Spoligotype Database of *Mycobacterium tuberculosis*: Biogeographic Distribution of Shared Types and Epidemiologic and Phylogenetic Perspectives

Christophe Sola,* Ingrid Filliol,* Maria Cristina Gutierrez,† Igor Mokrousov,* Véronique Vincent,† and Nalin Rastogi*

*Institut Pasteur de Guadeloupe, Pointe à Pitre, Guadeloupe; and †Centre National de Référence des Mycobactéries, Institut Pasteur, Paris, France

We give an update on the worldwide spoligotype database, which now contains 3,319 spoligotype patterns of *Mycobacterium tuberculosis* in 47 countries, with 259 shared types, i.e., identical spoligotypes shared by two or more patient isolates. The 259 shared types contained a total of 2,779 (84%) of all the isolates. Seven major genetic groups represented 37% of all clustered isolates. Two types (119 and 137) were found almost exclusively in the USA and accounted for 9% of clustered isolates. The remaining 1,517 isolates were scattered into 252 different spoligotypes. This database constitutes a tool for pattern comparison of *M. tuberculosis* clinical isolates for global epidemiologic studies and phylogenetic purposes.

The Database

Spoligotyping based on the variability of the Direct Repeat (DR) locus and analysis of a variable number of tandem DNA repeats (VNTR) of *M. tuberculosis* were performed according to the original protocols (7,8). For the construction of the database, spoligotyping results were entered into Excel spreadsheet files in chronological order, according to the availability of results from published articles and our own investigations. The database was searched regularly for new shared types, i.e., identical spoligotypes shared by two or more patient isolates. For phylogenetic reconstruction, the spoligotyping results were entered into Recognizer software of the Taxotron package (Taxolab, Institut Pasteur, Paris), as recommended (9). The “1-Jaccard” Index was calculated for each pairwise comparison of patterns (10), and the neighbor-joining algorithm was used for building trees (11).

The source of the data and its representativeness are shown in Table 1. Of 3,319 individual spoligotypes in our database, most (2,418 [73%]) were either from Europe (1,142 [34%]) or the USA (1,283 [39%]). Spoligotypes shared between the USA and Europe totaled 1,286 isolates distributed among 45 shared types (Europe, n=461; USA, n=825). A statistical
Table 1. Source of data for 3,319 spoligotypes of Mycobacterium tuberculosis used to generate the database of 259 shared types

| No. of isolates | Origina | Year | Reference |
|-----------------|----------|------|-----------|
| 3               | Peru     | 1998 | 35        |
| 18              | USA      | Unpublished | R. Frothingham |
| 105             | France   | 1997 | 36        |
| 167             | United Kingdom | 1997 | 37        |
| 296             | France   | Unpublished | This study |
| 28              | Zimbabwe | 1998 | 38        |
| 32              | Guinea-Bissau | 1999 | 25        |
| 118             | The Netherlands | 1997 | 7         |
| 68              | Various countries | 2000 | 15        |
| 58              | France   | Unpublished | J. Maisetti & B. Carbonnelle |
| 62              | Russia   | Unpublished | O. Narvskaya |
| 84              | West Africa | 1999 | 39        |
| 5               | Thailand | Unpublished | P. Palittapongarnpim |
| 14              | Romania  | 1997 | 40        |
| 17              | Brazil   | 1999 | 41        |
| 5b              | Spain    | Unpublished | S. Semper & C. Martin |
| 1,283           | USA      | 2000 | 12        |
| 1b              | United Kingdom | 1999 | 42        |
| 19              | The Netherlands | 1998 | 43        |
| 1b              | The Netherlands | 1999 | 19        |
| 69              | Far East Asia | 1995 | 44        |
| 69              | Caribbean | 1999 | 6         |
| 356             | Caribbean | Unpublished | This study |

aAlthough a potential sampling bias cannot be excluded, the sampling of isolates and their representativeness (in order of description) was as follows: Denmark, of 249 isolates described with a low copy number of IS6110 collected since 1992 (exhaustivity 98%); 24 shared types, representing 136 spoligotypes, were retained (9 other shared types, representing 49 isolates that were found exclusively in Denmark; 81,82,84,88,89,92,93,95-100); Brazil, of 17 spoligotypes out of 91 isolates from a São Paulo hospital in 1999 (unknown representativeness); Cuba, of 160 isolates typed obtained from a pool of 578 smear-positive sputa collected during 1994-1995; 157 spoligotypes described (exhaustivity 36%) were retained; Philippines, no data except for a single spoligotype available; Peru, of 29 strains isolated during 1995-1996 from the sputa of patients in Lima and Cuzco, only 3 were retained in this study since the remaining isolates shared spoligotypes with patients in Texas (12) and are included in the 1,283 Texan profiles; USA, 18 clinical isolates from the collection of R. Frothingham (representativeness unknown); France, 111 isolates from 155 hospitalized patients in Paris obtained during 1993 (patients were from three major hospitals that represented 5% of the total public hospital beds in Paris); United Kingdom, 167 isolates from all the culture-positive tuberculosis (TB) patients from three large hospitals in northeast London (without any indication of period of recruitment); France, 296 isolates sent for reference purposes during a 3-year period to the Centre National de Référence des Mycobactéries, Institut Pasteur, Paris; Zimbabwe, 28 spoligotypes obtained directly from sputum samples during a 1-month recruitment period (December 1995) of sputum-positive TB cases representing 20% of all cases; Guinea-Bissau, of 229 spoligotypes obtained from samples of 90 patients with suspected TB cases during 1989-1994, only 32 spoligotypes were fully described by the authors, and were retained for the analysis; the Netherlands, 118 isolates of unspecified representativeness from the collection of National Institute of Health (RIVM, Bilthoven); international multicenter study, 68 of 90 isolates from 38 countries representing the five continents; France, 58 isolates during a 1-year (1999) recruitment in the University Hospital of Angers; Russia, 62 isolates representing the St. Petersburg area collected during 1997-1999; West Africa, 84 isolates from Ivory Coast and around Dakar, Senegal, collected during 1994-1995; Thailand, 5 isolates from northern Thailand (unknown representativeness); Romania, 14 isolates of unknown representativeness; Brazil, 17 spoligotypes out of 91 isolates from a São Paulo hospital in 1995 (unknown representativeness); Spain, 5 multidrug-resistant isolates (unknown representativeness); USA, 1,429 clinical isolates from 1,286 isolates during 1994-1999 that are part of an ongoing population-based study in Houston, Texas; United Kingdom, a single spoligotype from ancient DNA extracted from a bone sample; the Netherlands, 19 spoligotypes obtained from paraffin-wax embedded tissue samples previously collected during 1983-1993 (unknown representativeness); the Netherlands, a single spoligotype from a previous study (unknown representativeness); Far East Asia, 69 isolates from China and Mongolia obtained during 1992-1994 (unknown representativeness); Carribbean, 425 clinical isolates from a population-based ongoing study that includes all cultures isolated in Guadeloupe, Martinique, and French Guiana since 1994 and covers a 1 million population (exhaustivity 100%). Some isolates in this pool came from patients from other countries (essentially neighboring countries such as Haiti, Dominican Republic, Brazil, Commonwealth of Dominica, Barbados, and Surinam).bDescription of a given spoligotype without precise number of isolates within this type.

description analysis was performed for the 1,286 isolates to evaluate the biogeographic specificity of the shared types and assess potential sampling bias by using a sample homogeneity test derived from the chi-square test (see below).

Results and Discussion

Description of Database

The 3,319 spoligotypes were grouped into 259 shared types containing 2,779 (84%) of the isolates and 540 (16%) orphan spoligotyping patterns (clinical isolates showing unique spoligotype; results not shown; see online graphic of database, http://www.cdc.gov/ncidod/EID/vol7no3/sola_data.htm). This gives a current total of 799 distinct spoligotype patterns in our database.

The distribution of shared types, their respective sizes, and their relative distribution in different locations (distinct countries or geographic regions) are summarized in Figure 1. The 24 most frequent shared types totaled 1,804 (65%) isolates (Figure 1A); 7 types were highly frequent, representing 1,250 (45%) isolates. The Beijing type (type 1) was most frequent and represented 18% of isolates. Two types (119 and 137), which were almost exclusively found in the USA, accounted for 9% of isolates and may be specific for American populations or outbreaks (12). Types 53 and 50 accounted for 8% and 6% of isolates and were found in 17 and 15 locations, respectively. Two other types (42 and 47)

Figure 1. Histograms derived from database (graphic online at http://www.cdc.gov/ncidod/EID/vol7no3/sola_data.htm) summarizing the distribution of shared types (A), their respective sizes (B), and their relative distribution in different locations (C).
accounted for 4% of the isolates and were found in 11 countries. The remaining isolates (n=1,517) were scattered into 235 types. Figure 1B shows the relative sizes of 259 shared types; 109 shared types (42%) contained only two patients each and 38 shared types contained only three patients each. Inversely, 24 shared types containing >20 patients totaled 1,804 (65%) isolates. Finally, the distribution “unique” versus “ubiquitous” shared types (reported in one location versus found in two or more locations) is shown in Figure 1C; 122 (47%) shared types were reported from a single location, 69 (26%) were from two locations, and 25 (10%) were from three locations. Inversely, the most ubiquitous types, in increasing order of distribution, were 33 and 37, 20, 52, 42, 50, and 53. Thus, most M. tuberculosis shared types contained a low number of patient isolates and were confined geographically, whereas a minority contained a high number of patient isolates and were highly disseminated. The finding of identical spoligotypes in distant countries may be explained either by recent or past transmission events or by phylogenetic convergence. However, the evolution of the DR locus relies on at least three independent mechanisms, namely, homologous recombination (13), replication slippage (14), and insertion sequence-mediated transposition (16-19), which does not favor a fortuitous convergence.

Geographic Distribution of Shared Types in the Database

Analysis of geographic distribution of the shared types (see online graphic of database, http://www.cdc.gov/ncidod/EID/vol7no3/sola_data.htm) permitted us to split our collection into two broad categories: those reported in a single area (n=122, Table 2) and those reported in two or more areas (n=137). In the latter category, matching analysis for 69 spoligotypes found in four broad geographic areas, namely, Africa, the Americas (North, Central and Caribbean, and South America), Europe, and Asia (Middle East, and Far East Asia), is shown in Table 3. Contrary to ubiquitous spoligotypes such as type 1, 53, and 50, which have been found in all regions, this is an attempt to define potential inter-regional and inter-continental flow of M. tuberculosis isolates so far confined to limited geographic areas. The most frequent matches were found for clusters in European countries (n=17), followed by Europe and North America (n=8), Europe and Central America and the Caribbean (n=5), and Europe and South America (n=4) (Table 3). These matches may underline both recent transcontinental transmission events and the history of TB spread in the New World through European settlers.

A total of 25 shared types were reported in three countries. Among these, 8 types were exclusively found either in Europe (types 10,22,161) or the Americas (types 5,67,70,93,130); 10 types were shared between two European countries and a country of another region (types 35,49,59,86,115,118,136,138,139,150); 5 types were shared between two countries of the Americas with a country in Europe (types 92,119,168,185,190); 1 type was shared between a European country and two African countries (type 125); and 1 type was shared between Asia, Europe, and the USA (type 124). Finally, 15 types were found in four countries; 1 type (type 41) was exclusively found in Europe and may be specific for this continent. Fourteen other types were distributed as follows: Europe + Americas, 8 types (types 3,7,19,31,40,51,137,152); Europe + Africa, 1 type (type 21); Europe + Asia + Americas, 3 types (type 8,89,167); Europe + Americas + Africa, 1 type (type 64); and Europe + Africa + Asia, 1 type (type 126). Finally, 28 types were reported in five or more countries, suggesting that these types are widespread and may constitute the ubiquitous types such as the Beijing type (type 1 in our database) or the Haarlem type (type 47). The only exception in this category was type 17, which was found in six countries in the Americas and may be specific for this region. Future population studies should focus on these ubiquitous types to better define their relative prevalence in each country.

Table 3. Total number of matches found in matching analysis of the shared types (n=69) found at two geographic locations*

| Region     | America + Asia | America + Caribbean | America + Europe | Asia + Middle East | Africa | America + Asia + Europe | Asia + Middle East |
|------------|---------------|---------------------|-----------------|-------------------|-------|-------------------------|-------------------|
| Africa     | 3a            | 2b                  | 4d              | 5a                | 0     | 0                       | 0                 |
| North America | NA+          | 6d                  | 4b              | 8b                | 0     | 1                       | 0                 |
| Central America | 2a          | 4b                  | 5k              | 0                 | 0     | 0                       | 0                 |
| South America | 3d          | 4m                  | 0               | 0                 | 0     | 0                       | 0                 |
| Europe     | 17a           | 1                   | 0               | 0                 | 0     | 0                       | 0                 |

*Indices a to n refer to the designation of the matching types. For full description of the matching shared type, see database (online graphic at http://www.cdc.gov/ncidod/EID/vol7no3/sola_data.htm). Spoligotyping data for isolates from Asia are scarce; hence, only two matches involving the Middle East and Far East were found (shared types 127 and 249, respectively). NA, not applicable (matches were searched only for shared types existing between two countries or regions; as no data were available for Canada, comparison of isolates within North America was not feasible).

Table 2. Geographic distribution of potentially specific shared types of Mycobacterium tuberculosis reported in a single location (n=122)

| Region | Country         | No. of types | Types |
|--------|-----------------|--------------|-------|
| Americas | Guadeloupe | 7 | 12,13,14,15,30,103,259 |
|        | French Guiana   | 4 | 66,76,94,96       |
|        | USA            | 46 | 192,194,197-199,201,202,205,206,208,210-217,219-235,237-239,241,243,246,248,256-258 |
| Europe | The Netherlands | 4 | 9,18,28,90 |
|        | United Kingdom | 6 | 16,23,27,38,43,100 |
|        | France         | 27 | 55,57,107-114,116,120,122,140,141,143-148,170,171,173,174,184,186 |
| Italy  |                | 9 | 155,157-160,163,165,166,169 |
| Spain  |                | 2 | 104,106 |
| Russia |                | 3 | 251,252,253 |
| Africa | Zimbabwe       | 6 | 79,82-85,87       |
|        | Guinea-Bissau  | 1 | 188               |
| Asia   | Philippines    | 1 | 69                |
|        | Mongolia       | 2 | 97, 98           |
Synopses

Biogeographic Analysis of European Versus American Spoligotypes

Several possible scenarios could account for the introduction and spread of TB in the Americas; however, documented contact with Europeans is considered too recent to account for the widespread distribution of the disease by AD 1000 (20). One hypothesis suggests that TB may have penetrated the Americas through human migration from Asia via the Bering Strait (21). Another scenario suggests TB's initial introduction as a zoonosis that became an anthroopozoonosis after cattle were domesticated (20,21). In this context, of the 259 shared types in our database, 59 were exclusively reported in the Americas, whereas 50 were found only in Europe (Table 2). This biogeographic dichotomy may signal the specific history of the disease in each continent. As enough data were present for the USA and Europe (2,418 [73%] isolates), a statistical analysis of distribution of shared types was performed. Of 45 shared types found in both USA and Europe (1,142 (11), 1,276 (12), and 3,319 individual spoligotypes reported, respectively, for Europe (p1), USA (p2), and the full database available for the world. The quotient d/p0 was due to sampling bias or the existence of two distinct populations, the percentage of individuals (p0) harboring shared-type “x” in the population studied was estimated by the equation p0=k1+k2/n1+n2=n1p1+n2p2/n1+n2. In this equation, individual sampling sizes are n1 and n2, the number of individuals within a given shared-type “x” are k1 and k2, and the representativeness for the two samples is p1=k1/n1 and p2=k2/n2.

Use of Database for Epidemiologic Studies

Essentially working in a Caribbean setting for last 6 years with systematic typing of all M. tuberculosis isolates from Guadeloupe, Martinique, and French Guiana, we initially focused on spoligotypes that may be specific to our region. Of 259 shared types, 85 types were present in the Caribbean. Of these, 69 were common to the Caribbean and the rest of the world, and 16 were reported only from the Caribbean (types 5,12,13,14,15,30,63,66,68,72,76,77,94,96,103,259). Although TB has a penchant to be latent for years with systematic typing of all drug-sensitive isolates (5,22), studies focusing on M. tuberculosis isolates from developing countries, where TB is highly prevalent, would improve understanding of the worldwide circulation of tubercle bacilli and provide insights into their epidemiology, phylogeny, and virulence.

Phylogenetic Reconstruction of M. tuberculosis

For phylogenetic analysis (23), a neighbor-joining tree was constructed by calculating the 1-Jaccard Index (10,24). This tree (Figure 2) incorporates the data for 252 M. tuberculosis shared types instead of the 259 allele types

---

Table 4. Analysis of distribution of shared types found in both USA and Europe

| Type (k1) | Europe (p1) | USA (p2) | World (p0) |
|----------|-------------|----------|------------|
| 1        | 21          | 1.8      | 326        | 25.5       | 476        | 14.4       | 15.3d |
| 2        | 6           | 0.5      | 2          | 0.5        | 27         | 0.8        | 1.6 |
| 8        | 10.9        | 0.5      | 7          | 0.5        | 18         | 0.6        | 0.9 |
| 19       | 1           | 0.1      | 23         | 1.8        | 27         | 0.8        | 4.2d |
| 20       | 8           | 0.7      | 2          | 0.2        | 20         | 0.6        | 2.1d |
| 25       | 13          | 1.1      | 3          | 0.2        | 17         | 0.5        | 2.7d |
| 26       | 22          | 1.9      | 5          | 0.4        | 28         | 0.8        | 3.6d |
| 33       | 13          | 1.1      | 10         | 0.8        | 38         | 1.2        | 0.9 |
| 34       | 6           | 0.5      | 9          | 0.7        | 21         | 0.6        | 0.6 |
| 37       | 17          | 1.5      | 2          | 0.2        | 28         | 0.8        | 3.7d |
| 44       | 12          | 1.1      | 1          | 0.1        | 15         | 0.5        | 3.3d |
| 47       | 25          | 2.2      | 23         | 1.8        | 20         | 0.7        | 0.7 |
| 48       | 34          | 3.0      | 7          | 0.5        | 41         | 1.2        | 4.6d |
| 50       | 56          | 4.9      | 32         | 2.5        | 155        | 4.7        | 3.1d |
| 52       | 29          | 2.5      | 7          | 0.5        | 40         | 1.2        | 4.0d |
| 53       | 79          | 6.9      | 46         | 3.6        | 218        | 6.6        | 3.6d |
| 54       | 8           | 0.4      | 7          | 0.5        | 17         | 0.5        | 0.7 |
| 62       | 7           | 0.6      | 4          | 0.3        | 15         | 0.5        | 1.1 |
| 92       | 2           | 0.2      | 8          | 0.6        | 14         | 0.4        | 1.7 |
| 118      | 8           | 0.7      | 1          | 0.1        | 9          | 0.3        | 2.5d |
| 119      | 2           | 0.2      | 110        | 8.6        | 115        | 3.5        | 9.6d |
| 137      | 10          | 0.9      | 134        | 10.5       | 146        | 4.4        | 9.7d |
| 138      | 5           | 1        | 1          | 0.1        | 6          | 0        | 1.8 |
| 139      | 19          | 1.7      | 19         | 1.5        | 38         | 1.2        | 0.3 |

1Results are given for 24 of 45 shared types that contained enough isolates to compare the results statistically.

2Percentages were calculated on the basis of 1,142 (n1), 1,276 (n2), and 3,319 individual spoligotypes reported, respectively, for Europe (p1), USA (p2), and the full database available for the world.

3The quotient d/p0 was calculated using the equation d=p1-p2/p0 q0/(n1+n2), where d is the absolute value of the difference between p1 and p2, q0 is the standard deviation of the repartition law of d which follows a normal distribution and can be calculated by the equation q0=√(p0 q0/n1+n02 q0/n2), and where p0 is best estimated by the equation p0=k1+k2/n1+n2. In this equation, individual sampling sizes are n1 and n2, the number of individuals within a given shared-type “x” are k1 and k2, and the representativeness for the two samples is p1=k1/n1 and p2=k2/n2.

4The absolute value of the quotient d/p0, the variations observed in the distribution of isolates for a given shared type were statistically significant and could be due to a sampling bias. Inversely, if d/p0>2, then the differences observed in the distribution of isolates for a given shared type were statistically significant and not due to a potential sample bias.

5Drug-sensitive isolates (5,22). Studies focusing on M. tuberculosis isolates from developing countries, where TB is highly prevalent, would improve understanding of the worldwide circulation of tubercle bacilli and provide insights into their epidemiology, phylogeny, and virulence.

For phylogenetic analysis (23), a neighbor-joining tree was constructed by calculating the 1-Jaccard Index (10,24). This tree (Figure 2) incorporates the data for 252 M. tuberculosis shared types instead of the 259 allele types.
described in the online database (types 253 to 259 were added recently after the completion of phylogenetic analysis). At an arbitrary distance of 0.2, one may easily distinguish nearly 15 branches that may contain significant phylogenetic information, as seen below for four selected branches (A to D) by combining results using independent genetic markers (Figure 3). As shown in Figure 2 and 3A, the homogeneous branch A (mainly present in Europe, West Africa, and South America) contains 20 types characterized by the absence of spacers 29 to 32 and 34. Such a family of isolates was recently described in Guinea-Bissau and also found to harbor a low copy number of IS6110 (25). Information concerning katG283-gyrA95 allele combination was available for 5 of these 20 types and showed that branch A belonged to the major genotypic group 1 as defined previously (26) and may represent an ancestral clone of M. tuberculosis isolates originating in Africa, Asia, or both (27; this work). For this branch, VNTR information was available for 3 of 20 types and showed a high exact tandem repeat (ETR)-A copy number (between 4 to 7; Figure 3A), which is common both for M. bovis and M. africanum (8,28).

Branch B shared a common root with branch A (Figure 2) but was clearly distinct from the population in branch A, an observation corroborated both by VNTR and katG283-gyrA95 types (Figure 3B). All the isolates in branches A and B were of the major genetic group 1, as defined (26), except for a single isolate of the major genetic group 2 in branch B (type 19); the significance of this observation is not clear. Branch C was composed of two subbranches, which are likely to be of different phylogenetic significance (Figure 3C); the upper part related to the Haarlem family, as previously defined (15), and was highly homogeneous upon VNTR typing (alleles 32333), whereas the lower part was quite heterogeneous (alleles 42431, 31333, 44553).

Finally, branch D comprised a subfamily of the spoligotypes that all missed spacers 33-36 (Figure 3D). This branch, which contained 30 different shared types, was easily

Figure 2. Phylogenetic tree of shared types of Mycobacterium tuberculosis constructed by pairwise comparison of patterns using the “1-Jaccard” index and the neighbor-joining algorithm. Approximately 15 branches may be visualized at an arbitrary distance of 0.2. The position of some reference strains (M. tuberculosis H37Rv, M. bovis BCG) or well-studied spoligotyping families of isolates (Beijing, Haarlem, and the M. africanum group) are also indicated.

Figure 3. Enlargement of branches A to D from the Mycobacterium tuberculosis phylogenetic tree (Figure 2). Numbers in standard characters refer to spoligotype numbers according to our database; those in boxes describe both the spoligotype number and variable number of tandem DNA repeats (VNTR) allele designations. Italicized numbers refer to spoligotype followed by the Houston spoligotype designation (12), and the major genetic groups 1 to 3 defined on the basis of katG283-gyrA95 allele combination (24). A and B show distinct branches belonging essentially to the major genetic group 1 with a high exact tandem repeat (ETR)-A copy number; C and D show branches that include some strains of the “Haarlem family” belonging to the major genetic group 2 with a low ETR-A copy number.
characterized by simultaneous absence of spacers 21-24 and 33-36, and constitutes a highly ramified but homogeneous family on the basis of its belonging to the major genetic group of Sreevatsan et al. (26), and the presence of two copies of the ETR-A allele upon VNTR typing. Frequently found in southern Europe and Central and South America, the ancestral type of this family (type 42) may have evolved by stepwise mutation to give, successively, types 20 and 17 (Figure 3D). This assumption is corroborated by the position of the respective types in the tree and their spoligotyping and VNTR patterns; type 42 (all spacers present except 21 to 24 and 33 to 36, VNTR 22433), type 20 (identical to type 42 plus a single missing spacer 3, VNTR identical to type 42), and type 17 (identical to type 20 plus a single missing spacer 13, VNTR 22321).

These results show that branches A and B are likely to be of an older evolutionary origin than branches C and D. Källenius et al. (25) hypothesized that branches A and B could find their evolutionary origin in West Africa, whereas branches C and D could be of European descent. However, since the global evolutionary rate of the DR locus may involve many independent mechanisms, this tree is likely to incorporate systematic yet unknown errors (6); therefore, a detailed analysis of the robustness of each potential phylogenetic link is under investigation.

Conclusion

We have presented an update of a database of

\textit{M. tuberculosis} spoligotypes with a detailed description of 259 shared types. This database may help to address major aspects linked to recent mycobacterial reemergence, evolutionary history, and future epidemiologic studies. Our results demonstrate that a few major families of conserved spoligotypes are well distributed throughout the world, whereas others are specific for certain geographic regions. Thus, the current epidemiologic picture of TB appears to be based both on the persistence of ancestral clones of \textit{M. tuberculosis} as well as those emerging more recently, e.g., the Beijing type (type 1 in our database), which also includes the MDR strain W from New York City. A future correlation between spoligotyping and VNTR typing; type 42 (all spacers present except 21 to 24 and 33 to 36, VNTR 22433), type 20 (identical to type 42 plus a single missing spacer 3, VNTR identical to type 42), and type 17 (identical to type 20 plus a single missing spacer 13, VNTR 22321).

These results show that branches A and B are likely to be of an older evolutionary origin than branches C and D. Källenius et al. (25) hypothesized that branches A and B could find their evolutionary origin in West Africa, whereas branches C and D could be of European descent. However, since the global evolutionary rate of the DR locus may involve many independent mechanisms, this tree is likely to incorporate systematic yet unknown errors (6); therefore, a detailed analysis of the robustness of each potential phylogenetic link is under investigation.

Conclusion

We have presented an update of a database of

\textit{M. tuberculosis} spoligotypes with a detailed description of 259 shared types. This database may help to address major aspects linked to recent mycobacterial reemergence, evolutionary history, and future epidemiologic studies. Our results demonstrate that a few major families of conserved spoligotypes are well distributed throughout the world, whereas others are specific for certain geographic regions. Thus, the current epidemiologic picture of TB appears to be based both on the persistence of ancestral clones of \textit{M. tuberculosis} as well as those emerging more recently, e.g., the Beijing type (type 1 in our database), which also includes the MDR strain W from New York City. A future correlation between spoligotyping and VNTR typing; type 42 (all spacers present except 21 to 24 and 33 to 36, VNTR 22433), type 20 (identical to type 42 plus a single missing spacer 3, VNTR identical to type 42), and type 17 (identical to type 20 plus a single missing spacer 13, VNTR 22321).

These results show that branches A and B are likely to be of an older evolutionary origin than branches C and D. Källenius et al. (25) hypothesized that branches A and B could find their evolutionary origin in West Africa, whereas branches C and D could be of European descent. However, since the global evolutionary rate of the DR locus may involve many independent mechanisms, this tree is likely to incorporate systematic yet unknown errors (6); therefore, a detailed analysis of the robustness of each potential phylogenetic link is under investigation.

Conclusion

We have presented an update of a database of

\textit{M. tuberculosis} spoligotypes with a detailed description of 259 shared types. This database may help to address major aspects linked to recent mycobacterial reemergence, evolutionary history, and future epidemiologic studies. Our results demonstrate that a few major families of conserved spoligotypes are well distributed throughout the world, whereas others are specific for certain geographic regions. Thus, the current epidemiologic picture of TB appears to be based both on the persistence of ancestral clones of \textit{M. tuberculosis} as well as those emerging more recently, e.g., the Beijing type (type 1 in our database), which also includes the MDR strain W from New York City. A future correlation between spoligotyping and VNTR typing; type 42 (all spacers present except 21 to 24 and 33 to 36, VNTR 22433), type 20 (identical to type 42 plus a single missing spacer 3, VNTR identical to type 42), and type 17 (identical to type 20 plus a single missing spacer 13, VNTR 22321).

These results show that branches A and B are likely to be of an older evolutionary origin than branches C and D. Källenius et al. (25) hypothesized that branches A and B could find their evolutionary origin in West Africa, whereas branches C and D could be of European descent. However, since the global evolutionary rate of the DR locus may involve many independent mechanisms, this tree is likely to incorporate systematic yet unknown errors (6); therefore, a detailed analysis of the robustness of each potential phylogenetic link is under investigation.

Conclusion

We have presented an update of a database of

\textit{M. tuberculosis} spoligotypes with a detailed description of 259 shared types. This database may help to address major aspects linked to recent mycobacterial reemergence, evolutionary history, and future epidemiologic studies. Our results demonstrate that a few major families of conserved spoligotypes are well distributed throughout the world, whereas others are specific for certain geographic regions. Thus, the current epidemiologic picture of TB appears to be based both on the persistence of ancestral clones of \textit{M. tuberculosis} as well as those emerging more recently, e.g., the Beijing type (type 1 in our database), which also includes the MDR strain W from New York City. A future correlation between spoligotyping and VNTR typing; type 42 (all spacers present except 21 to 24 and 33 to 36, VNTR 22433), type 20 (identical to type 42 plus a single missing spacer 3, VNTR identical to type 42), and type 17 (identical to type 20 plus a single missing spacer 13, VNTR 22321).

These results show that branches A and B are likely to be of an older evolutionary origin than branches C and D. Källenius et al. (25) hypothesized that branches A and B could find their evolutionary origin in West Africa, whereas branches C and D could be of European descent. However, since the global evolutionary rate of the DR locus may involve many independent mechanisms, this tree is likely to incorporate systematic yet unknown errors (6); therefore, a detailed analysis of the robustness of each potential phylogenetic link is under investigation.

Conclusion

We have presented an update of a database of

\textit{M. tuberculosis} spoligotypes with a detailed description of 259 shared types. This database may help to address major aspects linked to recent mycobacterial reemergence, evolutionary history, and future epidemiologic studies. Our results demonstrate that a few major families of conserved spoligotypes are well distributed throughout the world, whereas others are specific for certain geographic regions. Thus, the current epidemiologic picture of TB appears to be based both on the persistence of ancestral clones of \textit{M. tuberculosis} as well as those emerging more recently, e.g., the Beijing type (type 1 in our database), which also includes the MDR strain W from New York City. A future correlation between spoligotyping and VNTR typing; type 42 (all spacers present except 21 to 24 and 33 to 36, VNTR 22433), type 20 (identical to type 42 plus a single missing spacer 3, VNTR identical to type 42), and type 17 (identical to type 20 plus a single missing spacer 13, VNTR 22321).

These results show that branches A and B are likely to be of an older evolutionary origin than branches C and D. Källenius et al. (25) hypothesized that branches A and B could find their evolutionary origin in West Africa, whereas branches C and D could be of European descent. However, since the global evolutionary rate of the DR locus may involve many independent mechanisms, this tree is likely to incorporate systematic yet unknown errors (6); therefore, a detailed analysis of the robustness of each potential phylogenetic link is under investigation.
