MINIREVIEW

Importance of differential identification of *Mycobacterium tuberculosis* strains for understanding differences in their prevalence, treatment efficacy, and vaccine development

Hansong Chae and Sung Jae Shin[*]

Department of Microbiology, Institute for Immunology and Immunological Diseases, Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul 03722, Republic of Korea

(Received Jan 25, 2018 / Revised Feb 14, 2018 / Accepted Feb 14, 2018)

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (Mtb), remains a serious global health problem in the 21st century because of its high mortality. Mtb is an extremely successful human-adapted pathogen that displays a multifactorial ability to control the host immune response and to evade killing by drugs, resulting in the breakdown of BCG vaccine-conferred anti-TB immunity and development of multidrug-resistant (MDR) and extensively drug-resistant (XDR) Mtb. Although genetic components of the genomes of the Mtb complex strains are highly conserved, showing over 99% similarity to other bacterial genera, recently accumulated evidence suggests that the genetic diversity of the Mtb complex strains has implications for treatment outcomes, development of MDR/XDR Mtb, BCG vaccine efficacy, transmissibility, and epidemiological outbreaks. Thus, new insights into the pathophysiological features of the Mtb complex strains are required for development of novel vaccines and for control of MDR/XDR Mtb infection, eventually leading to refinement of treatment regimens and the health care system. Many studies have focused on the differential identification of Mtb complex strains belonging to different lineages because of differences in their virulence and geographical dominance. In this review, we discuss the impact of differing genetic characteristics among Mtb complex strains on vaccine efficacy, treatment outcome, development of MDR/XDR Mtb strains, and epidemiological outbreaks by focusing on the best-adapted human Mtb lineages. We further explore the rationale for differential identification of Mtb strains for more effective control of TB in clinical and laboratory settings by scrutinizing current diagnostic methods.

Keywords: *Mycobacterium tuberculosis*, *Mycobacterium tu-*

*For correspondence. E-mail: sjshin@yuhs.ac; Tel.: +82-2-2228-1813; Fax: +82-2-392-7088

berculosis lineage, genetic diversity, diagnosis, genotyping

Introduction

Tuberculosis (TB) is a distressing infectious disease among humans and remains one of the most important public health problems worldwide. The World Health Organization reports that 1–3 people die every minute from TB and the risk of multi-drug resistance (MDR) Mtb infection has increased over time (World Health Organization, 2016). The causative agents of TB in humans are obligate pathogenic bacteria belonging to the *Mycobacterium tuberculosis* complex (hereafter referred to as the Mtb complex), especially *Mycobacterium tuberculosis* (hereafter referred to as Mtb, the discovery of which was reported in 1882 by Robert Koch) (Koch, 1882; World Health Organization, 2016), *M. africanum* and *M. bovis*. Infection with a strain of the Mtb complex can result in variable outcomes including clearance via innate immunity, latent TB, or active extra-pulmonary or pulmonary TB (Coscolla and Gagneux, 2010). The variable disease outcomes of TB have been attributed more to variation in host immune responses and environmental factors than to genetic variation among Mtb strains, because of their low genetic diversity and the absence of conspicuous virulence factors in comparison to other bacterial genera (Coscolla and Gagneux, 2010; Gagneux, 2013). Nevertheless, particular strains belonging to specific Mtb complex species and Mtb lineages display pronounced differences in their virulence phenotypes and pathogenesis, including the emergence of drug resistance, breakdown of vaccine efficacy, and transmissibility among different ranges of hosts (Coscolla and Gagneux, 2014). Due to increasing emphasis on the importance of differential identification of the Mtb complex strains, several molecular genotyping methods have recently been introduced in clinical and experimental settings. They have led to a much better understanding of the global phylogenetic diversity of the Mtb complex and the potential factors influencing the different outcomes of TB (Coscolla and Gagneux, 2010, 2014; Couvin and Rastogi, 2015).

Different species and strains of the Mtb complex exhibit distinct responses in experimental studies and have been reported to be able to affect clinical presentation (Coscolla and Gagneux, 2010). Some clinical isolates have been reported to display differences in their virulence depending on their

lineage (Coscolla and Gagneux, 2010; Couvin and Rastogi, 2015). Especially, modern lineages such as the Beijing family (Lineage 2) and the Euro-American strains (Lineage 4) are believed to more virulent compared to the other ancient Mtb lineages such as the East-African-Indian (Lineage 1) and *M. africanum* strains (Lineages 5 and 6) (Merker *et al.,* 2015). In addition, it could be shown that specific features [including Bacillus Calmette-Guerin (BCG) vaccination efficacy and drug resistance] differ for each lineage or strain (Zhang *et al.,* 2016). However, how the genomic diversity of Mtb influences TB epidemiology in clinical settings continues to be debated.

In this review, we discuss the potential implications of the genomic and genetic diversity of Mtb lineages according to recent studies conducted on humans and in experimental settings. We also highlight the importance of differential iden-

tification of strains of the Mtb complex by summarizing data from recent methods and experimental results. These data support the relevance of studying the Mtb genetic diversity for understanding differences in treatment outcome and vaccine efficacy.

Genetic difference among the Mtb lineages

In many bacterial pathogens, genetic diversity increases at the species or subspecies level through recombination, duplication, insertion and deletion events. However, gene exchange in Mtb occurs rarely; therefore, Mtb displays a distinctly clonal evolution (Nicol and Wilkinson, 2008). Seven distinct Mtb lineages have been generally identified by large sequence polymorphisms (LSPs): Lineage 1, Indo-Oceanic or East-
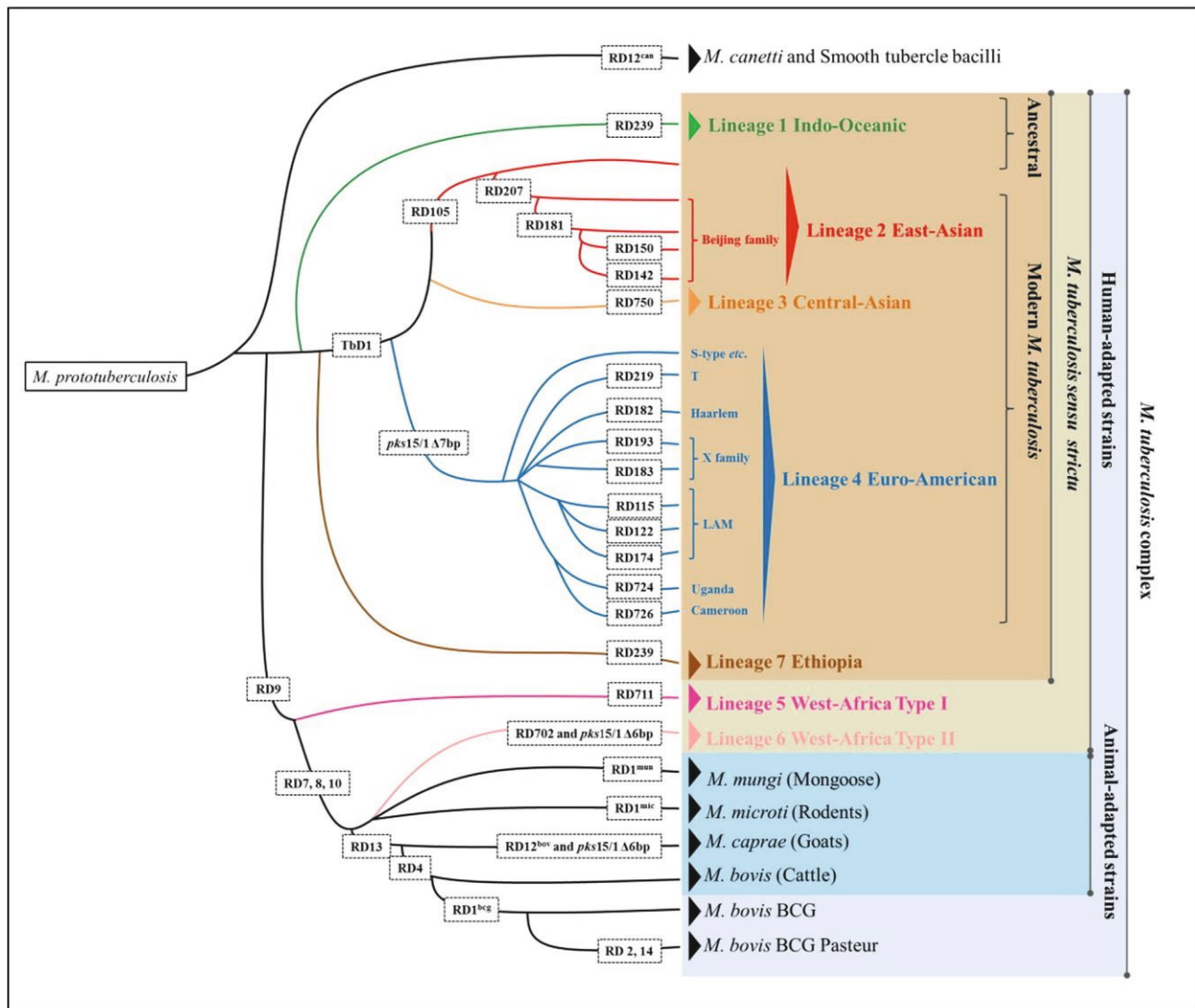

Fig. 1. Schematic diagram illustrating the evolutionary relationship among Mtb strains identified by large sequence polymorphisms [modified from Merker ***et al.* (2015)]**. Genomic regions of difference (RD) can be used to classify human- and animal-adapted strains of the *M. tuberculosis* complex based on the presence or absence of deletion sequences. The white boxes indicate which specific deletion sequence is associated with a lineage. *M. tuberculosis* strains are grouped into seven major lineages, which are in turn associated with specific geographic regions. However, lineages 2 and 4 are much more geographically widespread than other lineages.

Table 1. Genetic markers for differentiating major *Mycobacterium tuberculosis* lineages

Genetic markers and corresponding classification	*Mycobacterium tuberculosis* lineages						
	Lineage 1	Lineage 2	Lineage 3	Lineage 4	Lineage 5	Lineage 6	Lineage 7
TbD1	Ancestral	Modern	Modern	Modern	Ancestral	Ancestral	Intermediate
LSP marker Brosch *et al.* (2002)	ΔRD239	ΔRD105	ΔRD750	*pks*15/1 Δ7bp	ΔRD711	ΔRD702	ΔRD239
LSP-based classification	Indo-Oceanic	East Asian	East African-Indian	Euro-American	West-African 1	West-African 2	Ethiopian
Spoligotyping marker Brudey *et al.* (2006), Yimer *et al.* (2013), Couvin and Rastogi (2015)	Δ29-32 and 34	Δ1-34	Δ4-7 and Δ23-34	T: Δ33-36 H: Δ31 and 33-36 X: Δ18 and 33-36 LAM: Δ21-24 and 33-36	Δ7-9 and 39	Δ33-36, 39 and 43	Δ4-24, 28 and 29, Δ4-24 and 26-28
Classification by spoligotyping (Spoligotype)	East African-Indian	Beijing	Central Asian	Ghana, Haarlem, LAM and X *etc.*	*M. africanum*	*M. africanum*	Ethiopian
SNP marker Brosch *et al.* (2002)	OxyR: C37T	Rv3815c: G81A Rv0679c: C426G	RpoB: T2646G	KatG: T1388G RpoB: C3243T			
Geographical association	East Africa, South Asia, South India	East Asia, Former Soviet Union countries, South Africa	East Africa, North India, Central Asia	Americas, Europe, North Africa	Ghana, Cameroon	Senegal	The Horn of Africa

African; Lineage 2, East-Asian (includes Beijing family); Lineage 3, East-African-Indian or Central Asia strain; Lineage 4, Euro- American; Lineage 5, West-African or *M. africanum* type 1; Lineage 6, West-African or *M. africanum* type 2 and Lineage 7, Ethiopia (Fig. 1) (Coll *et al.,* 2014).

Of the seven Mtb lineages, Lineage 2, 3, and 4 have been designated as modern lineages and Lineage 1, 5, 6, and 7 have been designated as ancient lineages. This designation is based on the presence of the TbD1 sequence in their genome. TbD1-deleted lineages are considered evolutionarily modern (Lineages 2, 3, and 4), and lineages where this genomic region is conserved are considered evolutionarily ancient (Lineages 1, 5, 6, and 7; Table 1). The modern Mtb lineages form a monophyletic group in contrast to the ancient Mtb lineages (Coll *et al.,* 2014).

These human-adapted Mtb lineages are strongly associated with particular geographical regions (Hirsh *et al.,* Perez-Osorio *et al.,* 2012). Some lineages including the East-Asian

Lineage 2 and the Euro-American Lineage 4 are frequently isolated from patients with TB globally, while the prevalence of the other Mtb lineages is more regional (Gagneux, 2013; Galagan, 2014). The Indo-Oceanic Lineage 1 and the Central-Asian Lineage 3 are restricted to East Africa, Central, and South Asia. *M. africanum* type 1 and 2 (Lineages 5 and 6 respectively) are the most geographically restricted and occur almost exclusively in West Africa (Hirsh *et al.,* 2004; Gagneux *et al.,* 2006; Smith *et al.,* 2006). Similarly, the Ethiopian Lineage 7 was discovered recently and is limited to Ethiopia and recent Ethiopian immigrants. Lineages 1–4 comprise 99.2% of all cases of Mtb and the remaining percentage of cases is caused by other lineages that show restricted geographical and host distributions (Coscolla and Gagneux, 2014). The reasons for why these three lineages (Lineages 5, 6 and 7) are highly confined to specific regions are unknown (Coscolla and Gagneux, 2014).

Table 2. The major outbreak strains of *Mycobacterium tuberculosis*

Mtb Lineage	Outbreak strain	Location (time)	Characteristics	
Lineage 1	43-16836	Thailand (2010s)		Viratyosin *et al.* (2013)
	PR05	Malaysia (2013)		Ismail *et al.* (2013)
Lineage 2	HN878 (210)	Houston, Texas, and New York, USA (1990s)	▪ Expression of high lipid ▪ Rapid progression to death in mice	Sreevatsan *et al.* (1997), Barczak *et al.* (2005), Han *et al.* (2015)
	K	Kyunggi province, South Korea (1998)	▪ Most prevalent genotype in South Korea ▪ Drug-susceptible	Han *et al.* (2015)
	W	New York and other parts of USA (1990s)	▪ Strong propensity for drug resistance	Agerton *et al.* (1999), Glynn *et al.* (2002)
	CTRI-4	Tomsk, Siberia, Russian Federation (1998–1999)		
Lineage 3	CH	Leicester, UK (2001)		
Lineage 4	CDC1551	Rural area near the Kentucky-Tennessee border, USA (1994–1996)		
	KZN	Tugela Ferry, KwaZulu-Natal, South Africa (2005)	▪ Largest global outbreak of XDR TB	Cohen *et al.* (2015)
	F11	Western Cape communities of South Africa (1990s)		Victor *et al.* (2004)
	C	Los Angeles, USA (1994–1996)		Coscolla and Gagneux (2014)

Aspects of epidemiology of Mtb lineages

Differing virulence phenotypes of Mtb strains

Recent studies have explored how Mtb lineage variation is related to virulence, drug resistance, and BCG vaccination efficacy. Many experimental studies have produced evidence that clinical isolates of Mtb differ in virulence. Generally, Mtb virulence is directly linked to its transmission ability in comparison with many other bacteria; delayed immune response is one of the distinct features in Mtb infection (Ernst, 2012). More specifically, studies modeling infection with different Mtb lineages show that the modern Mtb Lineages 2, 3, and 4 induce a low level of inflammatory cytokines at early phase of infection (Table 2). They have shown that the expression of TNF-α, IFN-γ, IL-12, and IL-17 are important indicators of a protective immune response against TB, particularly against highly virulent Mtb lineages in contrast to ancestral Mtb Lineages 1 and 6 (Portevin *et al.,* 2011). These modern Mtb lineages are more widespread globally than other lineages. Therefore, many studies have reported that the modern Mtb strains are associated with the high virulence phenotype (i.e. they induce a delayed inflammatory response) and it is one of the factors linked to the global success of these lineages (Comas and Gagneux, 2011). Indeed, Reiling *et al.* (2013) have reported that modern Mtb lineages undergo fast replication in human macrophages and in an aerosol-infected mouse model. Other studies have also demonstrated induction of lower and more delayed pro-inflammatory response upon infection by modern Mtb lineages compared to infection by other lineages (Coscolla and Gagneux, 2014).

 Some studies have drawn comparisons within the group of modern Mtb lineages. One study found that Lineage 3 exhibited a high anti-inflammatory response compared to Lineage 4 (Mihret *et al.,* 2012; Newton *et al.,* 2006), and the Beijing family (that belongs to Lineage 2) commonly induces a low-level immune response (lower expression of cytokines TNF-α, IL-6, IL-10, and GRP-α) compared to Mtb H37Rv (that belongs to Lineage 4) by macrophages and dendritic cells (Wang *et al.,* 2010). In addition, Sarkar *et al.* (2012) have reported that Lineage 2 strains show a low pro-inflammatory response compared to the other two modern Mtb lineages. However, Krishnan *et al.* (2011) did not find any differences between Lineage 2 and Lineage 4. These disparate findings with respect to the immune response can be explained by differing experimental conditions and by differences in immune regulation by Mtb lineages that are expected to occur at the sub-lineage level. Such differences have been most widely studied in Lineage 2. Ribeiro *et al.* (2014) found that the more modern Beijing strains cause more cytotoxicity and lead to worse gross pathology in C57BL/6 mice. In this context, Mtb strains belonging to the W-Beijing family are the most widely studied because of their highly pathogenic phenotype. These strains are endemic in East Asia region but they have spread to New York, Russia, Latvia and South Africa since the late 1990s (Sreevatsan *et al.,* 1997). Many studies have adduced evidence from animal models that the W-Beijing family is more hyper-virulent, characterized by greater dissemination, a higher bacillary load in the sputum, and causes early death. For example of W-Beijing family, the HN878 strain belonging to Lineage 2 was a ma-

jor cause of an outbreak in Texas prisons in 1995 (Table 2) (Sreevatsan *et al.,* 1997). It is characterized by a more rapid progression of the disease to death in an immunocompetent mouse model in comparison with the clinical strain CDC-1551 (Lineage 4) and the standard laboratory strain H37Rv (Lineage 4). It is also characterized by an increased rate of relapse from the latent state (Barczak *et al.,* 2005; De Jong *et al.,* 2008; Thwaites *et al.,* 2008). Among the other strains in the Beijing family, the K strain caused a large outbreak of TB in South Korea in 1998 (Table 2). This strain is also characterized by a severe pathology and a high-level of reactivation. Moreover, Bacillus-Calmette-Guérin vaccination is less effective against the W-Beijing family than it is against other Mtb strains (including H37Rv) (Han *et al.,* 2015).

 The Mtb CH strain belonging to Lineage 3 was responsible for a large outbreak of TB in Leicester, UK (Table 2). This strain showed rapid progression to active infection within a year and the immune response of the infected macrophages is characterized by low expression of IL-12p40 and higher expression of IL-10. The ability to avoid detection by innate immune responses due to the deletion of RD750 (Rv1519-1520) is specific to Lineage 3.

 As mentioned above, many studies have found that different Mtb lineages exhibit different virulence profiles and the virulence is based on factors like bacterial uptake by host cells, intracellular growth, pathological progression, and cytokine induction. We expect that, in agreement with other studies, the high virulence Mtb strains are related to the more modern lineages of Mtb and they need to be monitored for effective TB control.

Differences in vaccine efficacy against Mtb strains

There has been no vaccine licensed for or found to be effective against TB except for the BCG vaccine. The BCG vaccine is one of the most widely used vaccines in the world and over 80% of new born children and infants receive BCG vaccination in countries where its immunization policy is introduced (World Health Organization, 2016). The efficacy of BCG vaccine against TB can vary from the absence of any protection to 80% fewer cases of TB in different geographical locations. This variation occurs due to a number of factors including BCG vaccine-host interaction, host nutritional status, environmental status, and genetic diversity of the Mtb strains (Abebe and Bjune, 2006). However, an exact mechanism explaining the variation in BCG vaccine efficacy has not been proposed to date. A putative reason for the variation in the protective efficacy of the BCG vaccine is the difference in genetic characteristics of the Mtb strains (Zhang *et al.,* 2016). The use of laboratory-adapted Mtb strains, including Mtb H37Rv and Mtb Erdman, for evaluation of vaccine efficacy, was reported to be of limited success in recent vaccine studies because of differences in virulence between laboratory-adapted strains and clinical strains (McShane and Williams, 2014; Henao-Tamayo *et al.,* 2015).

 In addition, there are two main approaches to the development of a TB vaccine: the development of a whole mycobacterial priming agent to replace the BCG agent and/or the development of a booster agent to be administered after the BCG or another vaccine (Hatherill *et al.,* 2016). Notably, the

Table 3. Characteristics of major *Mycobacterium tuberculosis* lineages

Lineage	Characteristics		
	In vitro characteristics[a]	*In vivo* and clinical characteristics[b]	
Lineage 1	■ Induce relatively weak levels of Th1-associated cytokines (IL-1β, IL-6, IL-12p40, except TNF) ■ Low uptake of bacteria by human macrophages. ■ Less virulent and more susceptible to oxidative stress than Lineage 4 ■ Lower growth in human alveolar lung cells than modern lineages	■ Lower bacillary load than Beijing family ■ Evolve into MDR[c] or XDR Mtb strains at the second highest rate following the Beijing family (Lineage 2)	Reiling *et al.* (2013), Nguyen *et al.* (2016)
Lineage 2	■ Induce relatively weak levels Th1-associated cytokines ■ Low uptake of bacteria by human macrophages ■ Higher intracellular growth in human alveolar lung cells than ancestral lineages ■ High levels of necrosis, and low levels of apoptosis in infected macrophages	■ Low vaccination efficacy ■ Rapid progression to death in animal models ■ High bacillary load and larger granulomatous lesions in lung of infected mice ■ More often isolated from the young who have been vaccinated before ■ Evolve into MDR or XDR strains at the highest rate ■ Outbreaks generally caused by MDR/XDR strains, especially from the modern Beijing family ■ Higher number of cellular infiltrates	Bifani *et al.* (2002), Abebe and Bjune (2006), Reiling *et al.* (2013), Ribeiro *et al.* (2014), Merker *et al.* (2015)
Lineage 3	■ Induce less IL-12p40 and more IL-10; rapid growth in immune cells.		Reiling *et al.* (2013)
Lineage 4	■ Induce high levels of Th1-associated cytokines ■ High uptake of bacteria by human macrophages. ■ High growth in human alveolar lung cells than ancestral lineages	■ Some strains (especially, CDC1551) less virulent than other clinical and laboratory strains, and show slow growth in infected animals. ■ KNZ strains known to evolve into MDR and XDR strains	Reiling *et al.* (2013)
Other lineages (5, 6, and 7)	■ Depending on the strains, different levels of cytokines may be induced; produce less IFN-γ than other lineages ■ Low uptake of bacteria by human macrophages ■ Low growth in human alveolar lung cells than modern lineages	■ Only confirmed outbreak of MDR strains occurred in France and Brazil (1989–1992) ■ Generally, no difference in propensity to evolve into MDR/ XDR strains than other lineages, or lower MDR/ XDR evolution in isolates ■ Lineage 5 and 6 strains have decreased virulence	Winglee *et al.* (2016)

[a] Including inflammatory response characteristics and intracellular growth
[b] Including pathology and virulence characteristics in animal models, and drug resistance
[c] MDR, Multidrug-resistant; XDR, Extensively drug-resistant

MVA85A vaccine given as a single-dose parenteral booster after BCG priming, which was being developed as a new-generation vaccine by Oxford University, showed insignificant prevention efficacy in infants. One of the reasons behind the failure might be the underestimation of Mtb genetic diversity (McShane and Williams, 2014). Especially, as exemplified by the Mtb Beijing family. As mentioned above, the Beijing family is related to one of the seven major Mtb lineages and is spread widely in Russia and East Asia, especially in countries neighboring China. Several studies covering these areas have reported that typical Beijing family isolates were more likely to be identified from patients immunized with the BCG vaccine and from areas where the BCG vaccine was prescribed (Rook *et al.,* 1981; Abebe and Bjune, 2006). Thus, several epidemiological studies have suggested that extensive BCG vaccination may be one of driving factors behind the emergence of the Beijing family (Abebe and Bjune, 2006; Jeon *et al.,* 2008).

Different lineages of Mtb are characterized by different aspects of infection in animal models. The Beijing family, which has been known to avoid BCG-induced immune responses, is an example of this. Results showed that infection by a strain of the Beijing family rapidly induces pneumonia, and TNF-α and iNOS expression when compared to infection by Mtb H37Rv. Furthermore, mice immunized with BCG immunization die rapidly and at a higher rate upon such an infection. On the other hand, when infected with *M. canettii* or a Lineage 1 strain, the mice showed limited pneumonia and continuous expression of TNF-α, and iNOS and an almost 100% survival rate (Abebe and Bjune, 2006;

Coscolla and Gagneux, 2014). In addition, the Beijing family strains consistently induce accelerated bacterial multiplication and death compared to the *M. canettii*, Haarlem, and Somali clades (Reiling *et al.,* 2013).

As mentioned above, various diagnostic techniques and genotyping methods are being commercialized and utilized against Mtb. However, most of these simply distinguish mycobacteria from other bacterial genera, or they only classify isolates as belonging to the Mtb complex, or as NTM. Techniques like multiplex PCR (both conventional PCR and real-time PCR) can be used to further identify the species of the isolates, but cannot be used to distinguish the lineage in Mtb. Therefore, a variety of diagnostic methods might yet be developed to take advantage of the specific properties of Mtb strains, for effective TB control.

Different proportions of drug resistance among Mtb strains

The molecular epidemiology of TB can be used to understand infection routes and outbreaks of drug-resistant TB. Multidrug-resistant (MDR) TB (which comprises 5% of TB cases globally) and extensively drug-resistant (XDR)-TB have been observed in various regions around the world, especially East Asia, countries of the former Soviet Union, and South Africa. Generally, development of MDR and XDR Mtb appears to result from the acquisition of genomic mutations (Lanzas *et al.,* 2013). Since drug-resistant (DR) Mtb develops upon mutations in individual resistance genes (including *rpo*A, *kat*G, and *rrs*) that cause a defect in the bacterium's life cycle, it is known that DR Mtb strains show low

infectivity (Ford *et al.,* 2013; Coscolla and Gagneux, 2014). However, some groups have reported that infectious DR Mtb strains do exist. It has been suggested that DR Mtb strains are able to spread (Table 3) (Ford *et al.,* 2013; Coscolla and Gagneux, 2014). Finally, though MDR-TB seems to be prevalent in areas with a high number of TB outbreaks, a WHO report has shown that MDR-TB occurs only in specific areas (Galagan, 2014; World Health Organization, 2016).

As mentioned above, since each Mtb lineage is almost exclusively associated with a specific region, it was hypothesized that DR, MDR, and XDR TB strains found in a specific area are related to a specific Mtb lineage. However, the global discovery of such strains has led to the belief that DR strains can be transmitted between people, and patients who carry DR strains need to be managed more thoroughly. Se-

veral studies have supported the view that strains of Mtb Lineage 2 and KwaZulu-Natal (KZN) strain have a higher propensity to develop drug-resistance than other Mtb lineages and strains (Table 3) (Buu *et al.,* 2012; Merker *et al.,* 2015). Merker *et al.* (2015) analyzed strains of the Beijing family found globally for drug resistance. Their data involved the calculation of genetic distances among Beijing sub-lineage genomes. The Beijing sub-lineages CC1 and CC2 exhibit a higher proportion of drug-resistant strains, and also a lower pairwise distance than other sub-lineages (Couvin and Rastogi, 2015; Merker *et al.,* 2015). In addition, the KZN strain belonging to Lineage 4 is known to be responsible for an outbreak of XDR-TB in 2005 in the KwaZulu-Natal region of South Africa. Genomic and proteomic analyses may provide opportunities for the discovery of poten-
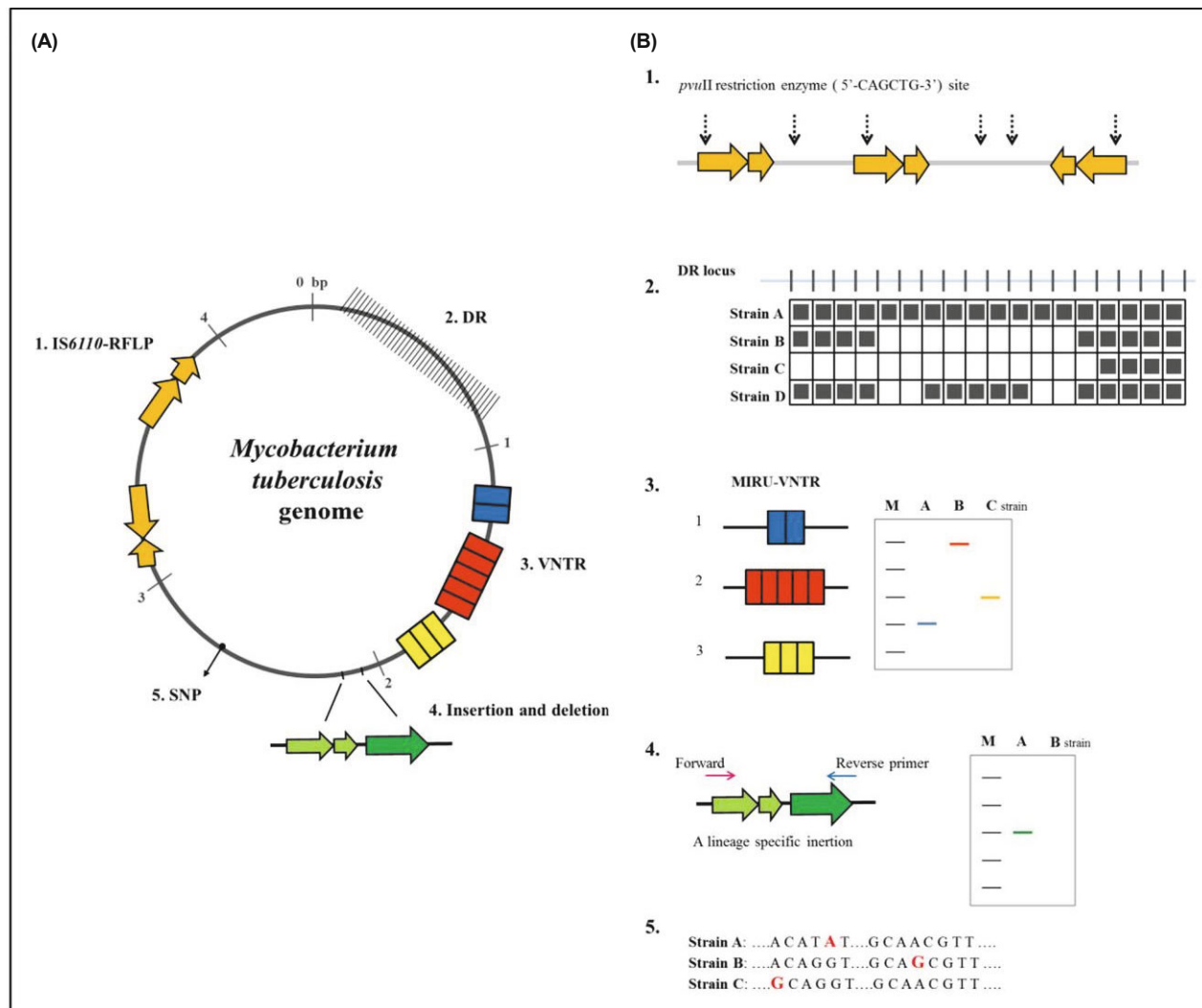

Fig. 2. Various molecular genotyping markers found in the *Mycobacterium* genome [modified from Comas *et al.* (2009)]. (1) Insertion sequence (IS) *6110*–Restriction fragment length polymorphisms arise among *M. tuberculosis* complex strains because of the variable occurrence of a specific transposase gene that carries a restriction enzyme site (*pvu*II); (2) Direct repeats (DR) of length 36 base pairs that are interspersed by unique spacer sequences (35–43 base pairs) are used in spoligotyping methods. Spoligotyping patterns are generally represented by black (signifying the presence of particular spacers) and white (signifying the absence of particular spacers) boxes; (3) Variable number tandem repeats (VNTR) of mycobacterial interspersed repetitive units (MIRU) are also employed as unique markers; (4) insertion and deletion of genes (or gene clusters) and (5) single nucleotide polymorphisms (nonsynonymous SNPs and synonymous SNPs) identified through whole-genome sequencing are also used for differentiating *M. tuberculosis* strains. Almost all methods for estimating genetic diversity are based on the use of these five genetic markers.

Table 4. Evaluation of various molecular genotyping methods

Genotyping method [a]	Principle [b]	Strength	Weakness
IS*6110*-RFLP Thierry *et al.* (1990), Kato-Maeda *et al.* (2011)	▪ Use of IS*6110*, a mobile element found frequently in the Mtb genome ▪ Analysis of digestion pattern of IS*6110* produced by the restriction enzyme *pvu*II	▪ Better resolution than the spoligotyping method ▪ Ability to identify Mtb lineages and strains	▪ Low resolution of Mtb strains harboring less than five copies of IS*6110* ▪ Requirement of large amounts of good quality genomic DNA (2–3 g), expensive computer software, and personnel with technical expertise ▪ Laborious and time-consuming
Spoligotyping Brudey *et al.* (2006), Kato-Maeda *et al.* (2011)	▪ Identification of Mtb genotype by length polymorphisms of the spacer regions between direct repeats (DRs) ▪ Analysis of binding patterns of probes specific to the spacer regions present between the GGCCA and GGCCA sequences (DRs)	▪ Relatively cost-effective and simple compared to other methods ▪ Highly reproducible, PCR-based method, requires small amount of DNA ▪ Free accessibility to the international spoligotyping database (SpolDB4) for referencing	▪ Low resolution compared to IS*6110*-RFLP ▪ Ineffective discrimination of Mtb strains belonging to the Beijing family ▪ Lack of ability to differentiate Lineage 7
MIRU-VNTR Weniger *et al.* (2010), Kato-Maeda *et al.* (2011)	▪ Analysis of copy number variation of repetitive sequences known as MIRU	▪ Better resolution than both IS*6110*-RFLP and spoligotyping ▪ A simple PCR-based method	▪ Requirement of specific equipment such as a sequencer, electro-capillary apparatus, and a specialized *in silico* automatic analytical system
Deletion analysis Coscolla and Gagneux (2014)	▪ Detection of LSPs, especially identification of irreversible deletions	▪ Fast and labor-saving ▪ PCR-based simple method ▪ Most frequently used method	▪ Occurrence of inconsistent results due individual settings and target preferences ▪ Requirement of confirmatory methods ▪ Relatively less reproducible than other methods ▪ High frequency of exceptional cases
WGS Coll *et al.* (2014)	▪ *In silico* identification of SNPs present either in the whole genome or in specific genes by comparison with sequences of reference strains	▪ The most advanced and reliable method ▪ Unrivalled resolution ▪ Provision of several adjunctive pieces of information such as drug resistance and genetic features of the targeted Mtb	▪ More expensive than other methods ▪ Requirement of specialized software and skilled personnel for data manipulation

[a] IS*6110*-RFLP, Insertion sequence *6110*-restriction fragment length polymorphism; Spoligotyping, Spacer oligonucleotide typing; MIRU-VNTR, Mycobacterial Interspersed Repetitive Unit-Variable Number of Tandem Repeat; WGS: Whole genome sequencing.
[b] Mtb, *Mycobacterium tuberculosis*; LSPs, Large sequence polymorphisms.

tial drug targets. In particular, the availability of genomic data of MDR and XDR Mtb strains (include the KZN strain) might make it possible to find a clue regarding the origin of drug resistance in bacteria that may lead to the discovery of new drug targets.

Differential identification of Mtb strains and genotyping methods

Genomic variations

Mobile elements and copy number repeats

A) Insertion Sequence (IS) *6110*
In the Mtb complex, IS*6110* is a mobile genetic element that has been utilized as the gold standard for genotyping since the 1990s (Fig. 2 and Table 4) (Thierry *et al.,* 1990). IS*6110* can be present in multiple copies in various parts of the whole genome of an Mtb complex strain. The number of copies of IS*6110* varies from strain to strain because IS is highly unstable. However, experimental evidence shows that although IS*6110* is an unstable element, transmission events involving it are very rare in mycobacteria (Warren *et al.,* 2002; Comas *et al.,* 2009). Based on this characteristic, many studies have established that IS*6110*-restriction fragment length polymorphisms (IS*6110*-RFLP) can be used to classify Mtb complex strains. RFLP classification using IS*6110* is based on the different number of copies and the different locations of IS*6110* insertion in the whole genome. Thus, classification based IS*6110* is performed by digestion with *pvu*II followed by electrophoretic separation and hybridization with an IS*6110*-derived probe (Thierry *et al.,* 1990; Comas *et al.,* 2009). This method is highly informative when there are over five IS*6110* copies in the whole genome, and it is used to determine whether isolates collected during the same outbreak come from the same strain or not (Comas *et al.,* 2009; Kato-Maeda *et al.,* 2011; Sinha *et al.,* 2016). However, IS*6110*-RFLP requires a large amount of good quality genomic DNA (about 2–3 g per reaction), and this assay is also time-consuming and difficult to reproduce because it requires experienced personnel with high technical expertise. Due to these limitations, PCR-based assays have been developed for genotyping Mtb complex strains that require only a small amount of DNA (Tortoli *et al.,* 2012; Yimer *et al.,* 2013; Chae *et al.,* 2017). Additionally, IS*6110*-RFLP cannot be performed for some strains that have less than five IS*6110* copies. Furthermore, the lack of reliability of this assay prevents the comparison of genotyping results among laboratories (Thierry *et al.,* 1990; Warren *et al.,* 2002).

B) Spacer oligonucleotide typing (spoligotyping)
Spoligotyping is a PCR-based technique for differentiating *M. tuberculosis* strains (Fig. 2 and Table 4). It is based on the existence of unique spacer regions between direct repeats,

which are known as the 'clustered regularly interspaced short palindromic repeats', in the Mtb genome (Fig. 2) (Groenen *et al.,* 1993). During spoligotyping, a primer pair amplifies the spacer present in the direct repeat locus and the products of the amplification are covalently hybridized to oligonucleotide probes derived from direct repeat spacers specific to a strain (Brudey *et al.,* 2006; Oelemann *et al.,* 2007). The "spoligo-type" pattern so generated is compared to patterns of spoligo-type found in the international spoligotyping database (SpolDB4) that includes thousands of spoligo-type patterns (Brudey *et al.,* 2006). This assay was originally developed to identify the route of infection by an Mtb complex strain. However, this genotyping method is of limited utility when drawing a phylogenetic tree and for accurate strain discrimination (particularly, discrimination of the Lineage 2/Beijing family) because many spoligo-types share similar patterns (especially Lineage 3/ CAS family) (Devi *et al.,* 2015).

C) Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU-VNTR)
The first whole genome sequence of Mtb was finished in 1998 (Cole *et al.,* 1998). Afterwards, comparative genome analysis was performed on several Mtb genomes leading to the discovery of VNTRs known as MIRUs in the Mtb complex (Fig. 2 and Table 4). These were soon utilized as a new method for molecular genotyping of Mtb. An MIRU is a repeating unit composed of dozens of base pairs and is used to identify strains by variation in its number (Fig. 2) (Weniger *et al.,* 2010; Couvin and Rastogi, 2015). These days, it has become a preferred genotyping technique over spoligotyping in many countries. Strain genotyping using VNTR has also developed rapidly as an alternative to IS*6110*-RFLP because it is a PCR-based assay and thus needs only a small amount of DNA. This simple process also provides high speed and accuracy (Weniger *et al.,* 2010). In addition, this technique lends itself to digitization and has allowed the development of an online database system called VNTR plus (Oelemann *et al.,* 2007). However, this technique possesses less discrimination ability than IS*6110*-RFLP, because different VNTR loci are used by each country and laboratory. This means that the Mtb complex strains found in each country are identified by different VNTR characteristics. Therefore, further experiments with the appropriate combination of VNTR loci are needed to establish the standard VNTR method (Kato-Maeda *et al.,* 2011).

Large Sequence Polymorphisms (LSPs)
The analysis results of Mtb genome sequencing led to the discovery of LSPs (including region of difference; RD) (Comas *et al.,* 2009). Even though horizontal gene transmission events have been detected in other mycobacteria species, the genetic components among the Mtb complex is highly conserved with 99.9% similarity at the nucleotide level (Galagan, 2014). Since horizontal gene transfer rarely occurs in the Mtb complex, LSPs could be suitable phylogenetic markers for lineage and sub-lineage classification. For example, RDs in mycobacteria have been used as major genetic markers to classify the strains of the Mtb complex into main lineages and sub-lineages (Fig. 2 and Table 4) (Hirsh *et al.,* 2004; Mostowy *et al.,* 2004a, 2004b). In additions, representatively Lineages 2, 3, and 4

are modern lineage and Lineages 1, 5, and 6 are ancient lineage based on the presence of TbD1 (Brosch *et al.,* 2002; Achtman, 2008).

Besides RDs, there are several new genetic markers known under the category of LSPs. For example, the insertion of the IS*6110* gene into a specific region of the genome can lead to the development of an LSP. In 1991, B0/W 148 strains were isolated in New York; the characteristic of these strains is that the IS*6110* gene is inserted into the noise transfer function (NTF) region of these strains. Within the Mtb Beijing family, the strains whose NTF region is found to carry the IS*6110* gene are classified as the modern type, whereas the other strains are classified as the ancient type (Bifani *et al.,* 2002; Tsolaki *et al.,* 2005; Merker *et al.,* 2015). In addition, our group recently found a unique gene cluster enabling to identify the Beijing family among the Mtb complex using whole genome sequencing. Since this region was highly conserved within the Beijing family, we developed a multiplex PCR assay for distinguishing the Beijing family not only from other major nontuberculous mycobacteria but also from other Mtb complex (Chae *et al.,* 2017).

Single nucleotide polymorphisms (SNPs)
LSPs can serve as robust markers for phylogenetic classification. However, LSPs are not different enough to allow distinction between closely related Mtb strains (Galagan, 2014). On the other hand, SNPs encompass all types of mutations in the whole genome and provide much better resolution than LSPs (Fig. 2 and Table 4) (Coll *et al.,* 2014). For example, as mention above, the Beijing family is associated with a unique spoligo-type pattern and the unique presence of RD 207. However, the use of these markers still seems to carry many disadvantages, since it relies on the unstable insertion and deletion of elements. Therefore, a mutation in the *rv0679*c locus allows distinguishing Lineage 2 in Mtb from other lineages which supports the evidence of LSPs marker with the reported advantages of SNPs (Nakajima *et al.,* 2013). The discovery of new SNPs by whole-genome sequencing approaches has the potential to produce more extensive insights into the molecular and epidemiological mechanisms of Mtb transmission, as well as allow the development of better diagnostic tools and therapies (Niemann and Supply, 2014). In addition, SNPs analysis can allow the calculation of genetic distances between Mtb strains and lead to the inference of their overall phylogeny, and the widespread occurrence of SNPs in the Mtb complex genome has been used for producing a phylogenetic classification of the Mtb complex (Coll *et al.,* 2014). Real-time PCR and sequencing are known techniques for SNP detection. These methods are important to understand the current state of the Mtb complex, as well as to help predict the direction of its evolution. Coscolla and Gagneux (2014) have reported that the strains belonging to the modern Lineage 2, 3, and 4 differ by 970 SNPs on average. Strains belonging to the ancient Lineage 1, 5, and 6 are more distantly related, with an average distance of 1500 SNPs between them. This information can be used to construct a phylogenetic tree and identify which Mtb strains have led to recent outbreaks. In addition, because horizontal gene transfer among Mtb complex strains is rare, SNPs can serve as a robust marker for identifying Mtb strains according to their

phylogenetic classification. The publication of whole genome sequencing (WGS) results by Rose *et al.* (2013) led to the identification of all SNPs specific to the different lineages of the Mtb complex. The number of SNPs ranges from 124 in Lineage 2 (a modern Mtb lineage) to 698 in Lineage 5 (an ancient Mtb lineage). In all cases, over 40% of the non-synonymous SNPs were fixed within their respective Mtb lineages and therefore they can be predicted to be involved in gene function. These data reveal that although the general genomic diversity of Mtb is lower than other bacterial genera (Achtman, 2008), a lot of SNPs have accumulated within each Mtb lineage and they are likely to contribute to phenotypic differences among them. The most common SNPs in Mtb are mutations associated with drug resistance and virulence (Perez-Osorio *et al.,* 2012; Casali *et al.,* 2014). Due to these mutations, certain gene products cannot be expressed, leading to an interruption in the metabolism of certain antibiotics and attenuation of virulence (Coll *et al.,* 2014).

The *katG* and *phoPR* genes provide particularly illustrative examples. First, the Ser375Thr mutation in *katG* confers resistance to isoniazid (Mokrousov *et al.,* 2002). *katG* encodes a catalase-peroxidase that activates the isoniazid pro-drug. However, the Ser375Thr mutation prevents activation of isoniazid, resulting in resistance to it. Second, *phoPR*, which regulates a two-component virulence system, is essential for Mtb complex virulence including the regulation of secretion of ESAT-6, acyltrehalose-based lipids, and the modulation of antigen export (Coscolla and Gagneux, 2014). SNPs in this gene have been reported to lead to critical Mtb phenotype changes. If an amino acid changes at position 219 of *phoP* in the Mtb H37Rv strain, its virulence is remarkably attenuated and it is then referred as the Mtb H37Ra strain. (Coll *et al.,* 2014; Coscolla and Gagneux, 2014).

Genotyping and diagnosis methods

Commercialized assays

The molecular diagnosis of TB directly from clinical isolates has been performed since the early 1990s. These methods compensate for defects in microscopy and culture-based methods, the major disadvantages of which are poor sensitivity (requiring 10^3 CFU/ml within sputum) and a narrow detection range (Sarmiento *et al.,* 2003). Many studies have introduced various genotyping methods and molecular diagnostic assays based on several different technologies.

A) PCR-based methods
One of the first commercialized assays was performed using multiplex PCR (both conventional PCR and real-time PCR), which was based on the amplification of housekeeping genes (e.g. 16S rRNA genes common to all mycobacteria) and Mtb-specific insertion sequences (e.g. IS*6110* or RD sequences) (Tevere *et al.,* 1996). Conventional multiplex PCR is based on the simultaneous use of multiple specific primers and real-time multiplex PCR use primers with fluorescent probes. Unlike microscopy-based tests, this assay can classify species and subspecies belonging to the Mtb complex and the NTM group (Kim *et al.,* 2013; Mutingwende *et al.,* 2015; Chae *et al.,* 2017).

In general, sensitivity and specificity of multiplex PCR as-

says are higher than that of other commercial methods. On average, a sensitivity of 75.5% is obtained with conventional PCR and a sensitivity of 77.3% is obtained with real-time PCR (not a significant difference). The specificity of both methods with sputum samples is 99.5%. However, cerebrospinal fluid, pleural fluid, and urine samples show a lower sensitivity because of inhibitory substances (< 50%) (Sarmiento *et al.,* 2003; Tortoli *et al.,* 2012).

GenXpert Mtb/RIF is a cartridge-based nucleotide amplification assay and simultaneously detects Mtb and rifampicin (RIF) resistance. Since 2010, this assay has been used in TB endemic countries (including China and countries in Africa) (Detjen *et al.,* 2015). This assay is based on the amplification of Mtb and RIF specific DNA sequences by PCR; genomic DNA from the sputum samples of patients is first purified and concentrated and then clinically relevant RIF resistance-inducing mutations (located in the RNA polymerase beta gene) are identified. The assay requires only 90 min for obtaining results and it reduces biohazard risk and technical training time (Detjen *et al.,* 2015).

The Line probe assay (LiPA) is similarly able to identify an Mtb complex strain and detect genetic mutations in the *katG* gene region, which is related to isoniazid (INH)-resistance, and the *rpoB* gene region, which is related to RIF-resistance. This assay is based on the use of several nitrocellulose paper strips layered with specific 10-oligonucleotide probes. After the *katG* and *rpoB* regions are amplified using PCR, the PCR products are hybridized with the oligonucleotide probes present on the paper strips. The results are determined by the absence or presence of various colored lines (Morgan *et al.,* 2005).

B) Enzyme-linked immunosorbent assay (ELISA)-based methods
Interferon-γ release assay (IGRAs) can detect latent TB infection based on the cellular immune response. After placing a sample of whole-blood in a tube coated with the peptides including ESAT-6, CFP-10, and TB7.7, interferon-γ released by T cells is measured. The assay has better sensitivity and specificity than the tuberculin skin test and shows no reaction with samples of BCG-vaccinated individuals (Lalvani and Pareek, 2010; Starke, 2014).

For diagnosis of TB in HIV-infected patients, recent studies have described the use of the lipoarabinomannan (LAM)-ELISA assay. LAM is a glycolipid found in the Mtb cell wall component and serves as a major virulence factor. Since this assay employs urine-based testing for identification of LAM in the urine of patients, samples for it are easier to collect and store than samples for sputum-based testing. The sensitivity of LAM-ELISA assay falls between 23–84%, and its specificity falls between 75–99%. On average, the specificity of this assay is higher than its sensitivity. A previous study tried to combine urine and sputum samples for performing the LAM assay. It reported a poorer specificity (15–21%) but higher sensitivity (86–90%) compared to LAM assays performed on urine-only samples (Dheda *et al.,* 2010). Although LAM is still considered unnecessary for diagnosing TB patients in a clinical setting, its utility has gradually advanced for diagnosing TB in HIV-infected patients.

Conclusions and perspectives

During the past decades, the importance of understanding the genetic diversity of Mtb lineages has been underscored by differences in their virulence and geographic dominance. It is plausible that with the development of more efficient molecular techniques and genetic tools, our knowledge of the genetic diversity of the Mtb complex strains and functions of the specific components of their genomes will expand. Unsurprisingly, human genetic variation is also known to play an important role in the spread of TB since the success of an Mtb infection is reciprocally influenced by genetic variations of the host and the pathogen. Accumulating evidence suggests that pathogenic, phenotypic, and genotypic variation among Mtb strains contributes to differences in vaccine efficacy, treatment outcomes, and the development of MDR/XDR Mtb strains. Thus, a better understanding of the genetic diversity of Mtb isolates found in clinical settings is needed. In addition, since most TB patients undergo latent infection prior to getting the active disease, further research is needed to find out the genotypes and lineages of Mtb which reactivate most frequently from the latent state; such research is important for controlling the spread of TB. Recently, several molecular genotyping methods have been introduced in experimental and clinical settings to identify specific genotypes of the Mtb complex. Although differential identification of Mtb genotypes and lineages has not been conducted routinely in clinical settings, some studies have reported on the association between the phylogenetic age of Mtb lineages and differences in their epidemiology. Epidemiological patterns show that the more modern lineages of Mtb are more virulent and exhibit more successful transmission globally. According to multiple criteria (immune response, virulence, drug resistance and BCG vaccine efficacy) Lineage 2 and Lineage 4 are more virulent phenotype than Lineage 1 and *M. africanum* (Lineage 6). Other lineages (Lineage 3, Lineage 5, and Lineage 7) need to undergo more studies.

A better understanding of the phenotypic and pathogenic variations arising from the genetic diversity of Mtb strains will facilitate the development of effective prevention and treatment strategies, including the development of new drugs and more effective vaccines against Mtb infection. Furthermore, the identification of genetic factors encoding virulence in clinically significant strains is necessary for differentiating Mtb strains and for improving public health measures.

Acknowledgements

References

Abebe, F. and Bjune, G. 2006. The emergence of Beijing family genotypes of *Mycobacterium tuberculosis* and low-level protection by bacille Calmette-Guérin (BCG) vaccines: Is there a link? *Clin. Exp. Immunol.* **145**, 389–397.

Achtman, M. 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu. Rev. Microbiol.* **62**, 53–70.

Agerton, T.B., Valway, S.E., Blinkhorn, R.J., Shilkret, K.L., Reves, R., Schluter, W.W., Gore, B., Pozsik, C.J., Plikaytis, B.B., Woodley, C., et al. 1999. Spread of strain w, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. *Clin. Infect. Dis.* **29**, 85–92; discussion 93–85.

Barczak, A.K., Domenech, P., Boshoff, H.I., Reed, M.B., Manca, C., Kaplan, G., and Barry, C.E. 3rd. 2005. *In vivo* phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J. Infect. Dis.* **192**, 600–606.

Bifani, P.J., Mathema, B., Kurepina, N.E., and Kreiswirth, B.N. 2002. Global dissemination of the *Mycobacterium tuberculosis* W-beijing family strains. *Trends Microbiol.* **10**, 45–52.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., et al. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. USA* **99**, 3684–3689.

Brudey, K., Driscoll, J.R., Rigouts, L., Prodinger, W.M., Gori, A., Al-Hajoj, S.A., Allix, C., Aristimuno, L., Arora, J., Baumanis, V., et al. 2006. *Mycobacterium tuberculosis* complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol.* **6**, 23.

Buu, T.N., van Soolingen, D., Huyen, M.N., Lan, N.T., Quy, H.T., Tiemersma, E.W., Kremer, K., Borgdorff, M.W., and Cobelens, F.G. 2012. Increased transmission of *Mycobacterium tuberculosis* beijing genotype strains associated with resistance to streptomycin: A population-based study. *PLoS One* **7**, e42323.

Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S.R., Ignatyeva, O., Kontsevaya, I., Corander, J., Bryant, J., Parkhill, J., Nejentsev, S., et al. 2014. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46**, 279–286.

Chae, H., Han, S.J., Kim, S.Y., Ki, C.S., Huh, H.J., Yong, D., Koh, W.J., and Shin, S.J. 2017. Development of a one-step multiplex PCR assay for differential detection of major *Mycobacterium* species. *J. Clin. Microbiol.* **55**, 2736–2751.

Cohen, K.A., Abeel, T., Manson McGuire, A., Desjardins, C.A., Munsamy, V., Shea, T.P., Walker, B.J., Bantubani, N., Almeida, D.V., Alvarado, L., et al. 2015. Evolution of extensively drug-resistant tuberculosis over four decades: Whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from Kwa-Zulu-Natal. *PLoS Med.* **12**, e1001880.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

Coll, F., McNerney, R., Guerra-Assuncao, J.A., Glynn, J.R., Perdigao, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., and Clark, T.G. 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5**, 4812.

Comas, I. and Gagneux, S. 2011. A role for systems epidemiology in tuberculosis research. *Trends Microbiol.* **19**, 492–500.

Comas, I., Homolka, S., Niemann, S., and Gagneux, S. 2009. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS One* **4**, e7815.

Coscolla, M. and Gagneux, S. 2010. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov. Today Dis. Mech.* **7**, e43–e59.

Coscolla, M. and Gagneux, S. 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* **26**, 431–444.

Couvin, D. and Rastogi, N. 2015. Tuberculosis - a global emergency:

Tools and methods to monitor, understand, and control the epidemic with specific example of the beijing lineage. *Tuberculosis (Edinb.).* **95 Suppl 1**, S177–S189.

De Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M., Jackson-Sillah, D.J., Fox, A., Deriemer, K., Gagneux, S., *et al.* 2008. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in the gambia. *J. Infect. Dis.* **198**, 1037–1043.

Detjen, A.K., DiNardo, A.R., Leyden, J., Steingart, K.R., Menzies, D., Schiller, I., Dendukuri, N., and Mandalakas, A.M. 2015. Xpert MTB/Rif assay for the diagnosis of pulmonary tuberculosis in children: A systematic review and meta-analysis. *Lancet Respir. Med.* **3**, 451–461.

Devi, K.R., Bhutia, R., Bhowmick, S., Mukherjee, K., Mahanta, J., and Narain, K. 2015. Genetic diversity of *Mycobacterium tuberculosis* isolates from Assam, India: Dominance of beijing family and discovery of two new clades related to CAS1_Delhi and EAI family based on spoligotyping and MIRU-VNTR typing. *PLoS One* **10**, e0145860.

Dheda, K., Davids, V., Lenders, L., Roberts, T., Meldau, R., Ling, D., Brunet, L., van Zyl Smit, R., Peter, J., Green, C., *et al.* 2010. Clinical utility of a commercial LAM-ELISA assay for TB diagnosis in HIV-infected patients using urine and sputum samples. *PLoS One* **5**, e9848.

Ernst, J.D. 2012. The immunological life cycle of tuberculosis. *Nat. Rev. Immunol.* **12**, 581–591.

Ford, C.B., Shah, R.R., Maeda, M.K., Gagneux, S., Murray, M.B., Cohen, T., Johnston, J.C., Gardy, J., Lipsitch, M., and Fortune, S.M. 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. 2013. Genetic diversity in *Mycobacterium tuberculosis*. *Curr. Top. Microbiol. Immunol.* **374**, 1–25.

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B.C., Narayanan, S., Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C., *et al.* 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **103**, 2869–2873.

Galagan, J.E. 2014. Genomic insights into tuberculosis. *Nat. Rev. Genet.* **15**, 307–320.

Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K., and van Soolingen, D. 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: A systematic review. *Emerg. Infect. Dis.* **8**, 843–849.

Groenen, P.M., Bunschoten, A.E., van Soolingen, D., and van Embden, J.D. 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* **10**, 1057–1065.

Han, S.J., Song, T., Cho, Y.J., Kim, J.S., Choi, S.Y., Bang, H.E., Chun, J., Bai, G.H., Cho, S.N., and Shin, S.J. 2015. Complete genome sequence of *Mycobacterium tuberculosis* K from a korean high school outbreak, belonging to the Beijing family. *Stand. Genomic Sci.* **10**, 78.

Hatherill, M., Tait, D., and McShane, H. 2016. Clinical testing of tuberculosis vaccine candidates. *Microbiol. Spectr.* **4**, 5.

Henao-Tamayo, M., Shanley, C.A., Verma, D., Zilavy, A., Stapleton, M.C., Furney, S.K., Podell, B., and Orme, I.M. 2015. The efficacy of the BCG vaccine against newly emerging clinical strains of *Mycobacterium tuberculosis*. *PLoS One* **10**, e0136500.

Hirsh, A.E., Tsolaki, A.G., DeRiemer, K., Feldman, M.W., and Small, P.M. 2004. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc. Natl. Acad. Sci. USA* **101**, 4871–4876.

Ismail, A., Teh, L.K., Ngeow, Y.F., Norazmi, M.N., Zainul, Z.F., Tang, T.H., Najimudin, N., and Salleh, M.Z. 2013. Draft genome sequence of a clinical isolate of *Mycobacterium tuberculosis* strain PR05. *Genome Announc.* **1**, 3.

Jeon, B.Y., Derrick, S.C., Lim, J., Kolibab, K., Dheenadhayalan, V., Yang, A.L., Kreiswirth, B., and Morris, S.L. 2008. *Mycobacterium bovis* BCG immunization induces protective immunity against nine different *Mycobacterium tuberculosis* strains in mice. *Infect. Immun.* **76**, 5173–5180.

Kato-Maeda, M., Metcalfe, J.Z., and Flores, L. 2011. Genotyping of *Mycobacterium tuberculosis*: Application in epidemiologic studies. *Future Microbiol.* **6**, 203–216.

Kim, Y., Choi, Y., Jeon, B.Y., Jin, H., Cho, S.N., and Lee, H. 2013. A simple and efficient multiplex PCR assay for the identification of *Mycobacterium* genus and *Mycobacterium tuberculosis* complex to the species level. *Yonsei Med. J.* **54**, 1220–1226.

Koch, R. 1882. Die Ätiologie der Tuberkulose. *Berliner Klin Wochenschr* **19**, 221–230.

Krishnan, N., Malaga, W., Constant, P., Caws, M., Tran, T.H., Salmons, J., Nguyen, T.N., Nguyen, D.B., Daffe, M., Young, D.B., *et al.* 2011. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLoS One* **6**, e23870.

Lalvani, A. and Pareek, M. 2010. Interferon gamma release assays: Principles and practice. *Enferm. Infecc. Microbiol. Clin.* **28**, 245–252.

Lanzas, F., Karakousis, P.C., Sacchettini, J.C., and Ioerger, T.R. 2013. Multidrug-resistant tuberculosis in panama is driven by clonal expansion of a multidrug-resistant *Mycobacterium tuberculosis* strain related to the KZN extensively drug-resistant *M. tuberculosis* strain from South Africa. *J. Clin. Microbiol.* **51**, 3277–3285.

McShane, H. and Williams, A. 2014. A review of preclinical animal models utilised for TB vaccine evaluation in the context of recent human efficacy data. *Tuberculosis (Edinb).* **94**, 105–110.

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., Blum, M.G., Rusch-Gerdes, S., Mokrousov, I., Aleksic, E., *et al.* 2015. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet.* **47**, 242–249.

Mihret, A., Bekele, Y., Loxton, A.G., Aseffa, A., Howe, R., and Walzl, G. 2012. Plasma level of IL-4 differs in patients infected with different modern lineages of *M. tuberculosis*. *J. Trop. Med.* **2012**, 518564.

Mokrousov, I., Otten, T., Filipenko, M., Vyazovaya, A., Chrapov, E., Limeschenko, E., Steklova, L., Vyshnevskiy, B., and Narvskaya, O. 2002. Detection of isoniazid-resistant *Mycobacterium tuberculosis* strains by a multiplex allele-specific PCR assay targeting *katG* codon 315 variation. *J. Clin. Microbiol.* **40**, 2509–2512.

Morgan, M., Kalantri, S., Flores, L., and Pai, M. 2005. A commercial line probe assay for the rapid detection of rifampicin resistance in *Mycobacterium tuberculosis*: A systematic review and meta-analysis. *BMC Infect. Dis.* **5**, 62.

Mostowy, S., Cousins, D., and Behr, M.A. 2004a. Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the *Mycobacterium tuberculosis* complex. *J. Bacteriol.* **186**, 104–109.

Mostowy, S., Onipede, A., Gagneux, S., Niemann, S., Kremer, K., Desmond, E.P., Kato-Maeda, M., and Behr, M. 2004b. Genomic analysis distinguishes *Mycobacterium africanum*. *J. Clin. Microbiol.* **42**, 3594–3599.

Mutingwende, I., Vermeulen, U., Steyn, F., Viljoen, H., and Grobler, A. 2015. Development and evaluation of a rapid multiplex-PCR based system for *Mycobacterium tuberculosis* diagnosis using sputum samples. *J. Microbiol. Methods* **116**, 37–43.

Nakajima, C., Tamaru, A., Rahim, Z., Poudel, A., Maharjan, B., Khin Saw, A., Ling, H., Hattori, T., Iwamoto, T., Fukushima, Y., *et al.* 2013. Simple multiplex PCR assay for identification of Beijing family *Mycobacterium tuberculosis* isolates with a lineage-specific mutation in *Rv0679c*. *J. Clin. Microbiol.* **51**, 2025–2032.

Newton, S.M., Smith, R.J., Wilkinson, K.A., Nicol, M.P., Garton, N.J., Staples, K.J., Stewart, G.R., Wain, J.R., Martineau, A.R., Fandrich, S., *et al.* 2006. A deletion defining a common asian lineage of

Mycobacterium tuberculosis associates with immune subversion. *Proc. Natl. Acad. Sci. USA* **103**, 15594–15598.

Nguyen, V.A., Banuls, A.L., Tran, T.H., Pham, K.L., Nguyen, T.S., Nguyen, H.V., Nguyen, N.L., Nguyen, N.L., Dang, D.A., Marks, G.B., *et al.* 2016. *Mycobacterium tuberculosis* lineages and anti-tuberculosis drug resistance in reference hospitals across Viet Nam. *BMC Microbiol.* **16**, 167.

Nicol, M.P. and Wilkinson, R.J. 2008. The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. *Trans. R. Soc. Trop. Med. Hyg.* **102**, 955–965.

Niemann, S. and Supply, P. 2014. Diversity and evolution of *Mycobacterium tuberculosis*: Moving to whole-genome-based approaches. *Cold Spring Harb. Perspect. Med.* **4**, a021188.

Oelemann, M.C., Diel, R., Vatin, V., Haas, W., Rusch-Gerdes, S., Locht, C., Niemann, S., and Supply, P. 2007. Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligo-typing for population-based molecular epidemiology studies of tuberculosis. *J. Clin. Microbiol.* **45**, 691–697.

Perez-Osorio, A.C., Boyle, D.S., Ingham, Z.K., Ostash, A., Gautom, R.K., Colombel, C., Houze, Y., and Leader, B.T. 2012. Rapid identification of mycobacteria and drug-resistant *Mycobacterium tuberculosis* by use of a single multiplex PCR and DNA sequencing. *J. Clin. Microbiol.* **50**, 326–336.

Portevin, D., Gagneux, S., Comas, I., and Young, D. 2011. Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog.* **7**, e1001307.

Reiling, N., Homolka, S., Walter, K., Brandenburg, J., Niwinski, L., Ernst, M., Herzmann, C., Lange, C., Diel, R., Ehlers, S., *et al.* 2013. Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. *mBio* **4**, e00250–13.

Ribeiro, S.C., Gomes, L.L., Amaral, E.P., Andrade, M.R., Almeida, F.M., Rezende, A.L., Lanes, V.R., Carvalho, E.C., Suffys, P.N., Mokrousov, I., *et al.* 2014. *Mycobacterium tuberculosis* strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. *J. Clin. Microbiol.* **52**, 2615–2624.

Rook, G.A., Bahr, G.M., and Stanford, J.L. 1981. The effect of two distinct forms of cell-mediated response to mycobacteria on the protective efficacy of BCG. *Tubercle.* **62**, 63–68.

Rose, G., Cortes, T., Comas, I., Coscolla, M., Gagneux, S., and Young, D.B. 2013. Mapping of genotype-phenotype diversity among clinical isolates of *Mycobacterium tuberculosis* by sequence-based transcriptional profiling. *Genome Biol. Evol.* **5**, 1849–1862.

Sarkar, R., Lenders, L., Wilkinson, K.A., Wilkinson, R.J., and Nicol, M.P. 2012. Modern lineages of *Mycobacterium tuberculosis* exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. *PLoS One* **7**, e43170.

Sarmiento, O.L., Weigle, K.A., Alexander, J., Weber, D.J., and Miller, W.C. 2003. Assessment by meta-analysis of PCR for diagnosis of smear-negative pulmonary tuberculosis. *J. Clin. Microbiol.* **41**, 3233–3240.

Sinha, P., Gupta, A., Prakash, P., Anupurba, S., Tripathi, R., and Srivastava, G.N. 2016. Differentiation of *Mycobacterium tuberculosis* complex from non-tubercular mycobacteria by nested multiplex PCR targeting *IS6110, MTP40* and *32kD alpha antigen* encoding gene fragments. *BMC Infect. Dis.* **16**, 123.

Smith, N.H., Kremer, K., Inwald, J., Dale, J., Driscoll, J.R., Gordon, S.V., van Soolingen, D., Hewinson, R.G., and Smith, J.M. 2006. Ecotypes of the *Mycobacterium tuberculosis* complex. *J. Theor. Biol.* **239**, 220–225.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., and Musser, J.M. 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl.*

Acad. Sci. USA **94**, 9869–9874.

Starke, J.R. 2014. Interferon-gamma release assays for diagnosis of tuberculosis infection and disease in children. *Pediatr.* **134**, e1763–e1773.

Tevere, V.J., Hewitt, P.L., Dare, A., Hocknell, P., Keen, A., Spadoro, J.P., and Young, K.K. 1996. Detection of *Mycobacterium tuberculosis* by PCR amplification with pan-*Mycobacterium* primers and hybridization to an *M. tuberculosis*-specific probe. *J. Clin. Microbiol.* **34**, 918–923.

Thierry, D., Brisson-Noel, A., Vincent-Levy-Frebault, V., Nguyen, S., Guesdon, J.L., and Gicquel, B. 1990. Characterization of a *Mycobacterium tuberculosis* insertion sequence, *IS6110*, and its application in diagnosis. *J. Clin. Microbiol.* **28**, 2668–2673.

Thwaites, G., Caws, M., Chau, T.T., D'Sa, A., Lan, N.T., Huyen, M.N., Gagneux, S., Anh, P.T., Tho, D.Q., Torok, E., *et al.* 2008. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J. Clin. Microbiol.* **46**, 1363–1368.

Tortoli, E., Urbano, P., Marcelli, F., Simonetti, T.M., and Cirillo, D.M. 2012. Is real-time PCR better than conventional PCR for *Mycobacterium tuberculosis* complex detection in clinical samples? *J. Clin. Microbiol.* **50**, 2810–2813.

Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y.O., Kreiswirth, B.N., Van Soolingen, D., and Small, P.M. 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **43**, 3185–3191.

Victor, T.C., de Haas, P.E., Jordaan, A.M., van der Spuy, G.D., Richardson, M., van Soolingen, D., van Helden, P.D., and Warren, R. 2004. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a western cape F11 genotype. *J. Clin. Microbiol.* **42**, 769–772.

Viratyosin, W., Kulawonganunchai, S., Smittipat, N., Juthayothin, T., Penpassakarn, P., Pasomsub, E., Chantratita, W., Chaiprasert, A., and 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**, e00801-13.

Wang, C., Peyron, P., Mestre, O., Kaplan, G., van Soolingen, D., Gao, Q., Gicquel, B., and Neyrolles, O. 2010. Innate immune response to *Mycobacterium tuberculosis* Beijing and other genotypes. *PLoS One* **5**, e13594.

Warren, R.M., van der Spuy, G.D., Richardson, M., Beyers, N., Borgdorff, M.W., Behr, M.A., and van Helden, P.D. 2002. Calculation of the stability of the *IS6110* banding pattern in patients with persistent *Mycobacterium tuberculosis* disease. *J. Clin. Microbiol.* **40**, 1705–1708.

Weniger, T., Krawczyk, J., Supply, P., Niemann, S., and Harmsen, D. 2010. MIRU-VNTR*plus*: A web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic Acids Res.* **38**, W326–W331.

Winglee, K., Manson McGuire, A., Maiga, M., Abeel, T., Shea, T., Desjardins, C.A., Diarra, B., Baya, B., Sanogo, M., Diallo, S., *et al.* 2016. Whole genome sequencing of *Mycobacterium africanum* strains from Mali provides insights into the mechanisms of geographic restriction. *PLoS Negl. Trop. Dis.* **10**, e0004332.

World Health Organization. 2016. Global tuberculosis report 2016.

Yimer, S.A., Hailu, E., Derese, Y., Bjune, G.A., and Holm-Hansen, C. 2013. Spoligotyping of *Mycobacterium tuberculosis* isolates among pulmonary tuberculosis patients in Amhara Region, Ethiopia. *Acta Pathol. Microbiol. Immunol. Scand.* **121**, 878–885.

Zhang, L., Ru, H.W., Chen, F.Z., Jin, C.Y., Sun, R.F., Fan, X.Y., Guo, M., Mai, J.T., Xu, W.X., Lin, Q.X., *et al.* 2016. Variable virulence and efficacy of BCG vaccine strains in mice and correlation with genome polymorphisms. *Mol. Ther.* **24**, 398–405.