

Highly polymorphic variable-number tandem repeats loci for differentiating Beijing genotype strains of *Mycobacterium tuberculosis* in Shanghai, China

Lu Zhang^{1,2,3}, Jing Chen^{1,2,3}, Xin Shen⁴, Xiaohong Gui⁴, Jian Mei⁴, Kathryn DeRiemer⁵ & Qian Gao^{1,2,3}

¹Key Laboratory of Medical Molecular Virology, Shanghai Medical College, Fudan University, Shanghai, China; ²Institute of Medical Microbiology, Fudan University, Shanghai, China; ³Institute of Biomedical Sciences, Fudan University, Shanghai, China; ⁴Department of TB Control, Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China; and ⁵School of Medicine, University of California, Davis, CA, USA

Correspondence: Qian Gao, Key Laboratory of Medical Molecular Virology Shanghai Medical College, Fudan University, 138 Yi Xue Yuan Road, Shanghai 200032, China. Tel.: +8621 5423 7195; fax: +8621 5423 7971; e-mail: qgao99@yahoo.com, qiangaos@shmu.edu.cn

Received 7 October 2007; accepted 4 January 2008.

First published online 10 March 2008.

DOI:10.1111/j.1574-6968.2008.01081.x

Editor: Roger Buxton

Keywords

Mycobacterium tuberculosis; Beijing genotype strain; variable-number tandem repeats; IS6110 restriction fragment length polymorphism.

Abstract

The PCR-based variable-number tandem repeats (VNTR) typing method is a very promising tool for the molecular epidemiological study of *Mycobacterium tuberculosis*. The discriminatory power of the VNTR loci that were optimized in many previous studies has not been evaluated in Shanghai, an area where Beijing genotype strains dominate. In the present study, we first performed a literature search to identify VNTR loci that were at least 45 bp in length. Second, we determined the Hunter–Gaston discriminatory index (HGI) values of each of the 45 VNTR loci that we identified, using Beijing genotype strains from a ‘test set’ of isolates from a population with low migration in Chongming Island, Shanghai, China. Third, we optimized two sets of VNTR loci, which we named VNTR-7 and VNTR-16. The HGI value of VNTR-7 was slightly lower than that of IS6110 restriction fragment length polymorphisms (RFLP), and the HGI values of VNTR-16 and IS6110 RFLP were comparable. Our results suggest that VNTR-7, followed by VNTR-16 and IS6110 RFLP, can be used routinely as a tool to discriminate between *M. tuberculosis* isolates in population-based epidemiologic studies of *M. tuberculosis* transmission in Shanghai, China.

Introduction

Tuberculosis is a major health threat globally, and China is one of 22 high-burden countries for tuberculosis disease. *Mycobacterium tuberculosis* Beijing genotype strains, identified by a characteristic spoligotyping pattern with spacers 1–34 absent (Kremer *et al.*, 2004), account for more than 80% of *M. tuberculosis* isolates in many provinces of China (van Soolingen *et al.*, 1995; Shen *et al.*, 2005) and are widespread in other world regions, e.g. East Asia and countries of the former Soviet Union (Bifani *et al.*, 1996, 2002; Anh *et al.*, 2000; Prodinger *et al.*, 2001; Drobniewski *et al.*, 2002, 2005; Toungousova *et al.*, 2002; Kremer *et al.*, 2005; Surikova *et al.*, 2005a,b). It is estimated that more than one fourth of tuberculosis cases worldwide are caused by Beijing genotype strains (van Soolingen *et al.*, 1995; Bifani *et al.*, 2002). Such strains have attracted a great deal of attention because of their reported association with multiple drug resistance and

their ability to grow rapidly in human macrophages (Bifani *et al.*, 1996; Drobniewski *et al.*, 2002, 2005; Li *et al.*, 2002; Toungousova *et al.*, 2002).

Beijing genotype strains are genetically closely related, so many genotyping methods with low discriminatory power have limited ability to track their transmission (Kremer *et al.*, 2004; Iwamoto *et al.*, 2007; Wada *et al.*, 2007; Yokoyama *et al.*, 2007). Although analysis by IS6110 restriction fragment length polymorphisms (RFLP) is the gold standard for DNA fingerprinting of *M. tuberculosis*, this method is time consuming and technically demanding. Furthermore, Beijing genotype strains contain high IS6110 copy numbers and exhibit highly similar RFLP patterns, making it difficult to discriminate between different Beijing genotype strains using this method. In contrast, variable-number tandem repeats (VNTR) typing is a promising, relatively easy and rapid real-time genotyping method (Le Fleche *et al.*, 2002; Mokrousov *et al.*, 2004; Kremer

et al., 2005; Smittipat *et al.*, 2005; Kam *et al.*, 2006). However, the discriminatory power of the VNTR loci used for Beijing genotype strains in previous studies often varied, and the methods were not fully evaluated in areas where Beijing genotype strains predominate (Supply *et al.*, 2006; Cardoso Oelemann *et al.*, 2007; Iwamoto *et al.*, 2007; Yokoyama *et al.*, 2007). Furthermore, one VNTR locus exhibited different discriminatory power among Beijing genotype strains from geographically distant areas (Smittipat *et al.*, 2005; Nikolayevskyy *et al.*, 2006). The purpose of our study was to evaluate the VNTR loci used in previously published studies and to develop a set of VNTR loci with high discriminatory power for the Beijing genotype strains that occur in Shanghai, China.

Materials and methods

Bacterial collection

In Shanghai, all *M. tuberculosis* isolates from new pulmonary tuberculosis patients, who were sputum smear-positive and/or culture-positive for *M. tuberculosis*, were sent to the Shanghai Municipal Center for Disease Control and Prevention for mycobacterial identification and drug susceptibility testing. Two hundred and sixty-nine isolates from the initial samples of patients diagnosed with pulmonary tuberculosis during 2003–2005 in Chongming Island, Shanghai, China, were included in this study. Chongming Island is the third largest island in China and is located in the estuary of the Yangtze River. The island has a population of 635 000 people, mostly farmers and some migrants.

The isolates were recovered from -70°C stock and were subcultured on solid Löwenstein–Jensen medium at 37°C for 4 weeks. Two hundred and twenty-four (83.3%, 224/269) *M. tuberculosis* clinical isolates were available. Of these, 81 isolates collected in 2004 were used for determining the Hunter–Gaston discriminatory index (HGI) values of VNTR loci, and a set of 81 isolates from 2004 combined with a set of 143 isolates from 2003 and 2005 were used to further optimize the highly polymorphic VNTR typing set.

Identification of VNTR loci

Using Google and PubMed (<http://www.ncbi.nlm.nih.gov>) and the terms '*Mycobacterium tuberculosis*', 'VNTR', 'mycobacterial interspersed repetitive unit (MIRU)', we performed an Internet search to identify all VNTR loci that had a repeat unit that was at least 45 bp long. We used the cutoff of 45 bp because DNA fragments of at least 45 bp can be easily measured on an agarose gel with a 50 or 100-bp DNA ladder maker.

Genotyping methods

Spoligotyping was conducted on all *M. tuberculosis* isolates using standardized methods (Kamerbeek *et al.*, 1997).

Deletion-targeted multiplex PCR (DTM-PCR) was used to identify Beijing genotype strains (Chen *et al.*, 2007). In addition, IS6110 RFLP typing was performed according to standardized protocols (van Embden *et al.*, 1993). PCR reactions were performed using the DNA extracted from each isolate. For each of the 49 VNTR loci, we used a total volume of 10 μL . Each PCR mixture contained the following: 1 μL of DNA template, a 0.4 μM concentration of each primer (Invitrogen), 1 \times Taq PCR MasterMix (Tiangen Biotech, Beijing, China), and double-distilled H_2O , to bring the total volume to 10 μL . In addition, we supplemented the reaction mixtures of some primer pairs with 4% dimethyl sulfoxide to improve amplification. The PCR reaction conditions were as follows: an initial denaturation of 5 min at 94°C was followed by 30 cycles of denaturation at 94°C for 30 s, annealing for 30 s at temperature range $55\text{--}64^{\circ}\text{C}$, and extension at 72°C for 30 s, followed by a final extension at 72°C for 7 min. The presence and the size of PCR products were analyzed by electrophoresis in 0.7–1.5% agarose gels with the 50bp DNA ladder Marker (TaKaRa Bio Inc., Dalian, China) or the 100 bp DNA ladder Marker (Dingguo Bio Inc., Beijing, China) as the reference. Double-distilled H_2O and H37Rv were used as negative and positive controls, respectively. The number of repeats was calculated using the following formula: (length of the PCR product – length of the flanking regions)/length of one repeat unit. Digital data from VNTR typing were entered in a database (Microsoft[®] Excel, Redmond, WA).

Cluster analysis

Genotyping results were analyzed using BIONUMERICS software (version 4.0, Applied Maths Inc., Austin, TX). IS6110 RFLP patterns were analyzed as fingerprint types, and VNTR types were analyzed as character types. Similarities between VNTR types were assessed using categorical coefficients, in which all VNTR loci were weighted equally. Dendrograms were constructed according to the unpaired group method using arithmetic averages (UPGMA). IS6110 RFLP patterns were analyzed using the Dice coefficient, and similarity was calculated using UPGMA. Band tolerance levels were set at 1%. If two or more isolates showed identical VNTR profiles or IS6110 RFLP patterns, they were considered clustered.

Statistical analysis

The discriminatory power of each typing method was calculated using the HGI (Hunter & Gaston, 1988). The percentage clustering was calculated by the following formula: $(n_c - C)/N$, where N is the total number of isolates, C is the number of clusters, and n_c is the total number of clustered isolates (Small *et al.*, 1994).

Table 1. PCR primer sequences and characteristics of 49 VNTR loci identified from published studies

VNTR locus	Alias	PCR primer sequence(5'–3')	Size* (bp)	Amplicons† in H ₃₇ Rv	References‡
MIRU2		CAGGTGCCCTATCTGCTGACG GTTGCGTCCGGCATAACCAAC	189+47	283 47 × 2	1
MIRU4		GTCAAACAGGTCAACGAGAGGAA CCTCCACAATCAACACACTGGTCAT	105+77	336 77 × 3	1
MIRU10		ACCGTCTTATCGGACTGCACTATCAA CACCTTGGTGATCAGCTACCTCGAT	219+53	378 53 × 3	1
MIRU16		CGGGTCCAGTCCAAGTACCTCAAT GATCCTCCTGATTGCCCTGACCTA	367+52	471 52 × 2	1
MIRU20		CCCTTCGAGTTAGTATCGTCGGTT CAATCACC GTTACATCGACGTCATC	220+72	364 72 × 2	1
MIRU23		CGAATTCTTCGGTGGTCTCGAGT ACC GTCTGACTCATGGTGTCCAA	79+52	391 52 × 6	1
MIRU24		GAAGGCTATCCGTCGATCGGTT GGGCGAGTTGAGCTCACAGAAC	312+53	365 53 × 1	1
MIRU26		GCGGATAGGTCTACCGTCGAAATC TCCGGGTCATACAGCATGATCA	243+48	387 48 × 3	1
MIRU27		TCTGCTTGCCAGTAAGAGCCA GTGATGGTGACTTCGGTGCCTT	269+52	425 52 × 3	1
MIRU31		CGTCGAAGAGAGCCTCATCAATCAT AACCTGCTGACCGATGGCAATATC	108+52	264 52 × 3	1
MIRU39		CGGTCAAGTTCAGCACCTTCTACATC GCGTCCGTACTTCCGGTTCAG	191+47	285 47 × 2	1
MIRU40		GATTCCAACAAGACGCAGATCAAGA TCAGGTCTTTCTCACGCTCTCG	226+50	276 50 × 1	1
H37Rv-0424-51-bp	Mtub04	GTCCAGGTTGCAAGAGATGG GGCCTCCTCAACAACGGTAG	137+51	269 51 × 2+30	2
H37Rv-0577-58-bp	ETR-C	GACTTCAATGCGTTGTTGGA GTCTTGACCTCCACGAGTGC	135+58	346 58 × 3+37	2, 3
H37Rv-1443-56-bp	Mtub16	GGTAATCCTGGTCGCTTGTGTC ACCCAAATTGCCCTGGTC	235+56	291 56 × 1	2
H37Rv-1451-57-bp	QUB-1451c	GGTAGCCGTCGTCGAGAAGC CGCCACCACCGCACTGGC	88+57	305 57 × 3+46	2, 4
H37Rv-1895-57 bp	QUB-1895	GGTGCACGGCCTCGGCTCC AAGCCCCGCCGCAATCAA	80+57	319 57 × 4+11	2, 4
H37Rv-1955-57-bp	Mtub21	AGATCCCAGTTGTCGTCGTC CAACATCGCCTGGTTCTGTA	92+57	206 57 × 2	2
H37Rv-1982-78-bp	QUB-18	ATCGTCAGCTGCGGAATAGT AATACCGGGGATATCGGTTT	182+78	621 78 × 5+49	2, 5
H37Rv-2074-56-bp	Mtub24	AAATTCAAAGAGTTTCTCGACAGTG GATCTTGAGAACCAAGATGTCCTT	607+56	805 56 × 3+30	2
H37Rv-2163-a-69-bp	QUB-11a	CCCATCCCGTTAGCACATTCTGTA TTCAGGGGGGATCCGGGA	159+69	305 69 × 2+8	2, 5
H37Rv-2163-b-69-bp	QUB-11b	CGTAAGGGGGATGCGGGAAATAGG CGAAGTGAATGGTGGCAT	67+69	412 69 × 5+10	2, 5
H37Rv-2165-75-bp	ETR-A	ATTCGATCGGGATGTTGAT TCGGTCCCATCACCTTTTA	149+75	397 75 × 3+23	3
H37Rv-2347-57-bp	Mtub29	AACCCATGTCAGCCAGGTTA ATGATGGCACACCGAAGAAC	131+57	350 57 × 3+48	2
H37Rv-2401-58-bp	Mtub30	AGTCACCTTTCTACCACTCGTAAC ATTAGTAGGGCACTAGCACCTCAAG	208+58	319 58+53	2
H37Rv-2461-57-bp	ETR-B	GCGAACACCAGGACAGCATCATG GGCATGCCGGTGATCGAGTGG	107+57	292 57 × 3+14	3
H37Rv-2990-55-bp	Mtub31	GTGACGTTTACCGTGCTCTATTTT GTCGTCGGACAGTTCTAGCTTT	143+55	257 55 × 2+4	2
H37Rv-3171-54-bp	Mtub34	GCAGATAACCCGAGGAATA GGAGAGGATACGTGGATTTGAG	120+54	279 54 × 2+51	2

Table 1. Continued.

VNTR locus	Alias	PCR primer sequence(5'–3')	Size* (bp)	Amplicons [†] in H ₃₇ Rv	References [‡]
H37Rv-3232-56-bp	QUB-3232	CAGACCCGGCGTCATCAAC CCAAGGGCGGCATTGTGTT	375+56	591 56 × 3+48	2, 4
H37Rv-3239-79-bp	ETR-F	CTCGGTGATGGTCCGGCCGGTCAC GGAAGTGCTCGACAACGCCATGCC	252+79	476 79 × 2+66	2, 3
H37Rv-3336-59-bp	QUB-3336	ATCCCCGCGGTACCCATC GCCAGCGGTGTCGACTATCC	112+59	407 59 × 5+5	2, 4
H37Rv-3663-63-bp	Mtub38	GCCAAAAAGCATGGGAACGTGCCCT GGTTGTCCCCCGCAGTATCTC	114+63	373 63 × 4+7	2
H37Rv-3690-58-bp	Mtub39	AATCACGGTAACCTGGGTTGTTT GATGCATGTTCCGACCCGTAG	221+58	341 58 × 2+4	2
H37Rv-4052-111-bp	QUB-26	AACGCTCAGCTGTCGGAT GGCCAGGTCTTCCCCGAT	129+111	708 111 × 5+24	2, 5
H37Rv-4156-59-bp	QUB-4156c	TGGTCGCTACGCATCGTGTCCGCCCGT TACCACCCGGGCAGTTTAC	55+59	224 59 × 2+51	2, 4
H37Rv-0569-56-bp		TCAGCGTGTGTTGTTACCC CTCTCGCCCATACCCGA	474+56	632 56 × 2+46	6
H37Rv-0595-58-bp	<i>msx.4</i>	GCGGATGTTGATCGGGAT TCTAGACGCCAATCACGC	321+58	437 58 × 2	6
H37Rv-0917-58-bp		CTTCGACTGTTCCAGCTGAC TATCTCGAGCGCGAGGACT	311+58	436 58 × 2+9	6
H37Rv-1281-60-bp	QUB-1281	CGGACTGGGACGACTACC AGCTGGACTCTTGCGGG	172+60	292 60 × 2	6
H37Rv-1305-62-bp		TGGAGGAGACCATCTCGC CCTTGTCGAGTCGGTTGC	177+62	344 62 × 2+43	6
H37Rv-1907-56-bp		GAACGTTGGAAGAGATCAGCC TACATCGGTACGCTCTCAACG	473+56	589 56 × 2+4	6
H37Rv-2372-57-bp		ACCTCCGTTCCGATAATC CAGCTTTCAGCTCCACA	172+57	298 57 × 2+12	6
H37Rv-2687-54-bp		TATCGTTCGATCGGTTGC CTCATGCAGCGCTGACAC		563	6
H37Rv-2703-57-bp		GTACAAAATCCGACACGG GAGGTCTTCGACGTGGTA	285+57	409 57 × 2+10	6
H37Rv-3155-54-bp	QUB-15	GGTGATCTGGTCCATCGC TGACCAGGGCCCAAGACG	392+54	599 54 × 3+45	6
H37Rv-3594-56-bp		ACCAGTACGAACCAACCTGC AACCGTGAGCTGAAGGCG	305+56	498 56 × 3+25	6
H37Rv-3820-57-bp		TGCGCGGTGAATGAGACG ACCTTCATCTTGCGGAC	247+57	444 57 × 3+26	6
H37Rv-4120-57-bp		GTTACCCGGAGCCAACC GAGGTGGTTTCGTGGTCCG	310+57	447 57 × 2+23	6
H37Rv-4155-57-bp	<i>mpp.8</i>	GCAGATCGAGTTTTCCAG TTCCGGTGATGCCTGAAG	635+57	758 57 × 2+9	6

*Flanking region's length plus one repeat unit length.

[†]Description of amplicon of each VNTR locus including: amplicon length (bp) and repeat unit size (bp) × copy number + the partial repeat size (bp).

[‡]References 1, Kwara *et al.* (2003); 2, Le Fleche *et al.* (2002); 3, Frothingham & Meeker-O'Connell (1998); 4, Roring *et al.* (2002); 5, Skuce *et al.* (2002); 6, Smittipat *et al.* (2005).

Results

Identification of VNTR loci

From the Internet search, we identified 49 VNTR loci, including 12 MIRU loci (Kwara *et al.*, 2003), which had a repeat unit that was not < 45 bp long (Table 1).

Identification of Beijing genotype strains

One hundred and eighty-nine of the total 224 isolates from 2003 to 2005 (189/224, 84.4%), including 65 (65/81, 80.3%) *M. tuberculosis* isolates from 2004, were Beijing genotype strains, based on the characteristic spoligotype pattern (Kremer *et al.*, 2005) and the deletion of RD105 region by a DTM-PCR method.

Table 2. Parameters of 45 VNTR loci and their polymorphisms in isolates of *Mycobacterium tuberculosis*, Chongming Island, 2004

VNTR locus	Size range (bp)	Copy number range	Allele no.	Typable strains (n = 81)	HGI for	
					Beijing genotype (n = 65)	All strains (n = 81)
VNTR3820	361–1501	2–6, 9–17, 19–21, 25	18	80	0.8500	0.8938
QUB-11b	136–550	1–7	7	78	0.6548	0.7359
QUB-18	260–1118	1, 3–12	11	80	0.6534	0.7415
MIRU26	339–723	2, 4–10	8	81	0.6120	0.7082
QUB-11a	228–1953	1, 3, 4–8, 11–12, 25–26	11	81	0.6082	0.6927
QUB26	351–1350	2–4, 6–8, 10, 11	8	81	0.5952	0.6179
Mtub21	149–377	1–5	5	81	0.5231	0.6485
QUB-4156c	224–350	3–5	3	81	0.4923	0.4562
QUB-1895	194–365	2–5	4	81	0.4442	0.4111
Mtub04	269–494	3–5, 7	4	81	0.2971	0.4415
MIRU39	285–379	2–4	3	81	0.2856	0.4802
Mtub24	663–831	1–4	4	81	0.2755	0.2664
MIRU31	212–420	2–6	5	81	0.2461	0.4466
MIRU16	471–523	2, 3	2	81	0.2423	0.3436
ETR-F	331–568	1–4	4	81	0.2005	0.3435
MIRU10	325–378	2, 3	2	81	0.1952	0.3636
VNTR2372	229–343	1–3	3	80	0.1771	0.3820
MIRU40	326–426	2–4	3	81	0.1471	0.3488
VNTR2703	399–456	2, 3	2	78	0.0936	0.0749
VNTR4120	447–880	2–5, 10	5	80	0.0918	0.3139
Mtub30	208–440	0, 2–4	4	81	0.0909	0.2620
Mtub38	177–373	1–4	4	81	0.0904	0.3336
Mtub34	228–282	2, 3	2	81	0.0894	0.0722
MIRU20	292–364	1, 2	2	81	0.0611	0.3210
VNTR1305	239–549	1–3, 6	4	81	0.0611	0.2046
<i>mpp.8</i>	692–806	1–3	3	81	0.0611	0.1824
MIRU4	105–413	0, 2, 3, 4	4	81	0.0606	0.1855
Mtub29	245–359	2–4	3	81	0.0606	0.0728
Mtub39	279–569	1–4, 6	5	81	0.0606	0.2293
VNTR0917	369–427	1, 2	2	81	0.0606	0.0488
MIRU23	287–443	4, 5, 7	3	80	0.0606	0.1206
QUB-15	500–608	2, 4	2	79	0.0317	0.0500
MIRU27	321–477	1–4	4	81	0.0308	0.2062
ETR-A	397–449	3, 4	2	81	0.0308	0.1802
Mtub31	198–308	1–3	3	81	0.0308	0.0963
VNTR1907	529–585	1, 2	2	81	0.0308	0.0247
MIRU2	283	2	1	81	0	0
MIRU24	365	1	1	78	0	0
QUB-1451c	202–316	2–4	3	81	0	0.2219
ETR-B	164–221	1, 2	2	81	0	0.1802
VNTR0569	530–642	1–3	3	79	0	0.0978
<i>msx.4</i>	437	2	1	81	0	0
QUB-1281	292–352	2, 3	2	81	0	0.0247
VNTR2687	563		1	81	0	0
VNTR3594	361–473	1, 3	2	80	0	0.025

IS6110 RFLP typing

We were able to perform IS6110 RFLP typing on 215 (215/224, 96.0%) *M. tuberculosis* isolates, including 181 (181/215, 84.2%) of the Beijing genotype strains. Nine isolates, including eight Beijing genotype isolates, lacked sufficient

quantity and quality of chromosomal DNA for IS6110 RFLP typing. Two hundred and nine isolates had six or more IS6110 copy numbers and six non-Beijing genotype strains had fewer than six IS6110 copies. IS6110 RFLP typing differentiated 153 distinct IS6110 RFLP patterns among 181 *M. tuberculosis* Beijing genotype strains (HGI=0.9977),

Table 3. Selection of top 16 highly polymorphic VNTR loci: the cumulative HGI increases with the successive addition of each VNTR locus

Order	VNTR locus	Beijing genotype strains (<i>n</i> = 65)		All isolates (<i>n</i> = 81)	
		HGI (individual locus)	HGI (cumulative)	HGI (individual locus)	HGI (cumulative)
1	VNTR 3820	0.8500	0.8500	0.8938	0.8938
2	QUB-11b	0.6548	0.9418	0.7359	0.9617
3	QUB-18	0.6534	0.9601	0.7415	0.9735
4	MIRU26	0.6120	0.9832	0.7082	0.9883
5	QUB-11a	0.6082	0.9870	0.6927	0.9907
6	QUB-26	0.5952	0.9875	0.6179	0.9910
7	Mtub21	0.5231	0.9899	0.6485	0.9926
8	QUB-4156	0.4923	0.9899	0.4562	0.9926
9	QUB-1895	0.4442	0.9899	0.4111	0.9926
10	Mtub04	0.2971	0.9918	0.4415	0.9938
11	MIRU39	0.2856	0.9928	0.4802	0.9947
12	Mtub24	0.2755	0.9928	0.2664	0.9947
13	MIRU31	0.2461	0.9947	0.4466	0.9960
14	MIRU16	0.2423	0.9952	0.3436	0.9963
15	ETR-F	0.2005	0.9952	0.3435	0.9963
16	MIRU10	0.1952	0.9961	0.3636	0.9969
17	VNTR 2372	0.1771	0.9961	0.382	0.9969
18	MIRU40	0.1471	0.9961	0.3488	0.9969
19	VNTR 2703	0.0936	0.9961	0.0749	0.9969
20	VNTR4120	0.0918	0.9961	0.3139	0.9969

Table 4. Optimization of sets of seven VNTR loci and 16 VNTR loci (VNTR-7 and VNTR-16, respectively)

VNTR loci	Beijing genotype strains (<i>n</i> = 189)				All isolates (<i>n</i> = 224)				
	HGI (individual locus)	HGI (cumulative)	No. of types	Percentage of clustering	HGI (individual locus)	HGI (cumulative)	No. of types	Percentage of clustering	
1	VNTR3820	0.8674	0.8674	34	82.0	0.8700	0.8700	28	87.5
2	QUB-11b	0.6888	0.9270	51	73.0	0.7431	0.9469	69	69.2
3	QUB26	0.6295	0.9587	84	55.6	0.6689	0.9701	109	51.3
4	MIRU26	0.6139	0.9827	111	41.3	0.7005	0.9881	143	36.2
5	QUB-18	0.6072	0.9889	126	33.3	0.6975	0.9918	153	31.7
6	Mtub21	0.5444	0.9912	132	30.2	0.6543	0.9935	162	27.7
7	QUB-11a	0.5383	0.9944	140	25.9	0.6355	0.9957	168	25.0
8	QUB-4156c	0.4691	0.9944	141	25.4	0.4587	0.9957	168	25.0
9	QUB-1895	0.3650	0.9950	143	24.3	0.3556	0.9962	171	23.7
10	MIRU31	0.3280	0.9960	147	22.2	0.4833	0.9968	174	22.3
11	ETR-F	0.2897	0.9961	148	21.7	0.3757	0.9969	175	21.9
12	Mtub04	0.2658	0.9970	149	21.2	0.4207	0.9975	176	21.4
13	MIRU10	0.2388	0.9974	154	18.5	0.3965	0.9980	183	18.3
14	Mtub24	0.2232	0.9976	156	17.5	0.2369	0.9981	186	17.0
15	MIRU39	0.1406	0.9977	158	16.4	0.3533	0.9981	186	17.0
16	MIRU16	0.1308	0.9979	159	15.9	0.2185	0.9982	188	16.1

The cumulative HGI increases with the successive addition of each VNTR locus.

including 129 unique genotypes and 24 clusters containing a total of 52 isolates.

Optimization of VNTR loci for differentiating Beijing genotype strains

To identify optimal sets of VNTR loci to differentiate Beijing genotype strains, we first optimized the PCR reaction condi-

tions of 49 VNTR loci. Four VNTR loci (ETR-C, Mtub-16, QUB-3232, QUB-3336) were excluded because PCR products were not consistently obtained despite repeated attempts, and 45 individual VNTR loci remained (Table 2). We used the 65 Beijing genotype strains from 2004, estimated the discriminatory power of each of the 45 individual VNTR loci (Table 2), and optimized 16 highly polymorphic VNTR loci for differentiating Beijing genotype strains (Table 3).

Table 5. Comparison of different genotyping methods using *Mycobacterium tuberculosis* isolates from Chongming Island, 2003–2005

Typing methods	Beijing genotype strains (<i>n</i> = 181)			All isolates (<i>n</i> = 215)		
	HGI	No. of types	Percentage of clustering	HGI	No. of types	Percentage of clustering
IS6110 RFLP	0.9977	153	15.5	0.9980	179	16.7
VNTR (7 loci)	0.9938	135	25.4	0.9953	162	24.7
VNTR (16 loci)	0.9979	155	14.4	0.9983	183	14.9

Table 6. Identification of clusters of patients with isolates of *Mycobacterium tuberculosis* that had identical genotype patterns, based on VNTR and IS6110 RFLP, Chongming Island, 2003–2005

Discriminatory parameter	Beijing genotype strains (<i>n</i> = 181)	All isolates (<i>n</i> = 215)
IS6110 RFLP		
No. of clusters	24	32
No. (%) of clusters further differentiated by VNTR-7	3 (12.5%)	3 (9.4%)
No. (%) of cluster further differentiated by VNTR-16	9 (37.5%)	10 (31.3%)
VNTR-7		
No. of clusters	28	35
No. (%) of clusters further differentiated by VNTR-16	10 (35.7%)	11 (31.4%)
No. (%) of clusters further differentiated by RFLP	12 (42.9%)	12 (34.3%)
VNTR-16		
No. of clusters	21	27
No. (%) of clusters further differentiated by RFLP	7 (33.3%)	7 (25.9%)

IS, insertion sequence.

Of the 45 VNTR loci, VNTR3820 had the most allelic diversity and the highest discriminatory power among Beijing genotype strains. Based on the HGI value of each VNTR loci for Beijing genotype strains, we added VNTR loci in order of decreasing discriminatory power and constructed a cumulative HGI table (Table 3). The cumulative HGI value of the top seven VNTR loci was equal to that of the top nine VNTR loci (HGI = 0.9899), and the cumulative HGI value of the top 16 VNTR loci was equal to that of the top 20 VNTR loci (HGI = 0.9961) (Table 3).

We further estimated the discriminatory power of the optimized VNTR-16 and VNTR-7 loci set using 189 Beijing genotype strains during the 2003–2005 collection. The HGI values and overall hierarchies of discrimination obtained were similar to the 65 Beijing genotype strains from the 2004 collection (Tables 3 and 4).

When the VNTR-7 set was used to genotype all 181 *M. tuberculosis* Beijing genotype strains for which we had IS6110 typing data available, the differentiation (HGI = 0.9938) was slightly lower than that obtained using IS6110 RFLP (HGI = 0.9977). Using the same panel of 181 Beijing genotype strains, VNTR-16 typing was marginally more discriminatory than the IS6110 RFLP (HGI = 0.9979 vs. 0.9977) (Table 5). Therefore, two optimized sets of loci, VNTR-7 and VNTR-16, were the most parsimonious and discriminatory sets of loci among Beijing genotype strains.

Some IS6110 RFLP clusters were further subdivided by VNTR-7 or VNTR-16. Using the total available 181 Beijing genotype strains, three IS6110 RFLP clusters with two

Beijing genotype strains each were differentiated by the VNTR-7 set as six unique strains, with one repeat unit variation at a VNTR locus. Twelve (12/28, 42.9%) VNTR-7 clusters were further subdivided by IS6110 RFLP typing (Table 6). Among the Beijing genotype strains, eight IS6110 RFLP clusters with two members each and one IS6110 RFLP cluster with six members were differentiated as 17 unique strains and one cluster with five members by VNTR-16 typing with one repeat unit variation at a VNTR locus, except for two strains with a two repeat unit difference at a VNTR locus. In contrast, five VNTR-16 clusters with two members each and two VNTR-16 clusters with three members each were differentiated as two clusters with two members each and 12 unique strains by IS6110 RFLP typing with very similar IS6110 RFLP patterns.

Discussion

A VNTR locus with different copy numbers in closely related isolates should have a high evolutionary rate and high discriminatory power among distantly related isolates (Klevytska *et al.*, 2001; Denoeud & Vergnaud, 2004). *Mycobacterium tuberculosis* isolates from the Chongming Island population, with limited migration, should be more closely related than *M. tuberculosis* isolates from large cities whose populations include migrants from other regions. We assumed that the VNTR loci that were highly polymorphic among Beijing genotype strains from Chongming Island should also be highly polymorphic and help discriminate

Table 7. VNTR loci for differentiating *Mycobacterium tuberculosis* Beijing genotype strains in different published studies

VNTR locus	Patients' country of origin for the Beijing genotype strains									
	South Africa*	Russia*	Russia [†]	Russia [‡]	Hongkong [§]	Hongkong [¶]	Thailand	Japan**	Japan ^{††}	Chongming China
VNTR 3820							0.642	0.8000	0.817	0.8205
QUB-11b			0.21		0.618	0.6690		0.7716	0.763	0.6888
QUB 26			0.78		0.299	0.3144	0.449	0.7409	0.215	0.6295
MIRU 26	0.25	0.49		0.445	0.200			0.3830	0.283	0.6139
QUB 18			0.74	0.489	0.74	0.4881			0.629	0.6072
Mtub 21				0.105			0.694	0.3927	0.537	0.5444
QUB-11a				0.177	0.384	0.5136		0.6854	0.535	0.5383
QUB 4156							0.472	0.6106	0.603	0.4691
QUB 1895					0.229		0.529	0.3637	0.468	0.3650
MIRU 31		0.20		0.176				0.3215	0.379	0.3280
ETR-F							0.331		0.499	0.2897
Mtub 04								0.4587	0.581	0.2658
MIRU 10	0.52				0.377			0.4189	0.291	0.2388
Mtub 24							0.308		0.614	0.2232
MIRU 39	0.44				0.320		0.346	0.2212	0.160	0.1406
MIRU 16								0.3104	0.421	0.1308
QUB 3232				0.621		0.8041	0.844	0.8799	0.813	
VNTR2372							0.463		0.345	
VNTR4120							0.580	0.9022	0.882	
Mtub 30								0.4034	0.210	
QUB 3336								0.4870	0.482	
MIRU 40								0.3268	0.473	
Mtub 39									0.271	
QUB15									0.629	

*Mokrousov *et al.* (2004).†Surikova *et al.* (2005a, b).‡Nikolayevskyy *et al.* (2006).§Kremer *et al.* (2005).¶Kam *et al.* (2006).||Smittipat *et al.* (2005).**Iwamoto *et al.* (2007).††Yokoyama *et al.* (2007).

VNTR, variable number tandem repeat.

M. tuberculosis Beijing genotype strains in Shanghai and other provinces of mainland China.

Many previously described VNTR loci showed variations in their ability to discriminate Beijing genotype strains from geographically distant areas (Mokrousov *et al.*, 2004; Kremer *et al.*, 2005; Smittipat *et al.*, 2005; Surikova *et al.*, 2005a, b; Nikolayevskyy *et al.*, 2006; Kam *et al.*, 2006; Iwamoto *et al.*, 2007; Yokoyama *et al.*, 2007) (Table 7). Highly polymorphic VNTR loci from Hong Kong, Thailand, Japan and Russia may not be able to discriminate between Beijing genotype strains in Chongming Island. We evaluated all VNTR loci with a repeat unit of at least 45 bp in length in the present study, excluding four loci that did not give consistent, reliable results. Using Beijing genotype strains from Chongming Island, we determined the HGI value for each of 45 VNTR loci, and used the highly polymorphic individual VNTR loci to identify the most parsimonious and discriminatory sets of loci, VNTR-7 and VNTR-16. We

further compared the VNTR loci we identified in the VNTR-7 and VNTR-16 sets with the VNTR loci from published studies using Beijing genotype strains from geographically distant areas. The top seven VNTR loci in our study also had high discriminatory power among Beijing genotype strains from other countries (Table 7) and *M. tuberculosis* isolates collected all over the world (Supply *et al.*, 2006). VNTR 2372, VNTR 4120, Mtub30, MIRU40, and QUB15 showed high or moderate discriminatory power among Beijing genotype strains from Thailand or Japan (Smittipat *et al.*, 2005; Iwamoto *et al.*, 2007; Yokoyama *et al.*, 2007), but had poor discriminatory power for the isolates in the present study. The differences in the discriminatory power of different loci can be attributed to differences in the *M. tuberculosis* strains from different populations in distinct geographic areas.

Several studies have suggested that IS6110 RFLP typing overestimates recent transmission of *M. tuberculosis*, even

among high IS6110 copy number strains (Kwara *et al.*, 2003; van Deutekom *et al.*, 2005; Supply *et al.*, 2006; Cardoso Oelemann *et al.*, 2007). Van Deutekom *et al.* (2005) suggested that MIRU-VNTR typing is more reliable than IS6110 RFLP for the analysis of *M. tuberculosis* transmission. In situations with no established epidemiological link between patients with Beijing genotype strains with identical IS6110 RFLP patterns, VNTR loci sets with high discriminatory power should be used to differentiate between the strains. Owing to the slightly lower discriminatory power of VNTR-7 typing, VNTR-7 could be used as a first-line typing method followed by IS6110 RFLP and VNTR-16 to efficiently differentiate Beijing genotype strains. VNTR-16 typing had discriminatory power comparable to that of IS6110 RFLP and nine of 24 (37.5%) IS6110 RFLP clusters were further divided by VNTR-16. Therefore, VNTR-16 could be used in addition to IS6110 RFLP typing as a secondary genotyping method to differentiate Beijing genotype strains. Because we lack detailed epidemiological data for the patients in the present study, the two sets of VNTR loci used to investigate tuberculosis transmission should be further evaluated in a population-based molecular epidemiologic study. We anticipate further study of the application and usefulness of the VNTR-7 and VNTR-16 typing methods in China and other countries where Beijing genotype strains are prevalent.

Acknowledgements

This work was supported by Key Project of Chinese National Programs for Fundamental Research and Development (973 program 2005CB523102 and 2002CB512804), the Chinese National Programs 863 (2006AA02Z423), and the Shanghai Key Medical Foundation Grant (05III029).

References

- Anh DD, Borgdorff MW, Van LN, Lan NT, van Gorkom T, Kremer K & van Soolingen D (2000) *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg Infect Dis* **6**: 302–305.
- Bifani PJ, Plikaytis BB, Kapur V *et al.* (1996) Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. *JAMA* **275**: 452–457.
- Bifani PJ, Mathema B, Kurepina NE & Kreiswirth BN (2002) Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol* **10**: 45–52.
- Cardoso Oelemann M, Diel R, Vatin V, Haas W, Rüscher-Gerdes S, Loch C, Niemann S & Supply P (2007) Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiologic studies of tuberculosis. *J Clin Microbiol* **45**: 691–697.
- Chen J, Tsolaki AG, Shen X, Jiang X, Mei J & Gao Q (2007) Deletion-targeted multiplex PCR (DTM-PCR) for identification of Beijing/W genotypes of *Mycobacterium tuberculosis*. *Tuberculosis (Edinburg)* **87**: 446–449.
- Denoeud F & Vergnaud G (2004) Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. *BMC Bioinform* **5**: 4.
- Drobniewski F, Balabanova Y, Ruddy M *et al.* (2002) Rifampin- and multidrug-resistant tuberculosis in Russian civilians and prison inmates: dominance of the Beijing strain family. *Emerg Infect Dis* **8**: 1320–1326.
- Drobniewski F, Balabanova Y, Nikolayevsky V, Ruddy M, Kuznetsov S, Zakharova S, Melentyev A & Fedorin I (2005) Drug-resistant tuberculosis, clinical virulence, and the dominance of the Beijing strain family in Russia. *JAMA* **293**: 2726–2731.
- Frothingham R & Meeker-O'Connell WA (1998) Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144**: (Pt 5): 1189–1196.
- Hunter PR & Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* **26**: 2465–2466.
- Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, Tanaka N, Kawakami Y & Ito M (2007) Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. *FEMS Microbiol Lett* **270**: 67–74.
- Kam KM, Yip CW, Tse LW, Leung KL, Wong KL, Ko WM & Wong WS (2006) Optimization of variable number tandem repeat typing set for differentiating *Mycobacterium tuberculosis* strains in the Beijing family. *FEMS Microbiol Lett* **256**: 258–265.
- Kamerbeek J, Schouls L, Kolk A *et al.* (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* **35**: 907–914.
- Klevytska AM, Price LB, Schupp JM, Worsham PL, Wong J & Keim P (2001) Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* **39**: 3179–3185.
- Kremer K, Glynn JR, Lillebaek T, Niemann S, Kurepina NE, Kreiswirth BN, Bifani PJ & van Soolingen D (2004) Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. *J Clin Microbiol* **42**: 4040–4049.
- Kremer K, Au BK, Yip PC, Skuce R, Supply P, Kam KM & van Soolingen D (2005) Use of variable-number tandem-repeat typing to differentiate *Mycobacterium tuberculosis* Beijing family isolates from Hong Kong and comparison with IS6110 restriction fragment length polymorphism typing and spoligotyping. *J Clin Microbiol* **43**: 314–320.
- Kwara A, Schiro R, Cowan LS, Hyslop NE, Wiser MF, Roahen Harrison S, Kissinger P, Diem L & Crawford JT (2003)

- Evaluation of the epidemiologic utility of secondary typing methods for differentiation of *Mycobacterium tuberculosis* isolates. *J Clin Microbiol* **41**: 2683–2685.
- Le Fleche P, Fabre M, Denoëuf F, Koeck JL & Vergnaud G (2002) High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiol* **2**: 37.
- Li Q, Whalen CC, Albert JM, Larkin R, Zukowski L, Cave MD & Silver RF (2002) Differences in rate and variability of intracellular growth of a panel of *Mycobacterium tuberculosis* clinical isolates within a human monocyte model. *Infect Immun* **70**: 6489–6493.
- Mokrousov I, Narvskaya O, Limeschenko E, Vyazovaya A, Otten T & Vyshnevskiy B (2004) Analysis of the allelic diversity of the mycobacterial interspersed repetitive units in *Mycobacterium tuberculosis* strains of the Beijing family: practical implications and evolutionary considerations. *J Clin Microbiol* **42**: 2438–2444.
- Nikolayevskiy V, Gopaul K, Balabanova Y, Brown T, Fedorin I & Drobniewski F (2006) Differentiation of tuberculosis strains in a population with mainly Beijing-family strains. *Emerg Infect Dis* **12**: 1406–1413.
- Prodinger WM, Bunyaratvej P, Prachaktam R & Pavlic M (2001) *Mycobacterium tuberculosis* isolates of Beijing genotype in Thailand. *Emerg Infect Dis* **7**: 483–484.
- Roring S, Scott A, Brittain D, Walker I, Hewinson G, Neill S & Skuce R (2002) Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. *J Clin Microbiol* **40**: 2126–2133.
- Shen GM, Zha J, Xu L, Sun B, Gui XH, Wang YF, Mei J & Gao Q (2005) Evaluation of the mycobacterial interspersed repetitive units typing as a practical approach in molecular epidemiology of *Mycobacterium tuberculosis*. *Zhonghua Jie He He Hu Xi Za Zhi* **28**: 292–296.
- Skuce RA, McCorry TP, McCarroll JF, Roring SM, Scott AN, Brittain D, Hughes SL, Hewinson RG & Neill SD (2002) Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* **148**: 519–528.
- Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL & Schoolnik GK (1994) The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* **330**: 1703–1709.
- Smittipat N, Billamas P, Palittapongarnpim M, Thong-On A, Temu MM, Thanakijcharoen P, Karnkawinpong O & Palittapongarnpim P (2005) Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*. *J Clin Microbiol* **43**: 5034–5043.
- Supply P, Allix C, Lesjean S *et al.* (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* **44**: 4498–4510.
- Surikova OV, Voitykh DV, Kuz'micheva GA, Tat'kov SI, Mokrousov IV, Narvskaya OV & Filippenko ML (2005a) Differentiation of the *Mycobacterium tuberculosis* W-Beijing family strains from the Russian Federation by the VNTR-typing. *Mol Gen Mikrobiol Virusol* **3**: 22–29.
- Surikova OV, Voitech DS, Kuzmicheva G, Tatkov SI, Mokrousov IV, Narvskaya OV, Rot MA, van Soolingen D & Filipenko ML (2005b) Efficient differentiation of *Mycobacterium tuberculosis* strains of the W-Beijing family from Russia using highly polymorphic VNTR loci. *Eur J Epidemiol* **20**: 963–974.
- Toungousova OS, Sandven P, Mariandyshev AO, Nizovtseva NI, Bjune G & Caugant DA (2002) Spread of drug-resistant *Mycobacterium tuberculosis* strains of the Beijing genotype in the Archangel Oblast, Russia. *J Clin Microbiol* **40**: 1930–1937.
- van Deutekom H, Supply P, de Haas PE, Willery E, Hoijing SP, Loch C, Coutinho RA & van Soolingen D (2005) Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *J Clin Microbiol* **43**: 4473–4479.
- van Embden JD, Cave MD, Crawford JT *et al.* (1993) Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* **31**: 406–409.
- van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, Qing HZ, Enkhsaikan D, Nymadawa P & van Embden JD (1995) Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J Clin Microbiol* **33**: 3234–3238.
- Wada T, Maeda S, Hase A & Kobayashi K (2007) Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan. *J Med Microbiol* **56**: 1052–1057.
- Yokoyama E, Kishida K, Uchimura M & Ichinohe S (2007) Improved differentiation of *Mycobacterium tuberculosis* strains, including many Beijing genotype strains, using a new combination of variable number of tandem repeats loci. *Infect Genet Evol* **7**: 499–508.