. Revisiting the role of phospholipases C in virulence and the lifecycle of *Mycobacterium tuberculosis*. *Sci Rep.* **2015**;5:16918.