Spoligotypes of *Mycobacterium tuberculosis* from Different Provinces of China\(^\dagger\)

Haiyan Dong,† Zhiguang Liu,‡ Bing Lv,‡ Yuanyuan Zhang,‡ Jie Liu, Xiuqin Zhao, Jinghua Liu, and Kanglin Wan*

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, China

Received 14 March 2010/Returned for modification 17 June 2010/Accepted 30 July 2010

A total of 2,346 *Mycobacterium tuberculosis* isolates from 13 provinces in China were genotyped by spoligotyping. Two hundred seventy-eight spoligotypes were identified: 2,153 isolates were grouped into 85 clusters, and the remaining 193 isolates were orphans. Comparison with the SpolDB4.0 database revealed that 118 spoligotypes had shared international type numbers in the database and the other 160 were novel. These 160 novel spoligotypes were assigned to families and subfamilies using the SpotClust program. The most prevalent family was the Beijing family (74.08%), followed by the T family (14.11%). CAS family strains were found only in the Xinjiang and Tibet regions, while EAI family strains were found only in Fujian Province. In conclusion, the present study of the *M. tuberculosis* population in China demonstrated that Beijing family isolates are the most prevalent strains in China and that they exhibit geographical variation. Furthermore, many new spoligotypes were found in this study.

Tuberculosis (TB) continues to be a major public health problem in China. Based on the data from a nationwide random survey of the epidemiology of TB in China in 2000, there were probably 4.51 million active-TB patients in the country, including 1.50 million smear-positive cases, which were the infectious sources (16). From 2006 to 2009, more than 1 million new TB cases emerged each year. Consequently, the task of controlling TB in China remains difficult.

The genotyping of *Mycobacterium tuberculosis* strains is important for TB control because it allows the detection of suspected outbreaks and the tracing of transmission chains. It is also important to monitor species diversity, as well as to identify secondary infections (4, 7, 13, 19). Insertion sequence (IS) 6110 restriction fragment length polymorphism (IS6110 RFLP) is thought of as the gold standard genotyping method for *M. tuberculosis* strain genotype identification (6, 13, 21). However, the method is time-consuming, labor-intensive, and costly. Furthermore, it is difficult to compare results between laboratories. Spacer oligonucleotide typing (spoligotyping), which is based on the analysis of polymorphisms of direct-repeat (DR) regions comprised of 36-bp DRs interspersed with 35- to 41-bp unique spacer sequences, is a good alternative to traditional IS6110 RFLP fingerprinting because of its simplicity, speed, and reliability (9, 11). Spoligotyping is useful for classifying *M. tuberculosis* strains into spoligotype families and subfamilies according to the presence or absence of spacer regions (24). Brosch et al. reported that *M. tuberculosis* can be divided into ancestral or modern strains based on *M. tuberculosis*-specific deletion 1 (TbD1) region analysis. The TbD1 region is present in ancestral *M. tuberculosis* strains but is absent from modern ones. These ancestral strains predominantly originated from endemic foci, whereas modern *M. tuberculosis* strains represent epidemic strains that were introduced into the same geographical regions more recently as a consequence of the worldwide spread of the tuberculosis epidemic (4).

Presently, an international spoligotype database, SpolDB4.0, has been established. Although the updated SpolDB4.0 version reflects the global distribution of *M. tuberculosis* spoligotypes, it contains little information regarding *M. tuberculosis* strains in China (5). In this study, we typed 2,346 *M. tuberculosis* clinical isolates from 13 different provinces across China between 2005 and 2007 using spoligotyping to study *M. tuberculosis* diversity in China.

**MATERIALS AND METHODS**

*M. tuberculosis* strains and DNA isolation. A total of 2,346 *M. tuberculosis* isolates were randomly collected between 2005 and 2007 from 2,346 patients at 13 different provincial tuberculosis hospitals across China; provincial tuberculosis hospitals are responsible for the delivery of health care to TB patients in each province (Fig. 1). Demographic, epidemiologic, and clinical information was obtained from the medical records of all patients, including sex, age, contact (family/close contact) data, as well as results of mycobacterial smears, signs, symptoms, previous TB history, present address, and associated medical data from local doctors working in TB hospitals using uniform epidemiological methods. Mycobacterial genomic DNA was extracted from mycobacterial colonies grown on Löwenstein-Jensen medium by resuspending one loopful of mycobacterial colonies in 200 μL TE buffer (10 mM Tris-Cl, 1 mM EDTA) and was incubated at 85°C for 30 min. Then, the supernatant containing the DNA was collected by centrifugation at 12,000 × g for 10 min and stored at −20°C for further use (25).

Spoligotyping. In order to investigate the population structure of the 2,346 *M. tuberculosis* strains, spoligotyping was carried out using a homemade membrane with 43 covalently bound oligonucleotides derived from the spacer sequences of *M. tuberculosis* H37Rv and *Mycobacterium bovis* BCG P3 as previously described by Kamerbeek et al. (11). Briefly, the DR region was amplified by PCR using

\(\dagger\) Corresponding author. Mailing address: National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, P.O. Box 5, Changping, Beijing 102206, People’s Republic of China. Phone and fax: 086 10 58900779. E-mail: wankanglin@icdc.cn.

‡ H.D., Z.L., B.L., and Y.Z. contributed equally to this study.

† Supplemental material for this article may be found at http://jcm.asm.org/.

\(\dagger\) Published ahead of print on 25 August 2010.
primers DRa and DRb. The amplified DNA was subsequently hybridized to a set of 43 oligonucleotide probes by reverse line blotting.

TbD1 analysis. TbD1 analysis was performed using the methods described by Brosch et al. (4). Briefly, the presence or absence of TbD1 was analyzed by 2 PCR assays using primers complementary to the sequences of the deleted region or the internal sequences of the intact region. The PCR products were analyzed by electrophoresis on 2% agarose gels. For isolates containing the TbD1 region (TbD1+/H11001) or lacking the TbD1 region (TbD1-/H11002), a PCR product was obtained with either internal or flanking primers, respectively.

Database comparison. The spoligotyping results were entered in octal and binary formats into Microsoft Excel spreadsheets; spoligotype patterns were designated as 43-character-long strings consisting of black and white squares representing the presence or the absence of an individual spacer, respectively. Spoligotype designations were determined by comparing the spoligotyping results with already existing designations in the international database SITVIT2 (Institut Pasteur de Guadeloupe), an updated version of the international spoligotyping database, SpolDB4.0 (5) (http://www.pasteur-guadeloupe.fr/spotclust.html). At the time of matching analysis, the updated SpolDB4.0 contained 39,609 patterns distributed among 2,881 shared types in 121 countries. Patterns that were not found in SpolDB4.0 were assigned to families and subfamilies using the SpotClust program (23), which was built on the SpolDB3 database (http://cgi2.cs.rpi.edu/~bennek/SPOTCLUST.html).

RESULTS

Genetic diversity and family assignment. Among the 2,346 typed isolates, 2,149 (91.60%) were classified into one of the 118 shared international types (SITs) according to SpolDB4.0 (the data are summarized in Table 1 and detailed in Table S2 in the supplemental material). The remaining 197 isolates generated 160 different spoligotypes that had not been previously described in the database (see Table S3 in the supplemental material). A total of 2,153 isolates were grouped into 85 clus-

![FIG. 1. Map of China showing the distribution of M. tuberculosis isolates included in this study (the numbers indicate the absolute number of isolates per province).](image_url)
TABLE 2. Numbers and frequencies of isolates clustered by spoligotyping

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of isolates studied</td>
<td>2,346</td>
</tr>
<tr>
<td>No. of clusters</td>
<td>85</td>
</tr>
<tr>
<td>Mean no. of isolates per cluster</td>
<td>25.33</td>
</tr>
<tr>
<td>Median no. of isolates per cluster</td>
<td>3</td>
</tr>
<tr>
<td>No. (%) of clustered isolates</td>
<td>2,153 (91.77)</td>
</tr>
<tr>
<td>No. (%) of unclustered isolates</td>
<td>193 (8.23)</td>
</tr>
</tbody>
</table>

Among the 85 isolate clusters, we found 76 minor spoligotypes (including 2 to 9 isolates) and 9 major spoligotypes (>10 isolates). Isolates ST1, ST53, ST190, ST52, ST50, ST54, ST37, ST742, and ST265 represent more than 80% of the total number of isolates in this study (Table 1).

Family assignment revealed that the most frequent strain was the Beijing family ST1 (68.41%), followed by ST53 (4.52%) of the T1 family and ST190 (1.75%) of the Beijing family. The Beijing family was the most prevalent genotype in China. These novel spoligotypes were assigned to families and subfamilies using the SpotClust program. Based on the results of these new spoligotypes, the most prevalent strains were the T family, followed by family 33 and the Beijing family. Out of all the studied isolates, the most prevalent family was the Beijing family (1,738/2,346; 74.08%), followed by the T family (14.11%) and the Haarlem family (4.48%). In addition, strains from other family, such as the CAS family, EAI family, S family, MANU2 family, and X family, were found in China. It is interesting that CAS family strains were found only in the Xinjiang and Tibet regions, while EAI family strains were found only in Fujian Province.

Geographical distribution of Beijing family strains in 13 provinces. The geographical distribution of the Beijing family strains in 13 provinces is shown in Table 3. The mean percentage of Beijing family strains was 74.08%, and percentages varied greatly among the 13 provinces (~54.50 to 92.59%). The highest prevalence was found in Beijing and its surrounding areas, while a lower prevalence was found in central and southern China. The geographical distribution of Beijing family strains provides evidence for the recent clonal expansion of this particular family of strains.

**TB1 analysis.** All isolates were analyzed for the presence or absence of the TbD1 region by PCR. Seven isolates had this region intact, whereas all other isolates lacked the region. Based on data from SpolDB4.0 and SpotClust, these 7 isolates belonged to the EAI family (5 isolates had SIT numbers, and 2 were true orphans).

**DISCUSSION**

The present study aimed to describe the genetic diversity of *M. tuberculosis* strains collected from 13 provinces in China using the spoligotyping method. Although these strains are not representative of all strains present in these areas, they provide preliminary insight into the population structure of *M. tuberculosis* spoligotypes in China. Our study identified 118 previously described (according to SpolDB4.0) spoligotypes and 160 novel spoligotypes. The most prevalent *M. tuberculosis* strain was ST1 (68.41%; Beijing family), followed by ST53 (4.52%; T1 family) and ST190 (1.75%; Beijing family). Based on the SpolDB4.0 spoligotype database and SpotClust results, the most prevalent lineage was the Beijing family (74.08%). Strains of this family are genetically closely related, have a characteristic spoligotype pattern, and have been identified in many countries worldwide (3, 8, 12) since they were first described by van Soolingen et al. in the Beijing area in 1995 (22). However, there is little available information as to whether the observations made in the Beijing area are representative of the whole country. This study may provide preliminary insight into the population structure of *M. tuberculosis* genotypes in China.

The prevalence of the Beijing genotype in the strains analyzed exhibited geographical variation ranging from 54.50% to 92.59%; the highest prevalence was found in northern China, followed by central and southern China. However, Xinjiang might represent a special case, because its particular situation might be related to geographic, climatic, or ethnic differences. Rates of infection by Beijing family strains in the regions neighboring Beijing are higher than those in the most distant regions—this supports the hypothesis that strains of the family might originate from the Beijing area. Beijing family strains are aggressively expanding clones that are found in a number of populations worldwide (8). Beijing family strains are dominant in countries neighboring China, such as Thailand, South Korea, Mongolia, and Vietnam. Beijing family strains have also been found in Europe, Africa, and the United States. Based on epidemiologic data, Beijing family strains may have a selective advantage over other strains, enabling better communication among the population (8, 14, 15). The low diversity of the highly prevalent Beijing family in this study may indicate that the Beijing family is spreading rapidly in China. It may also
reflect slower evolution of the DR region in the Beijing family strains.

In addition to the prevalent Beijing family strains, in this study we also detected strains from other families: Haarlem, CAS, EAI, LAM, X, S, and MANU2. The remaining strains belonged to the T family, which was the second most frequently occurring of all tested strains, and to 160 novel spoligotypes. Although the T family is one of the most prevalent, it remains an ill-defined family of *M. tuberculosis* that is found worldwide (17, 18). It has been suggested that strains of this family have started spreading across China; of course, this requires further testing using more extended typing of more clinical isolates across more provinces of China.

One very interesting finding of this study is that 12 of the strains tested belong to the CAS family, which was previously found primarily in India (2, 17)—these 12 CAS family strains were found only in Tibet and Xinjiang. Among the 12 CAS family strains, 3 and 9 strains originated from patients of Tibetan and Uyghur ethnicity, respectively. This finding suggests that the CAS family strains might have biogeographical specificity. Moreover, Tibet and Xinjiang share geographic borders with India, where the CAS family is dominant. It has also been suggested that strains of this family might be transported by trade, tourism, or migration from India.

The TbD1 analysis employed in this study shows that the majority of isolates (99.70%) belonged to the TbD1 modern strains, whereas only 7 strains (0.03%) were TbD1+/EAI isolates. These 7 TbD1+/EAI strains were found only in Fujian Province, which is located in southeastern China. TbD1 was specifically lacking in the mmpL6 genes of nearly all *M. tuberculosis* strains. On the other hand, TbD1 was present in all other members of the *M. tuberculosis* complex. However, the few *M. tuberculosis* strains that contained the TbD1 region were designated ancestral *M. tuberculosis* strains because they belong to a lineage that split from all other *M. tuberculosis* strains before the deletion of TbD1 occurred. EAI strains have been the focus of attention for some time due to their presumed low-virulence characteristics and links to ancestral African predecessors (1, 10, 20). The predominance of modern *M. tuberculosis* strains and the relatively poor representation of ancestral strains in China support the hypothesis that China has a relatively modern endemic focus of TB. Since the ancestral TbD1+/EAI strains were found only in Fujian Province, this suggests that they are endemic there. This finding corroborates the biogeographical specificity of most of the spoligotype clades.

In conclusion, this study provides preliminary insight into the population structure of *M. tuberculosis* circulating in China as well as the distribution of Beijing family strains in various provinces. This study will be continued and extended to other provinces in China in order to provide a better estimation of the Beijing family as a risk factor for the development of TB in China.

**ACKNOWLEDGMENTS**

We thank Lishui Zhang, Xinjun Yang, Chongxiang Tong, Li Shi, Feiying Liu, Yingcheng Qi, Qing Wang, Xuanmin Zhang, Xiaohui Cao, Yunhong Tan, Haitao Li, Xiaomeng Wang, and Jun Yang for supplying strains. We also thank Christine Pourcel, Gilles Vergnaud, Philippe Le Fleche, Christophe Sola, Dick van Soolingen, and Kristin Kremer for their helpful advice. Lastly, we thank the hospital staffs of the 13 provinces for their invaluable contributions to this study.

This work was financially supported by the Transmission Mode of Tuberculosis project of the National Key Programme of Mega Infectious Diseases (2008ZX1003-0101-02) and the Study on the Genetic Polymorphism and Strain Level Identification of *Mycobacterium tuberculosis* in China of the National Natural Science Funding of China (31771853).

**REFERENCES**


