# **Original Article**

# Evaluation of Four Candidate VNTR Loci for Genotyping 225 Chinese Clinical *Mycobacterium Tuberculosis* Complex Strains<sup>\*</sup>

JIANG Yi<sup>1,‡</sup>, LIU Hai Can<sup>1,‡</sup>, ZHENG Hua Jun<sup>2,‡</sup>, TANG Biao<sup>2,‡</sup>, DOU Xiang Feng<sup>3,‡</sup>, ZHAO Xiu Qin<sup>1</sup>, ZHU Yong Qiang<sup>2</sup>, LU Bing<sup>1</sup>, WANG Sheng Yue<sup>2</sup>, DONG Hai Yan<sup>1</sup>, ZHAO Guo Ping<sup>2</sup>, ZHANG Yuan Yuan<sup>1</sup>, KAN Biao<sup>1</sup>, and WAN Kang Lin<sup>1,#</sup>

 National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention/State Key Laboratory for Infectious Disease Prevention and Control, Beijing 102206, China;
 Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China;
 Beijing Center for Diseases Control and Prevention, Beijing 100013, China

# Abstract

**Objective** To evaluate four candidate variable number tandem repeat (VNTR) loci for genotyping *Mycobacterium tuberculosis* complex strains.

**Methods** Genomic sequences for two *M. tuberculosis* strains (CCDC5079 and CCDC5180) were generated, and using published sequence data, four candidate VNTR loci were identified. The VNTRs were used to genotype 225 Chinese clinical *M. tuberculosis* complex strains. The discriminatory power of the VNTRs was evaluated using BioNumerics 5.0 software.

**Results** The Hunter-Gaston Index (HGI) for BJ1, BJ2, BJ3, and BJ4 loci was 0.634, 0.917, 0.697, and 0.910, respectively. Combining all four loci gave an HGI value of 0.995, thus confirming that the genotyping had good discriminatory power. The HGI values for BJ1, BJ2, BJ3, and BJ4, obtained from Beijing family strain genotyping, were 0.447, 0.878, 0.315, and 0.850, respectively. Combining all four loci produced an HGI value of 0.988 for genotyping the Beijing family strains. We observed unique patterns for *M. bovis* and *M. africanum* strains from the four loci.

**Conclusion** We have shown that the four VNTR loci can be successfully used for genotyping *M*. *tuberculosis* complex strains. Notably, these new loci may provide additional information about Chinese *M*. *tuberculosis* isolates than that currently afforded by established VNTR loci typing.

Key words: VNTR loci; Mycobacterium tuberculosis; Genotype

Biomed Environ Sci, 2012; 25(1):82-90	doi: 10.3967/0895-3988.2	012.01.012	ISSN:0895-3988
www.besjournal.com(full text)	CN: 11-2816/Q	Copyright ©2	012 by China CDC

# INTRODUCTION

NA fingerprinting Mycobacterium tuberculosis clinical isolates is useful for determining the extent of recent transmission within a community and the potential risk factors for future transmission. The technique can also be used for identifying previously unsuspected transmission, for monitoring the transmission of drug-resistant strains, and for confirming laboratory cross contamination.

<sup>&</sup>lt;sup>\*</sup>This work was supported by the project "Transmission Mode of Tuberculosis" (2008ZX100/03-010-02) and "Warning Mode of Tuberculosis" (2008ZX10003-008) of the National Key Programme of Mega Infectious Disease. It also supported by the project "Mycobacterium tuberculosis genome SNP analysis and research on the origin of the Beijing family strains" (2011SKLID208) of State Key Laboratory for Infectious Disease Prevention and Control.

<sup>&</sup>lt;sup>#</sup>Corresponding author: WAN Kang Lin. Tel: 86-10-58900776. Fax: 86-10-58900779. E-mail:wankanglin@icdc.cn

 $<sup>^{</sup>m s}$ JIANG Yi, LIU Hai Can, ZHENG Hua Jun, TANG Biao, and DOU Xiang Feng contributed equally to this study.

Biographical note for the first author: JIANG Yi, female, born in 1978, assistant research fellow, majoring in pathogen biology.

Insertion sequence 6110-restriction fragment length polymorphism (IS6110-RFLP) typing has been the gold standard genotyping method for more than a decade. However, the method is labor-intensive and requires isolate culturing for several weeks prior to DNA extraction; it also suffers from problems associated with interpretation of the data and reliability of the complex banding patterns. In addition, it provides insufficient discrimination among isolates with low IS6110 copy numbers, an issue that is only partly overcome by using PCR-based spacer oligonucleotide typing (Spoligotyping) as a secondary screening method.

Genotyping based on the variable number tandem repeats (VNTRs) of different classes of interspersed genetic elements designated mycobacterial interspersed repetitive units (MIRUs)<sup>[1-7]</sup> is increasingly used to solve such problems. The method relies on PCR amplification of multiple loci using primers specific for the flanking sequences of each repeat locus; with this method, the amplicon size corresponds to the targeted VNTR copy number. Multiple Locus VNTR Analysis (MLVA) typing is technically flexible because the tandem repeat sizing can be determined using capillary<sup>[8-9]</sup> or gel<sup>[10]</sup> electrophoresis, or non-denaturing high-performance liquid chromatography<sup>[11]</sup>. MLVA is considerably faster than IS6110-RFLP typing, is applicable to crude DNA extracts from early stage mycobacterial cultures, and has been adapted to high-throughput conditions<sup>[12]</sup>. Moreover, the results are expressed as numerical codes that are easy to use.

Different combinations of multiple VNTR loci produce different genotype clusters that have different discriminatory powers. Currently, 12-locus VNTR<sup>[10]</sup> is the most widely used typing method and is integrated into TB control systems at a national scale. Subsequently, 15-locus<sup>[13]</sup> and 24-locus<sup>[7]</sup> sets have been proposed to allow standardization of MIRU-VNTR typing of *M. tuberculosis*. In 2008<sup>[14]</sup>, 24-locus MIRU-VNTR was used for genotyping 41 *M. tuberculosis* isolates; the results indicated that it was a useful tool for studying the molecular epidemiology of tuberculosis (TB) in populations where TB has a high incidence.

We sequenced two Chinese *M. tuberculosis* strains, namely, CCDC5079 (CP002884) and CCDC5180 (CP002885), and using published sequence data, identified four promising VNTR loci that we used to genotype 225 Chinese clinical *M. tuberculosis* complex strains. The power to discriminate *M. tuberculosis* strains was evaluated. Our results indicate that the 4 VNTR loci could prove very useful tools for genotyping *M. tuberculosis complex* strains. Furthermore, these loci may provide additional information about *M. tuberculosis* isolates than that currently obtained by established VNTR markers. Interestingly, unique patterns characteristic of *M. bovis* strains and *M. africanum* strains were also identified using the 4 loci.

# MATERIALS AND METHODS

# Strains and DNA Isolation

A total of 225 isolates were used in this study, of which 223 were M. tuberculosis strains; one strain (FJ07111) was BCG, whilst another (FJ06057) was M. africanum. All strains were randomly selected from 2354 TB strains which were genotyped by Spoligotyping in a previous study<sup>[15]</sup>. 225 strains were collected from 13 provinces, municipalities, and autonomous regions, including 12 from Anhui Province, 18 from Shanxi Province, 15 from Beijing Municipality, 34 from Fujian Province, 16 from Gansu Province. 35 from Guangxi Zhuang Autonomous Region, 12 from Sichuan Province, 17 from Henan Province, 7 from Hunan Province, 16 from Xizang (Tibet) Autonomous Region, 18 from Xinjiang Autonomous Region, 14 from Jilin Province, and 11 from Zhejiang Province. Figure 1 shows the geographical distribution of the isolates.

The strains were cultured using a standard Löwenstein-Jensen medium method, heat inactivated and then used directly in polymerase chain reactions (PCRs).



**Figure 1.** Map of China showing the distribution of the isolates used in this study (black areas represent the provinces, municipalities, and autonomous regions).

#### Identification of 4 Candidate Tandem Repeat Loci

We sequenced the genomes of the two *M. tuberculosis* strains (CCDC5079 and CCDC5180) from

Fujian Province, China. We next compared the genomic sequences of nine different M. tuberculosis complex strains using TRF software (http://tandem. bu.edu/trf/trf.html); seven of the strains that had been sequenced elsewhere were downloaded from the NCBI (http://www.ncbi.nlm.nih.gov/). Four specific VNTR loci were identified. The repeat numbers for each locus corresponding to the nine strains are shown in Table 1. The 4 loci were designated BJ1, BJ2, BJ3, and BJ4. Table 1, shows that Bovis, BCG-Pasteur and BCG-Tokyo have a single copy of BJ1 and similar copy numbers for BJ2. CCDC5079 and CCDC5180 have much higher copy numbers than the other strains for BJ2. BJ3 can distinguish H37Rv from H37Ra. High copy numbers for BJ4 are characteristic of the CCDC5079 and CCDC5180 strains. Taken together, these results confirmed that the 4 loci could be used to distinguish the Chinese strains from the other strains (Table 1). On this basis, these 4 loci were selected for genotyping the Chinese strains. The locations of each of the 4 loci in the different genomes and the primers used to amplify each locus and repeat unit of each locus are listed in the Table 2. The primers were designed using the H37Rv genomic sequence.

# PCR

PCRs were performed in 20  $\mu L$  volumes. PCR mixtures contained 10  $\mu L$  PCR GC Buffer I (BJ1, BJ2,

and BJ3) or buffer II (BJ4), 100 nmol/L of each primer, 200 µnmol/L of each of the four dNTPs and 0.5 U DNA Taq Polymerase (Takara). An initial denaturation step of 5 min at 95 °C was followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 65 °C (BJ1, BJ2, and BJ3) or 57 °C (BJ4) for 40 s and extension at 72 °C for 1 min, followed by a final extension at 72 °C for 10 min. Negative controls (reagents only, no DNA) were included each time PCR was performed. The positive control was 500 pg of DNA from *M. tuberculosis* H37Rv. Amplicon sizes were determined by agarose gel electrophoresis in Tris/boric acid/EDTA buffer followed by ethidium bromide staining. Primers for each locus are shown in Table 2. PCRs were performed in duplicate.

# **Table 1.** Copy Numbers of 4 Loci in Different Mycobacterium Tuberculosis Complex Genome

genome	BJ1	BJ2	BJ3	BJ4
CCDC5079	4	13.5	7	16
CCDC5180	3	13.5	7	14
H37Rv	2	3.5	5	3
H37Ra	2	3.5	8	3
F11	5	2.5	9	6
Bovis	1	5.5	3	8
BCG-Pasteur	1	5.5	11	7
BCG-Tokyo	1	5.5	8	6
CDC1551	1	6.5	6	3

VNTR	Position in Genome		ne	Brimors for Each Locus	Repeat	Size of PCR Products (bp)		
Locus	CCDC5079	CCDC5180	H37Rv		Unit (bp)	CCDC5079	CCDC5180	H37Rv
DI1	BJ1 2373672 2369450 23724	2222448	F : TGCCGTACTCGACCTCCGTTCCGATAA	57	339	282	225	
DJI		2572440	R : GGGCCCTGCTTGATTGTCCACCT	57				
010	BJ2 3221297 3218182 323	221207 2210102 222262	F : CCAGAACCATGCGCACTAGTTGTTG	56	980	980	420	
DJZ		5252005	R : GTTCCGCGTTCAACAGGAACTTGG					
210	2225706 222500 22	2226510	F : CGCGATCCAGGAATTGTCAGACGTT	F 0	520	520	F01	
BJ3 3325706 3322590 33	2220210	R : GCGACGATGATCGAGACGGCATTG	20	529	529	591		
DI4	DIA 2010747 2010000	47 2010000 2020400	F : CGACGTCACCTTCATCCT	59	1326	1210	573	
DJ4 3010/4/ 381890	2010900	00 3820408	R : CGAGTCGACCGATCTAGC					

# Table 2. Genomic Locations of the 4 Loci and the Repeat Unit Size of Each Locus

# **Gel Analysis**

The molecular weights of the PCR products were determined using BioNumerics5.0 software. Once the length of the PCR fragments was accurately calculated, the number of copies for each locus was deduced by comparison with the reference strain (H37Rv). PCR fragments with sizes above 1 000 bp were electrophoresed at 12 cm-length gel at 150 V for 2 h.

# Spoligotyping

Spoligotyping was performed according to a previously described method<sup>[16]</sup>. Spoligotype families were assigned using criteria described elsewhere<sup>[16]</sup>.

# **Clustering Analysis**

All data were analyzed by BioNumerics5.0 software. The discriminatory power for each of the typing methods, or for each locus was calculated

using the Hunter-Gaston index (HGI) as previously described<sup>[18]</sup>.

#### RESULTS

#### VNTR Analysis of M. Tuberculosis Complex Strains

The 4 VNTR loci described here were first identified by other groups<sup>[7,17]</sup> (Table 3), but have not been widely used for genotyping *M. tuberculosis* complex strains. In this study, the 4 loci showed high discriminatory power when used to genotype the 225 Chinese clinical *M. tuberculosis* complex strains.

Among the 225 strains tested, most strains had copy numbers of 1-4 in BJ1, a result that is consistent with other studies where the same locus was shown to have low copy numbers. However, in this study, seven of the strains had copy numbers that were higher than 4; these included one *M. africanum* strain, two U family strains, one MANU strain and two Beijing family strains.

For the BJ3 locus, 115 strains had identical copy numbers (i.e. 51.1% of all the strains). Among the 225 strains, a large number of strains had high copy numbers for BJ2 and BJ4. The highest copy numbers seen in this study were 22 (BJ4) and 31.5 (BJ2).

**Table 3.** No. of Alleles and Copy Number atEach Locus

Locus Name	Designation by Other Researchers	No. of Alleles	Copy Number
BJ1	VNTR2372	8	1-8
BJ2	VNTR3232, QUB-3232	19	0.5, 1.5-16.5, 19.5, 23.5, 31.5
BJ3	VNTR3336, QUB-3336	13	2-14
BJ4	VNTR3820	17	1-9, 11-16, 18, 19, 22

#### Typing, Clustering, and Discriminatory Power Analysis

Of the 225 strains, 165 had distinct profiles that divided them into 6 clusters by BioNumerics5.0 (Figure 2). The spoligotyping pattern of each strain is shown in Figure 2. The predominant cluster (cluster V) included 139 strains. Most of them were from the Beijing family (111). However, 13 others belonged to the U family, 8 were MANU, 4 were T strains, and the remaining 2 strains belonged to EAI and Haarlem, respectively. Another strain was found to be novel and was not, therefore, included in the SpoIDB4<sup>[19]</sup> database. The second largest cluster (number II)

consisted of 49 strains. Within this cluster, H37Rv and two H37Rv family strains were present. Cluster I contained 7 strains and each strain exhibited one spoligotyping pattern. Cluster III had 10 strains; 1 BCG (FJ07111), and 2 *M. bovis*-BCG family strains belonged to this cluster. Cluster IV included 11 strains; this cluster contained 8 Beijing family strains and one T strain. The profiles of the 4 new VNTR loci are shown in Figure 3.

The HGI results for BJ1, BJ2, BJ3, and BJ4 were 0.634, 0.917, 0.697, and 0.910, respectively. The HGI for the combination of all 4 loci was 0.995, thus suggesting that the analysis had good discriminatory power.

# Subclusters within the Beijing Family

Of the 225 *M. tuberculosis* strains, 126 were Beijing strains. All of the Beijing strains could be divided into 15 subclusters by the 4 new VNTR loci, whilst 99 non-Beijing family strains were divided into 6 subclusters. Within the Beijing family, the major subclusters consisted of 88 strains. The HGI values for BJ1, BJ2, BJ3, and BJ4 for genotyping the Beijing strains were 0.447, 0.878, 0.315, and 0.850, respectively. The HGI value of the combination of the 4 loci used in genotyping the Beijing family strains was 0.988.

# Unique Patterns among the Strains

Among the 225 isolates, we found that FJ07111 was BCG, and that FJ07113 and JL06005 belonged to the *M. bovis*-BCG family (according to spoligotyping results). These 3 strains displayed unusual patterns at these 4 loci, with values of 1.0, 5.5, 10.0, and 7.0 in BJ1, BJ2, BJ3, and BJ4, respectively. According to the Spoligotyping results, FJ06057 is an *M. africanum* strain which has a high copy number (8.0, BJ1) and a unique pattern.

#### Spoligotyping

Spoligotyping of the same set of strains showed 16 different genotypes with cluster sizes varying from 1 to 126 in the strains. Twelve of the 16 spoligotypes were found in the international SpoIDB4<sup>[18]</sup> database, and 4 patterns found in only 1 of strain each, were not present in the database and could therefore be classified as 4 new spoligotypes. The HGI for the Spoligotyping was calculated at 0.660. Hence the discriminatory power of the 4 new loci VNTR genotyping techniques was much higher than that observed using Spoligotyping.

Categorica N-VNTR **BJ1** 3J3 BJ4 3.12 Key Sp ligotyping Num FJ05395 4.0 6.5 9.0 6.0 E.I05425 3.0 6.5 9.0 3.0 s GX06044 4.0 8.0 12.5 9.0 15.0 Beijing family I FJ06057 M.africar 6.5 8.0 5.0 3.0 6.5 12.0 5.0 SC06005 ш JL06151 4.0 4.0 6.5 6.0 5.0 MANU XJ06153 CAS 4.5 12.0 5.0 2.0 4.5 GX06066 U 9.0 6.0 2.0 2.0 ShanX05260 13.5 9.0 5.0 υ FJ05349 7.5 2.0 5.0 2.0 5.0 GX06137 7.5 13.0 2.0 2.0 ShanX05092 7.5 3.0 5.0 т ShanX05125 7.5 3.0 6.0 2.0 9.5 11.0 5.0 FJ07042 Ha U 2.0 2.0 GX06027 9.5 11.0 5.0 FJ05490 υ 9.5 12.0 5.0 2.0 12.0 E.I06038 Has 9.5 5.0 2.0 2.0 9.5 11.0 2.0 GX06088 LAM ShanX05177 9.5 6.0 7.0 2.0 6.0 XJ06001 No 8.5 5.0 2.0 2.0 XJ06060 8.5 12.0 5.0 т FJ05199 На 2 8.5 11.0 5.0 2.0 FJ07033 H37Rv family 8.5 11.0 3.0 XJ06015 2.0 2.0 8.5 11.0 7.0 Beijing family XJ06025 υ 8.5 11.0 6.0 2.0 23.5 11.0 7.0 E 106025 2.0 2.0 4.5 11.0 1.0 SC06003 т II GS05040 6.5 10.0 5.0 2.0 10.0 5.0 XZ06107 LAM 12.5 2.0 2.0 GX06019 6.5 13.0 5.0 υ FJ05063 6.5 4.0 1.0 2.0 6.5 4.0 5.0 FJ05132 2.0 2.0 6.5 11.0 4.0 AH04125 s GX06087 6.5 10.0 4.0 2.0 8.5 10.0 4.0 HeN05007 s 2.0 2.0 XZ06006 2 6.5 14.0 2.0 Beijing family XZ06008 6.5 14.0 2.0 2.0 4.5 10.0 2.0 XJ06013 2.0 2.0 10.5 11.0 6.0 GX06117 υ XJ06177 υ 10.5 11.0 5.0 2.0 BJ05052 10.5 10.0 3.0 GX06187 2.0 2.0 H37Rv family 10.5 11.0 3.0 HuN06101 10.5 9.0 9.0 2.0 E.I05357 s 7.5 9.0 3.0 2.0 2.0 FJ07031 7.5 8.0 3.0 Haarle JL06183 7.5 Nev 11.0 3.0 2.0 8.0 3.0 H37Rv 3.5 2.0 2.0 1.5 9.0 3.0 GX06204 s GX06160 6.5 10.0 3.0 2.0 10.0 3.0 XJ06033 CAS 6.5 GX06014 2 2.0 6.5 11.0 3.0 SC06055 2.0 6.5 11.0 3.0 2.0 2.0 E 105554 Π. 6.5 3.0 2.0 1.0 5.5 11.0 3.0 GX06020 Т GX06043 υ 6.5 10.0 6.0 1.0 6.5 10.0 8.0 SC06002 Beijing family FJ07111 1.0 1.0 BCG 5.5 10.0 7.0 M.bovis-BCG family 2 JL06005 10.0 5.5 7.0 1.0 6.5 10.0 7.0 GX06130 Ш III 1.0 1.0 FJ05406 EAI 0.5 3.0 8.0 XJ06116 υ 8.5 5.0 6.0 1.0 XJ06183 u 9.5 5.0 6.0 1.0 4.0 ZJ06003 14.5 5.0 13.0 Beijing family XJ06018 CAS 8.5 10.0 5.0 7.0 7.5 10.0 5.0 XJ06188 CAS 4.0 4.0 7.5 10.0 3.0 XZ06003 CAS XZ06023 CAS 6.5 10.0 3.0 3.0 FJ05009 EAI 7.5 10.0 3.0 IV 3.0 4.0 10.5 10.0 3.0 GX06140 s AH03031 Beijing fa 10.5 10.0 8.0 2.0 10.0 8.0 GX06042 10.5 Ŧ 5.0 5.0 GS05138 10.5 10.0 12.0 MANU FJ05413 8.5 10.0 8.0 6.0 9.0 FJ05484 U 10.5 10.0 3.0 3.0 BJ05013 6 12.5 7.0 12.0 Beijing family XJ06106 υ 12.5 7.0 12.0 3.0 12.5 7.0 12.0 XJ06112 MANU 3.0 3.0 11.5 7.0 12.0 AH03026 Beijing family 3 XJ06087 υ 11.5 7.0 12.0 3.0 FJ05180 Beijing family 3 13.5 7.0 12.0 3.0 3.0 15.5 7.0 12.0 HeN06043 Beijing family SC06117 Beijing family 7.0 5.5 12.0 3.0 ShanX0509 Beijing family 9.5 7.0 12.0 GS05112 3.0 2.0 12.5 9.0 12.0 Beijing family BJ05015 Beijing family 12.5 7.0 12.0 3.0 13.5 6.0 12.0 E-105086 Haarlem 3.0 3.0 GS05121 10.5 8.0 12.0 Beijing family JL06035 MANU 11.5 8.0 12.0 3.0 GS05125 Beijing family 2 13.5 7.0 14.0 3.0 3.0 13.5 7.0 8.0 HeN06026 Beijing family BJ05014 Beijing family 5.5 7.0 15.0

3.0 13.5 7.0 15.0 BJ05074

3.0 16.5 7.0 15.0

3.0 8.5 7.0 15.0

Beijing family

Beijing family

Beijing family

GX06025

XZ06011

2

2



**Figure 2.** Cluster analysis of 225 strains using the 4 new VNTR loci. Six clusters are defined. The spoligotyping of each strain is shown on the right.

Biomed Environ Sci, 2012; 25(1): 82-90



Figure 3. PCR analysis of the four new polymorphic VNTR loci. M: Marker; Rv: H37Rv; 1:ShanX05111; 2:ShanX05115; 3:ShanX05118; 4:ShanX05124; 5:ShanX05125; 6:ShanX05177; 7:ShanX05178; 8:ShanX05260; 9:ShanX05290; 10:ShanX05296.

# DISCUSSION

In this study, 4 VNTR loci were identified by sequencing two Chinese TB strains (CCDC5079 and CCDC5180) and comparing their genomic sequences with published sequences from known isolates. This information allowed us to genotype 225 M. tuberculosis complex strains from 13 provinces in China. During the course of this work, it became apparent that the 4 loci had previously been identified<sup>[7,17]</sup>, but had not been recommended for use in M. tuberculosis genotyping because of the difficulties associated with interpreting the results (i.e. amplification of multiple alleles and hypervariability). However, our results demonstrate that these loci are suitable for genotyping M. tuberculosis complex strains and may also provide useful additional information about Chinese M. tuberculosis isolates. The strains genotyped in this study originated from a very large geographical area and have different spoligotyping patterns; hence the data provided by them could be representative of the type of genetic diversity that might be present within China, at least to some extent. A large number of the strains exhibited very high allele copy numbers for BJ2 and BJ4. 135 (60%) of the isolates displayed high allele numbers (i.e. >10) for BJ2, whereas 128 (56.9%) of the isolates had high allele number (i.e. >10) for BJ4. In addition, the allele numbers for BJ3 were identical in more than 50% of the strains (7.0). Nevertheless, other sequenced M. tuberculosis strains exhibit different patterns at these 4 loci; specifically H37Rv and H37Ra (Table 1). We also found that F11 and CDC1551 strains had low allele numbers at BJ2 and BJ4. Interestingly, the allele numbers for BJ3 in these 4 strains are distinct from the two Chinese strains. The data, therefore, invite speculation that the 4 loci could provide specific information about the Chinese strains that would enable researchers to determine which strains are predominant in China. Additional

isolates from other countries and areas within China should be tested to validate this idea.

To address the technical problems associated with use of the 4 loci (as reported in previous studies<sup>[7,17]</sup>), we took the following measures. We performed all of the PCRs at least twice to validate the reproducibility of the method. Our data showed that the amplicons had reproducible sizes when specific PCR conditions were used (i.e. GC buffer I/II and optimal annealing temperatures). In addition, a 12 cm-length agarose gel and 200-2000 bp markers were used to make it easier to estimate the PCR product sizes.

Among the 225 strains, FJ07111 is BCG, FJ07113, and JL06005 belong to the *M. bovis*-BCG family. We found that all of these strains had unique allele numbers at these 4 loci. Coincidently, published data shows that *M. bovis*, BCG-Pasteur and BCG-Tokyo have the same allele numbers for BJ1 and BJ2. Thus it appears likely that the *M. bovis* strains and *M. bovis*-BCG family strains have unique copy numbers for BJ1 and BJ2; this may facilitate separation of *M. bovis* strains from *M. tuberculosis* strains and differentiation of *M. bovis*-BCG family strains. In addition, the BJ1 pattern for FJ06057 has a unique high copy number of 8.0, which is a special characteristic of this *M. africanum* strain.

We found that at >0.6, the diversity index of the 4 loci provided high discriminatory power similar to that observed in previous studies<sup>[7,17]</sup>. Two hundred twenty five strains were divided into 6 clusters by the 4 VNTR loci. One BCG and 2 M. bovis-BCG family strains had same patterns at these 4 loci and all belong to cluster III (Figure 2). Furthermore, 2 H37Rv family strains and the reference H37Rv strain fall into same cluster. Hence there is agreement between the 4 loci VNTR genotyping results and Spoligotyping. Although use of the 4 loci alone for genotype M. tuberculosis complex strains would give insufficient resolution, it might still yield useful information about particular genotypes in some specific lineages. Considering the utility of these 4 loci, it could prove advantageous to combine them with 12-locus MIRU-VNTR, 15-locus MIRU-VNTR, or 24-locus MIRU-VNTR genotyping to obtain sufficient discriminatory power to study the molecular epidemiology of tuberculosis.

Previous data has shown that the Beijing strains are monophyletic within *M. tuberculosis* and that subpopulations exist within this family. By using the 4-loci VNTR to genotype the Beijing family, we were able to divide all 126 Beijing family strains into 15 subclusters with an HGI value of 0.988; this indicates that a high degree of genetic diversity exists within this family. Hence the VNTR loci have potential for subgrouping Beijing family strains. However, because we are currently lacking comparisons against other methods, the 4 loci VNTR typing method requires further evaluation before it can be applied to genotyping the Beijing family.

The published data available from NCBI shows that H37Rv and H37Ra have a different allele number for BJ3, the former being 5, the latter 8. In this study, we used H37Rv as a reference strain and sequenced PCR products derived from it. Our data showed that the copy number for BJ3 in H37Rv was 8 instead of 5. This discrepancy may due to difficulties sequencing the repeat region because of high GC content. Thus, H37Rv, and H37Ra have exactly same profile at these 4 loci, a result that is not unreasonable as the two strains were derived from a single *M. tuberculosis* strain in the 1930s.

In conclusion, 4 VNTR loci with high discriminatory power were identified in this study. These 4 loci have strong potential for genotyping *M. tuberculosis complex* strains and may provide important new information about the characteristics of *M. tuberculosis* isolates in China. In addition, the 4 loci generated unique patterns in *M. bovis* strains and *M. africanum* strains. Nevertheless, the 4 tandem repeats should be further evaluated using more strains before they can be widely used.

#### ACKNOWLEDGMENTS

We thank the staff from the 13 provinces, municipalities, and autonomous regions of China for their valuable contributions to this study.

# REFERENCES

- Gutierrez MC, Brisse S, Brosch R, et al. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog, 2005 Sep; 1(1), e5.
- Le Fleche P, Fabre M, Denoeud F, et al. High resolution, on-line identification of strains from the Mycobacterium tuberculosis complex based on tandem repeat typing. BMC Microbiol.,2002; 2, 37.
- Roring S, Scott A, Brittain D, et al. Development of variablenumber tandem repeat typing of Mycobacterium bovis: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J. Clin. Microbiol, 2002; 40, 2126-33.
- Skuce RA, McCorry TP, McCarroll JF, et al. Discrimination of Mycobacterium tuberculosis complex bacteria using novel VNTR-PCR targets. Microbiology, 2002; 148, 519-28.
- 5. Smittipat N, Palittapongarnpim P. Identification of possible loci of variable number of tandem repeats in Mycobacterium

tuberculosis. Tuber. Lung Dis, 2000; 80, 69-74.

- Supply P, Magdalena J, Himpens S, et al. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol. Microbiol, 1997; 26, 991-1003.
- Supply P, Mazars E, Lesjean S, et al. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Mol. Microbiol, 2000; 36, 762-71.
- Allix C, Supply P, Fauville-Dufaux M. Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis. Clin Infect Dis, 2004; 39, 783-9.
- Kwara A, Schiro R, Cowan LS, et al. Evaluation of the epidemiologic utility of secondary typing methods for differentiation of Mycobacterium tuberculosis isolates. J Clin Microbiol, 2003; 41, 2683-5.
- 10.Mazars E, Lesjean S, Banuls AL, et al. High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. Proc Natl Acad Sci U S A, 2001; 98, 1901-6.
- 11.Evans JT, Hawkey PM, Smith EG, et al. Automated high-throughput mycobacterial interspersed repetitive unit typing of Mycobacterium tuberculosis strains by a combination of PCR and nondenaturing high-performance liquid chromatography. J Clin Microbiol, 2004; 42, 4175-80.
- 12.Supply P, Lesjean S, Savine E, et al. Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive

units. J Clin Microbiol, 2001; 39, 3563-71.

- 13.Supply P, Allix C, Lesjean S, et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. J Clin Microbiol, 2006; 44(12), 4498-510.
- 14.Maes M, Kremer K, Soolingen D, et al. 24-Locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis, 2008; 88, 490-4.
- Dong H, Liu Z, Lv B, et al. Spoligotypes of Mycobacterium tuberculosis from different Provinces of China. J Clin Microbiol, 2010; 48(11), 4102-6.
- 16.Kamerbeek J, Schouls L, Kolk A, et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol, 1997; 35, 907-14.
- Smittipat N, Billamas P, Palittapongarnpim M, et al. Polymorphism of Variable-Number Tandem Repeats at Multiple Loci in Mycobacterium tuberculosis. J Clin Microbiol, 2005; 43(10), 5034-43.
- Hunter P, Gaston M. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol, 1988; 26, 2465-6.
- 19.Brudey K, Driscoll JR, Rigouts L, et al. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database(SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol, 2006 Mar; 6, 23.