ACTA ABP BIOCHIMICA POLONICA

Vol. 62, No 2/2015
317–322
http://dx.doi.org/10.18388/abp.2015_1037

Regular paper

A novel method of *Mycobacterium tuberculosis* complex strain differentiation using polymorphic GC-rich gene sequences

Roman Kotłowski✉

Department of Molecular Biotechnology and Microbiology, Gdansk University of Technology, Gdańsk, Poland

Tuberculosis is one of the leading infectious diseases. In this work, a new genotyping method of *Mycobacterium tuberculosis* (Mtb) complex strain is presented. 27 Mtb genomes were analyzed for the presence of length polymorphism within polymorphic GC-rich gene sequences. Four genes, Rv3345c, Rv3507, Rv0747 and Rv3511, showing variation in length depending on the Mtb strain were selected for designing primer sequences flanking variable regions for the PCR method. Identification of 16 genotypes among 27 analyzed genomes demonstrated usefulness of our genotyping method in differentiation of Mtb genomes based on sequence polymorphism in the four PGRS genes.

Key words: *Mycobacterium tuberculosis* complex, PGRS, genotyping

Received: 23 April, 2015; revised: 21 May, 2015; accepted: 26 May, 2015; available on-line: 22 June, 2015

INTRODUCTION

The IS6110-RFLP strain identification method (van Embden *et al.,* 1993) has been used for many years as a golden standard in studying diversity of the *Mycobacterium tuberculosis* complex strains, mainly because of high discriminatory power. Currently, the most popular and internationally recognized genotyping methods for studying molecular epidemiology of *Mycobacterium tuberculosis* complex strain outbreak are Spoligotyping (Kamerbeek *et al.,* 1997) and MIRU-VNTR (Supply *et al.,* 2001). The main advantages of these two methods are simplicity of performance and reproducibility of results between different laboratories, as well as access to international databases that contain genotyping results from different countries. This is of a particular importance in the analysis of the route of transmission of drug-resistant strains that are crossing borders between countries and continents. Since Spoligotyping is limited to the differentiation of Beijing type strains, MIRU-VNTR has a higher discriminatory power. However, in spite of many papers describing overall technical robustness, resolution power and clonal stability of the individual MIRU-VNTR loci, there is still a need for new genotyping methods that are better in reflecting variation in the genes encoding virulence features and antigenic properties of the Mtb strains. For this reason, a new genotyping method targeting the PGRS genes is presented in this study (Ramakrishnan *et al.,* 2000; Brennan *et al.,* 2001; Delgou *et al.,* 2001; Singh *et al.,* 2001; Banu *et al.,* 2002; Brennan & Delgou, 2002; Lamichhane *et al.,* 2003; Delogu *et al.,* 2004; Chaitra *et al.,* 2005; Talarico *et al.,* 2005). We hypothesized that identification of specific genotypes will be possible based on the DNA sequence variation within the PGRS genes that are targeted in our new genotyping method. Analysis of defined lengths of amplicons will allow differentiation of strains using combination of agarose and polyacrylamide gel electrophoresis.

METHODS

In our study, 27 *M. tuberculosis* complex genomes were analyzed. For collection of genome sequences, GenBank databases of the National Center for Biotechnology Information, DNA Data Bank of Japan, Wellcome Trust Sanger Institute and Broad Institute were used (Benson *et al.,* 2010). Blast, ClustalX (Larkin *et al.,* 2007) and Clone Manager 7 computer programs were applied for DNA comparison and designing of primers. Methodology presented in this work relayed on application of six novel pairs of primers designed for the PCR method specific to the PGRS regions (Table 1).

RESULTS

Three different genotyping methods were evaluated in this work, a new genotyping method (Table 2), Spoligotyping (Table 3) and MIRU-VNTR (Table 4). For the new genotyping method, six sets of primers were selected based on theoretical analysis of 27 genomic sequences of the *M. tuberculosis* complex strains (Table 1, Fig. 1). The annealing temperature calculated for the designed primer sequences is about 68°C. Initial size-polymorphism results between the analyzed strains, presented in Table 2, showed that the designed pairs of primers flanking variable regions allowed for differentiation of 16 out of 27 genotypes (Fig. 2) For MIRU-VNTR and Spoligotyping methods the numbers of genotypes identified were 22 (Fig. 3) and 18 (Fig. 4), respectively. HGDI indexes for all genotyping methods are presented in table V. In addition, genotyping analysis showed that for *M. tuberculosis* Beijing type strains CCDC5079 and CCDC5180, two different banding patterns were obtained using our new method and MIRU-VNTR technique in contrast to Spoligotyping. The first strain was sensitive to all four first line drugs while the second one was resistant (Zhang *et al.,* 2011). Also *M. africanum* with the specific size of one

✉e-mail: romkotlo@pg.gda.pl

Abbreviations: DRE-PCR, double repetitive-element-polymerase chain reaction; HGDI, Hunter–Gaston discriminatory index; IS6110-RFLP, IS6110-based restriction fragment length polymorphism; MIRU-VNTR, variable-number tandem repeats of mycobacterial interspersed repetitive units; PGRS, polymorphic GC-rich sequences; PGRS-RFLP, restriction fragment length polymorphism of polymorphic GC-rich sequences.

Table 1. List of designed PCR primers

Variable regions	Sets of PCR primers	Tm [°C]	Number of occurrence in Mtb 7199-99 genome per number of mismatches allowed in primers				
			0	1	2	3	4
I	IF 5′CGTTGCCGGCCGAGCCA	68	1	2	8T	140T	4858T
	IR 5′GCGTCGTCAACGTCACCGCC	68.5	1	1	1	14	169
II	IIF 5′TGCCACCCTGGCCGCCGTTA	71	1	1	3	45T	262T
	IIR 5′AAGGGCGCCATTCCAGCCATGAA	67.8	1	1	1	1	1
III	IIIF 5′TTGACGGCCAAGGTCACATCACC	66.3	1	1	1	1	1
	IIIR 5′GATTCGACGCCGCCACCTTG	65.4	1	1	1	4	25
IV	IVF 5′CTCAACCCCGACACCCCCG	66.7	1	1	1	14	131
	IVR 5′GCCGCCTTCGCCACCGAC	68.6	1	2	28	259	1631
V	VF 5′CGGCCTCGGCGGGATTGG	67.8	1	1	5	86	779
	VR 5′GAAACTCCGGCGGCGGTGCTAT	68.6	1	1	1	1	1
VI	VIF 5′AGCCAAGGCAACGGCGGCA	69	1	1	3	24	208
	VIR 5′GTTGCCGCCCTTACCCCCAT	66.2	1	1	3	27	155

T, means the presence of tandem repeats among repeated hybridization places of PCR primers

Table 2. Theoretical results for Spoligotyping method

Strains Spacers	Spoligotyping 1 => 43
7199_99	00111111111111111111111101000000100000111111
CCDC5079_Beijing	00000000000000000000000000000000000111111111
CCDC5180_Beijing_MDR	000000000000000000000000000000000000111111111
CDC1551	11100000000011111011110111111111000001111111
CITRI_2	11111111111111111111000011111111000011111111
EAI5	11111111111111111111001111111111000011111111
EAI5_NITR206	11111011111111111110001101111111000000111111
F11	11111111010111111111000011111111000011111111
M_africanum	10011100011111111111111111111111111111101111
H37Ra	11110111111111111111001111111111000011111111
H37Rv	111111111111111111110011111111111000011111111
Harlem	11111110111111111111111111111111100001111011
Harlem3_NITR202	11011111111111111111001111111111100000111111
KZN605.B_XDR	11111111111111111111000011110111000011110111
KZN1435_MDR	11111111111111111111000011111111000011110111
RGTB327	11011111111011111111001101111110100001101111
PanR0209	11111111000111111111000010111111000011111111
PanR0405	11111111111111111111000011111111000011111111
B1_Beijing	00000000000000000000000000000000000111111111
B2_Beijing	000000000000000000000000000000000000111111111
ZMC13-88_XDR-TB_Beijing	000000000000000000000000000000000000001111111
ZMC13-264_XDR-TB_Beijing	00000000000000000000000000000000000001111111
HGDI index = 0.9610	
Mycobacterium bovis	
AM408590_BCG Pasteur 1173P2	11011111011111101111111111111111111111100000
BX248333_AF2122/97	11011010000011101111111111111111111111100000
AP010918_BCG str. Tokyo 172	11011111011111101111111111111111111111100000
NC_016804_BCG str. Mexico	110111110111111011111111111111111111111100000
NC_020245_BCG str. Korea 1168P	11011111011111101111111111111111111111100000
HGDI index for all strains = 0.9573	

Table 3. Theoretical results for MIRU-VNTR method

Strains	MIRU 2	Mtub04	ETR C	MIRU4	MIRU40	MIRU10	MIRU16	Mtub21	MIRU 20	QUB-11b	ETR A	Mtub29	Mtub30	ETR B	MIRU 23	MIRU 24	MIRU26	MIRU 27	Mtub34	MIRU31	Mtub39	QUB-26	Qub4156	MIRU 39
7199.99	2	2	3	2	2	4	3	3	2	6	3	4	4	2	2	1	5	3	3	3	4	7	3	2
CCDC5079_Beijing	2	4	4	2	2	3	3	5	2	6	4	2	4	1	5	1	7	3	3	5	3	7	2	3
CCDC5180_Beijing_MDR	2	4	4	2	3	3	3	5	2	5	3	4	2	2	6	1	7	3	3	5	2	7	2	3
CDC1551	2	4	3	2	5	5	3	3	2	3	2	3	4	2	5	1	5	3	2	3	3	6	3	2
CITRI.2	1	3	2	2	5	4	2	3	2	1	2	4	1	2	5	1	5	3	3	2	2	6	2	2
EAI5	2	2	4	2	1	3	2	2	2	5	3	4	2	3	6	1	3	3	3	3	2	5	2	2
F11	2	4	4	2	1	3	3	4	2	4	2	4	2	2	6	1	3	3	1	3	1	7	2	2
M. africanum	2	2	5	2	3	3	2	4	2	4	7	3	4	3	4	2	4	*	3	3	4	4	3	2
H37Ra	2	2	4	2	1	7	2	2	2	5	3	4	2	3	6	1	4	3	3	3	3	3	2	2
H37Rv	2	2	4	2	1	3	2	2	2	5	3	4	2	3	6	1	3	3	3	3	2	5	2	2
Harlem	2	2	3	2	1	3	3	3	2	4	3	4	4	2	5	1	3	3	3	3	3	2	3	2
Harlem3_NITR202	2	2	4	2	3	5	2	2	2	5	3	4	2	3	6	*	5	3	3	3	2	5	2	2
KZN605.B_XDR	1	3	4	2	4	4	3	3	2	2	2	4	1	2	6	1	5	3	3	3	2	5	3	2
KZN1435_MDR	1	3	4	2	5	4	3	3	2	2	2	4	1	2	6	1	5	3	3	3	2	4	3	2
RGTB327	2	2	4	2	1	3	2	2	2	5	3	4	2	3	6	1	3	3	3	3	2	5	*	2
PanR0209	1	4	4	2	3	3	3	4	2	4	2	4	1	2	6	1	4	3	1	3	1	7	2	2
PanR0405	0	2	4	2	0	3	2	2	2	1	0	1	0	3	6	1	3	3	3	3	2	4	2	2
B1_Beijing	2	4	3	2	3	5	3	3	2	4	4	4	4	2	5	1	5	3	3	3	3	6	2	3
B2_Beijing	2	3	4	3	3	3	3	5	2	*	*	4	4	2	5	1	10	3	3	5	3	6	2	3
ZMC13-88_XDR_Beijing	2	2	4	2	1	3	2	2	2	5	3	4	2	3	6	1	3	3	3	3	2	5	2	2
ZMC13-264_XDR_Beijing	2	2	4	2	1	3	2	2	2	5	3	4	2	3	6	1	3	3	3	3	2	5	2	2
AM408590_BCG_Pasteur_1173P2	2	0	6	1	2	2	3	1	2	3	5	2	2	5	4	2	5	*	3	3	2	5	0	2
BX248333_AF2122/97	2	2	5	4	2	2	2	3	2	4	8	3	4	5	4	1	5	*	3	3	2	4	1	2
AP010918_BCG_Tokyo_172	2	0	5	2	2	2	3	1	2	3	5	2	2	5	4	2	5	*	3	3	2	4	0	2
NC_016804_BCG_Mexico	2	0	6	1	2	2	3	1	2	3	5	2	2	5	4	2	5	*	3	3	2	5	0	2
NC_020245_BCG_Korea_1168P	2	0	6	1	2	2	3	1	2	3	3	2	2	5	4	2	5	*	3	3	2	5	0	2
Alleles	154	424	577	580	802	960	1644	1955	2059	2163b	2165	2347	2401	2461	2531	2687	2996	3007	3171	3192	3690	4052	4156	4348

* not present
HGDI = 0.9785

amplicon for sets of primers IV was differentiated from other 26 strains.

DISCUSSION

Molecular biology methods used for tracking the evolution of an Mtb strain during outbreak, when passed from person to person, are very helpful in the early detection and mapping of the transmission of strains and allow for determination of whether the recurring tuberculosis is due to relapse or recurrence. In this work, a novel genotyping method of Mtb strain differentiation is presented. The method relies on a single-locus amplification reaction using PCR, and the analysis of results is based on determination of the length of amplicons. Reproducible sizes of the PCR products among 27 analyzed genomes depending on strains and relatively big differences in the length of amplicons within I, II, III and V pairs of primers, allow for precise determination of the size of PCR products based on DNA ladder standards. However, for the IV and VI pairs of primers, the polyacrylamide gel separation or sequencing techniques are recommended because of similarity of PCR band sizes.

The method proposed, like other single-locus amplification PCR techniques, will allow for detection of similar intensity of PCR products in contrast to the PCR-fingerprinting techniques that yield an unexpected number and size of bands as a result, varying from dozens to a few thousands of nucleotide base pairs (Friedman *et al.,* 1995; Kotlowski *et al.,* 2004). Ligation of oligonucleotide adapters to the enzymatically digested fragments of the genomic DNA allows

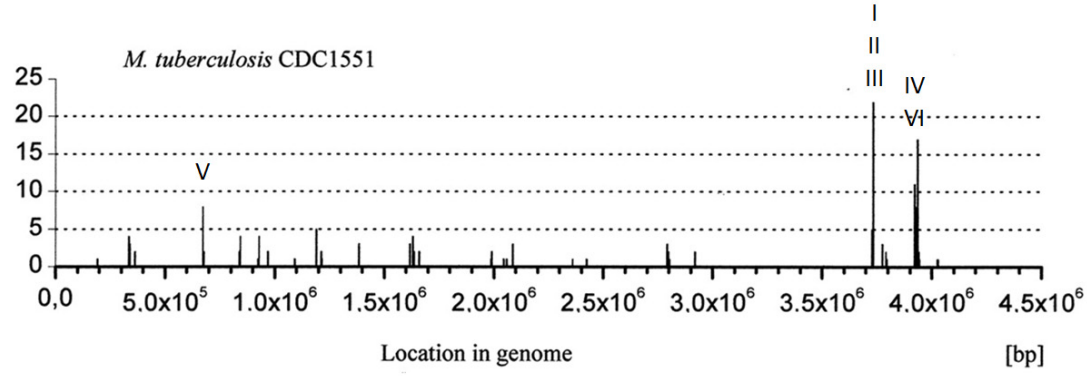

Figure 1. Histogram presenting the distribution of high-GC content Mtb2 sequence: 5′ CGG-CGG-CAA-CGG-CGG-C in the genome of *M. tuberculosis*.
Location of amplicons for each set of primers is indicated by Roman numbers.

Table 4. Results of *M. tuberculosis* and *M. africanum* genotyping using the new genotyping method

Strains Primers	I	II	III	IV	V	VI
7199.99	451	355	211	219	438	1920
CCDC5079_Beijing	433	355	133	219	438	877
CCDC5180_Beijing_MDR	433	355	211	219	354	876
CDC1551	451	355	211	210	390	1230
CITRI.2	451	169	211	219	438	894
EAI5	433	355	133	219	438	877
EAI5_NITR206	433	355	133	219	438	877
F11	451	355	211	219	438	1584
M. africanum	450	355	211	183	306	885
H37Ra	433	355	133	221	438	886
H37Rv	433	355	133	219	438	877
Harlem	451	355	211	219	438	1575
Harlem3_NITR202	433	355	133	219	438	877
KZN605.B_XDR	451	169	211	219	438	894
KZN1435_MDR	451	169	211	219	438	894
RGTB327	434	355	133	221	438	874
PanR0209	451	355	211	219	438	1584
PanR0405	433	169	133	225	438	877
B1_Beijing	465	355	211	192	306	531
B2_Beijing	459	355	211	201	306	1911
ZMC13-88_XDR-TB_Beijing	433	355	133	219	438	877
ZMC13-264_XDR-TB_Beijing	433	355	133	219	438	877
HGDI total index = 0.8918,						
Mycobacterium bovis						
AM408590_BCG Pasteur 1173P2	330	355	211	192	390	1230
BX248333_AF2122/97	330	355	211	165	438	1230
AP010918_BCG str. Tokyo 172	330	355	211	192	390	885
NC_016804_BCG str. Mexico	330	355	211	192	390	1230
NC_020245_BCG str. Korea 1168P	330	355	211	192	390	885
HGDI total index = 0.9231						

Table 5. HGDI indexes for *M. tuberculosis* complex strains

Method	HGDI index		
	M. africanum and *M. tuberculosis*	All *M. tuberculosis* complex strains	*M. tuberculosis* Beijing-type
Spoligotyping	0.961	0.9573	0.533
MIRU-VNTR	0.971	0.9785	0.933
New method	0.892	0.9231	0.800

to improve discriminatory power of the fingerprinting techniques and reproducibility of results (Goulding *et al.,* 2000), however, there are still significant differences between the intensities of PCR products in each sample which makes it difficult to identify the right number of amplicons for correct and reproducible banding pattern analysis.

The first genotyping method using high GC-rich sequences concerning PCR amplification of variable regions between *IS6110* insertion sequences and PGRS regions was DRE-PCR (Friedman *et al.,* in 1995). A comparative study using 90 *M. tuberculosis* strains has shown that the DRE-PCR method has a slightly higher discriminatory power than Spoligotyping, however, the main disadvantage of this method was reproducibility of results which equals 58% (Kremer *et al.,* 1999). Interestingly, for another PGRS-RFLP(*Alu*I) hybridization method using PGRS probe against *Alu*I digested chromosomes (Kremer *et al.,* 1999), highly reproducible results were obtained, however, in this case sample preparation procedure is very laborious and more difficult in contrast to the single-locus amplification methods like MIRU-VNTR.

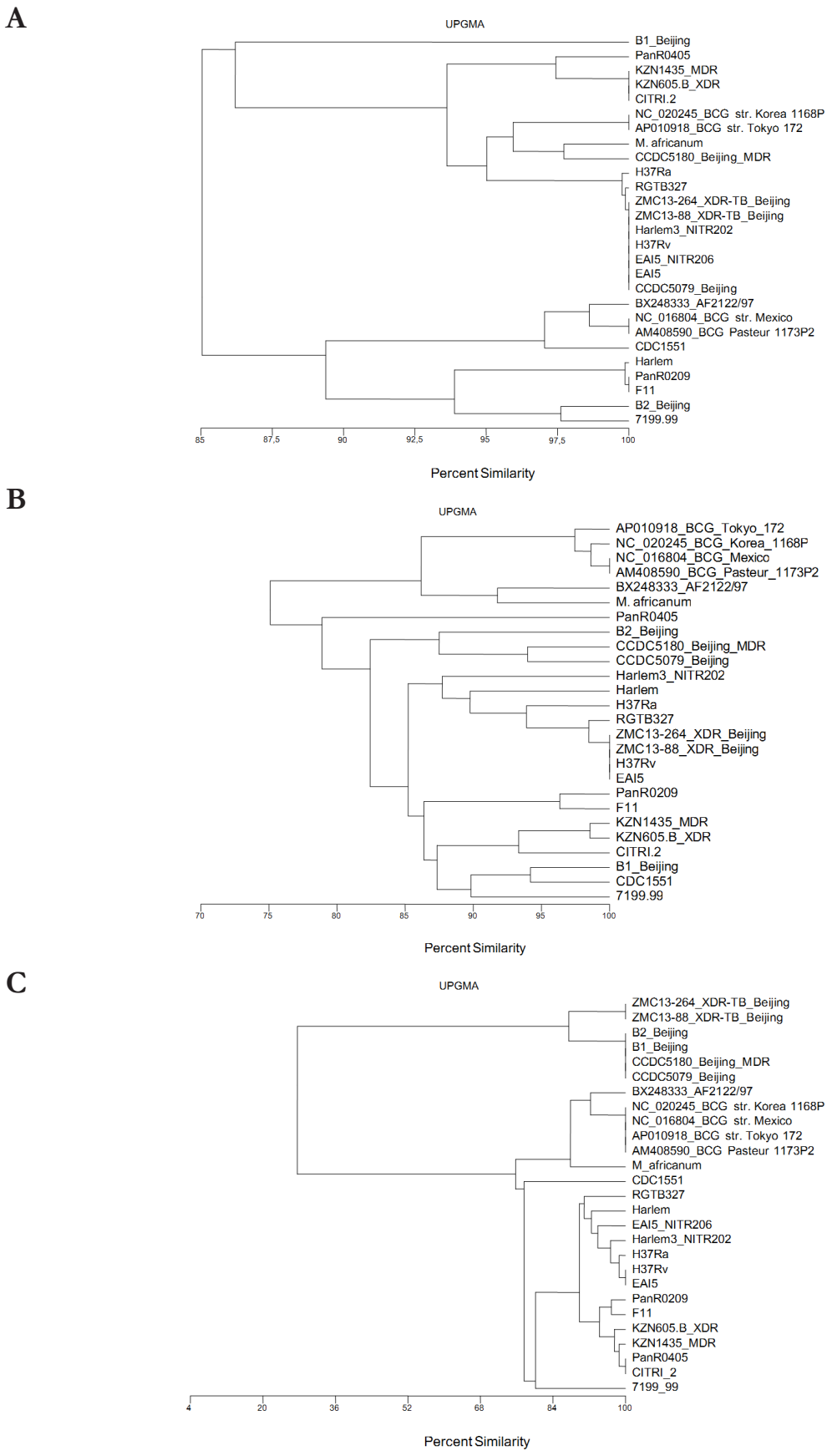

Figure 2. Differentiation of *M. tuberculosis* complex genomes using the new genotyping metod (A), MIRU-VNTR (B) and Spoligotyping (C).

The method presented in this study is the first single-locus PCR amplification technique utilizing the PGRS regions. Although the discriminatory power of the method presented is still relatively low in comparison to the MIRU-VNTR technique, the number of variable GC-rich amplicons can possibly be extended, improving the discriminatory power, after obtaining more data from sequencing results of a greater number of *Mycobacterium tuberculosis* complex genomes.

REFERENCES

van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, Small PM. **(**1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* **31**: 406–409.

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

Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Locht C (2001) Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol* **39**: 3563–3571.

Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: application to *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **100**: 7213–7218.

Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol Microbiol* **44**: 9–19.

Delogu G, Pusceddu C, Bua A, Fadda G, Brennan MJ, Zanetti S (2004) Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. *Mol Microbiol* **52**: 725–733.

Brennan MJ, Delogu G (2002) The PE multigene family: A 'molecular mantra' for mycobacteria. *Trends Microbiol* **10**: 246–249.

Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, Jacobs WR Jr (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect Immun* **69**: 7326–7333.

Delogu G, Brennan MJ (2001) Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. *Infect Immun* **69**: 5606–5611.

Singh KK, Zhang X, Patibandla AS, Chien P Jr, Laal S (2001) Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: Serological immunodominance of proteins with repetitive amino acid sequences. *Infect Immun* **69**: 4185–4191.

Talarico S, Cave MD, Marrs CF, Foxman B, Zhang L, Yang Z (2005) Variation of the *Mycobacterium tuberculosis* PE_PGRS 33 gene among clinical isolates. *J Clin Microbiol* **43**: 4954–4960.

Chaitra MG, Hariharaputran S, Chandra NR, Shaila MS, Nayak R (2005) Defining putative T cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential. *Vaccine* **23**: 1265–1272.

Ramakrishnan L, Federspiel NA, Falkow S (2000) Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* **288**: 1436–1439.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2010) GenBank. *Nucleic Acids Res* **38**: 46–51.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Zhang Y, Chen C, Liu J, Deng H, Pan A, Zhang L, Zhao X, Huang M, Lu B, Dong H, Du P, Chen W, Wan K (2011) Complete genome sequences of *Mycobacterium tuberculosis* strains CCDC5079 and CCDC5080, which belong to the Beijing family. *J Bacteriol* **193**: 5591–5592.

Friedman CR, Stoeckle MY, Johnson WD Jr, Riley LW (1995) Double-repetitive- element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *J Clin Microbiol* **33**: 1383–1384.

Kotlowski R, Shamputa IC, El Aila NA, Sajduda A, Rigouts L, van Deun A, Portaels F (2004) PCR-based genotyping of *Mycobacterium tuberculosis* with new GC-rich repeated sequences and *IS6110* inverted repeats used as primers. *J Clin Microbiol* **42**: 372–377.

Goulding JN, Stanley J, Saunders N, Arnold C (2000) Genome-sequence-base fluorescent amplified-fragment length polymorphism analysis of *Mycobacterium tuberculosis*. *J Clin Microbiol* **38**: 1121–1126.

Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martín C, Palittapongarnpim P, Plikaytis BB, Riley LW, Yakrus MA, Musser JM, van Embden JD (1999) Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* **37**: 2607–2618.