Evaluation of Four Candidate VNTR Loci for Genotyping 225 Chinese Clinical
*Mycobacterium Tuberculosis* Complex Strains

JIANG Yi¹², LIU Hai Can¹², ZHENG Hua Jun¹², TANG Biao¹², DOU Xiang Feng³⁴, ZHAO Xiu Qin¹, ZHU Yong Qiang², LU Bing¹, WANG Sheng Yue², DONG Hai Yan¹, ZHAO Guo Ping², ZHANG Yuan Yuan¹, KAN Biao¹, and WAN Kang Lin¹,⁹

1. 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; 2. Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China; 3. 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

INTRODUCTION

DNA 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.

¹This 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.

²Corresponding author: WAN Kang Lin. Tel: 86-10-58900776. Fax: 86-10-58900779. E-mail: wankanglin@icdc.cn

³Jiang 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.

Received: August 29, 2011; Accepted: October 25, 2011
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)[1-7] 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[8-9] or gel[10] electrophoresis, or non-denaturing high-performance liquid chromatography[11]. 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.[12] 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[10] is the most widely used typing method and is integrated into TB control systems at a national scale. Subsequently, 15-locus[13] and 24-locus[7] sets have been proposed to allow standardization of MIRU-VNTR typing of M. tuberculosis. In 2008[14], 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[15]. 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 μL volumes. PCR mixtures contained 10 μL PCR GC Buffer I (BJ1, BJ2, and BJ3) or buffer II (BJ4), 100 nmol/L of each primer, 200 μmol/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 2. Genomic Locations of the 4 Loci and the Repeat Unit Size of Each Locus**

<table>
<thead>
<tr>
<th>VNTR Locus</th>
<th>Position in Genome</th>
<th>Primers for Each Locus</th>
<th>Repeat Unit (bp)</th>
<th>Size of PCR Products (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCDC5079</td>
<td>CCDC5180</td>
<td>H37Rv</td>
<td>F : TGGCGTACTCGACCTCCTCCGTCCGATAA</td>
</tr>
<tr>
<td>BJ1 2373672</td>
<td>2369450</td>
<td>2372448</td>
<td>57</td>
<td>339</td>
</tr>
<tr>
<td>BJ2 3221297</td>
<td>3218182</td>
<td>322663</td>
<td>56</td>
<td>980</td>
</tr>
<tr>
<td>BJ3 3325706</td>
<td>3322590</td>
<td>3336510</td>
<td>58</td>
<td>529</td>
</tr>
<tr>
<td>BJ4 3818747</td>
<td>3818900</td>
<td>3820408</td>
<td>59</td>
<td>1326</td>
</tr>
</tbody>
</table>

**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[16]. Spoligotype families were assigned using criteria described elsewhere[16].

**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\textsuperscript{18}.

**RESULTS**

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

The 4 VNTR loci described here were first identified by other groups\textsuperscript{7,17} (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 at Each Locus**

<table>
<thead>
<tr>
<th>Locus Name</th>
<th>Designation by Other Researchers</th>
<th>No. of Alleles</th>
<th>Copy Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>BJ1</td>
<td>VNR2372</td>
<td>8</td>
<td>1-8</td>
</tr>
<tr>
<td>BJ2</td>
<td>VNR3323, QUB-3232</td>
<td>19</td>
<td>0.5, 1.5-16.5, 19.5, 23.5, 31.5</td>
</tr>
<tr>
<td>BJ3</td>
<td>VNR3336, QUB-3336</td>
<td>13</td>
<td>2-14</td>
</tr>
<tr>
<td>BJ4</td>
<td>VNR3820</td>
<td>17</td>
<td>1-9, 11-16, 18, 19, 22</td>
</tr>
</tbody>
</table>

**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 SpolDB4\textsuperscript{19} 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 JLO6005 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. africam* 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 SpolDB4\textsuperscript{18} database, and 4 patterns found in only 1 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.
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.
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\(^7,17\), 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\[7,17\]), 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 JLO6005 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 from other M. tuberculosis 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\[7,17\]. 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

5. Smittipat N, Palittapongarnpim P. Identification of possible loci of variable number of tandem repeats in Mycobacterium


