Analysis of the Population Structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and the Netherlands: Usefulness of DNA Typing for Global Tuberculosis Epidemiology

Peter W. M. Hermans, Feriele Messadi, Haimanot Guebrexabher, Dick van Sooijingen, Petra E. W. de Haas, Herre Heersma, Han de Neeling, Abdelkader Ayoub, Françoise Portaels, Dominique Frommel, Mohamed Zribi, and Jan D. A. van Embden

The genetic heterogeneity among *Mycobacterium tuberculosis* isolates from 501 patients in Ethiopia, Tunisia, and the Netherlands was compared by analysis of DNA polymorphism driven by insertion element IS6110. The percentage of isolates displaying two or more identical patterns differed greatly in the three countries: It was highest among Tunisian isolates and lowest in Dutch isolates. In contrast to isolates from Dutch subjects infected with *M. tuberculosis*, the majority of strains from Ethiopia and Tunisia were from a few families of genetically highly related strains. Furthermore, little overlap was observed among isolates from the three countries, indicating strict isolation of the bacterial reservoirs in the countries. A few strains from the Netherlands matched strains from Ethiopia and Tunisia. Those strains were invariably isolated from refugees, immigrants, or persons who visited Ethiopia or Tunisia.

The characterization of repetitive DNA patterns in *Mycobacterium tuberculosis* has led to reliable methods to differentiate clinical isolates. Many investigators are using these techniques to answer long-standing questions about the epidemiology of tuberculosis, such as the relative prevalence of re-infection versus reactivation, and to estimate risk factors for recently acquired infections in large populations [1, 2]. Most studies have focused on tracing tuberculosis outbreaks within a community, hospital, or other high-risk environment, dissemination of multidrug-resistant strains, and evaluation of the efficacy of chemotherapy [3–13]. Only three thorough population-based studies have been published; each was done in large cities in highly industrialized countries [14–16]. These studies showed that only a fraction of the main transmission routes of tuberculosis are disclosed by classical contact-tracing practices, even in locations where an optimal infrastructure for tuberculosis control exists. Therefore, DNA restriction fragment length polymorphism (RFLP) typing of *M. tuberculosis* may be extremely useful for detecting unsuspected cases of transmission and for identifying the main sources and routes of infection [17].

The genetic elements that have been used to differentiate *M. tuberculosis* are insertion elements and short, repetitive DNA sequences with no known function [17]. The most widely used element is IS6110, a transposable sequence usually present in 5–20 copies. Because of the high discriminative power to differentiate *M. tuberculosis* isolates using the IS6110-driven RFLP technique, this tool permits the study of transmission in restricted geographic regions and transmission by the intercontinental movement of large numbers of persons. Preliminary studies have suggested an association between the DNA type of *M. tuberculosis* strains and their geographic origin [12, 14]. Therefore, DNA RFLP analysis of *M. tuberculosis* may allow investigators to trace the origin of tuberculosis infections when transmission crosses international borders.

Data also suggest that strains from countries with a high prevalence of *M. tuberculosis*, such as the Central African Republic, exhibit less DNA polymorphism than do strains in countries with a low prevalence of infection, such as the Netherlands [12]. This raises the question whether strain differentiation based on IS6110 DNA polymorphism can be used to study the epidemiology of tuberculosis in countries where infection is prevalent. Assuming that the mobility of IS6110 reflects a molecular clock [17] and that this element is not transferred horizontally to any significant extent, the IS6110-driven RFLP in a population reflects the genetic distances between the isolates within the population. We will refer to this RFLP pattern as the “population structure.”

This study was undertaken to compare the population
structure of *M. tuberculosis* in a high-prevalence country with that in countries with medium and low prevalences. We analyzed *M. tuberculosis* isolates from Ethiopia, which has an estimated tuberculosis incidence of ~300 per 100,000 people [18] and from Tunisia and the Netherlands, which have respective rates of 30 and 10 per 100,000 population.

**Materials and Methods**

*Mycobacterial strains and bacteriologic analysis.* We cultured 158 *M. tuberculosis* isolates from Ethiopian clinical specimens (sputum samples) collected in 1992 and 1993. Of these, 134 were collected at eight health centers from persons living in Addis Ababa and 24 were obtained at Sidamo Regional Hospital from residents of southern Ethiopia (Sidamo Province). The 190 Tunisian isolates were obtained over 1 year (end of 1992 to end of 1993) by 12 hospitals, mainly in eastern Tunisia. The 153 Dutch isolates were obtained from 34 public health laboratories over 1 year (1993–1994). These laboratories cover all geographic regions of the country. All 501 isolates were randomly sampled from the participating health centers, hospitals, and laboratories and are considered representative of isolates from the three countries. We also analyzed 13 isolates from Eritrean refugees who had lived in the Netherlands several years.

DNA techniques. Genomic DNA extraction and Southern blot analysis were done as described previously [12, 19]. All *M. tuberculosis* strains were typed by IS6110 using the standardized procedure described by van Embden et al. [20]. A fraction of the Ethiopian isolates was further typed by direct repeat (DR) and repetitive polymorphic GC-rich sequence (PGRS) techniques as described [21, 22]. The DNA probes were labeled using the enhanced chemoluminescence gene detection system (Amerham International, Amersham, UK). RFLP for Ethiopian, Tunisian, and Dutch isolates was done in Addis Ababa, Antwerp, and Bilthoven, respectively.

Comparison of DNA RFLP patterns. IS6110 DNA patterns were analyzed by Gelcompar software (Windows version 3.10; Applied Math, Kortrijk, Belgium) after autoradiographs were scanned at 74.8 dots/cm (190 dots/inch) (HP Scanjet IiC/T; Hewlett-Packard, Camas, WA). The mobilities of the IS6110-containing DNA fragments were compared to a set of internal molecular weight markers by superimposing the autoradiographs containing the IS6110 DNA patterns and the autoradiographs of the internal markers of known molecular size [19, 20]. This procedure normalizes the position of each IS6110-containing fragment, regardless of the autoradiograph and gel distortions. The accuracy of the procedure was evaluated by comparing the IS6110 banding patterns of strain Mt14323, which was present in a single lane on each autoradiograph [20]. Patterns were compared by the UPGMA clustering method using the Dice coefficient, following the instructions of the Gelcompar manufacturer.

**Results**

RFLP analysis of Ethiopian *M. tuberculosis* strains using IS6110 as a genetic marker. Isolates from 158 Ethiopian tuberculosis patients were analyzed by RFLP. The IS6110 patterns were highly variable (figure 1). All 158 IS6110 banding patterns were analyzed for similarity by computer analysis (figure 2A). Dendrograms were constructed to show the degree of relatedness among strains according to a previously described algorithm [12]. Ninety-six distinct IS6110 patterns were found; 14 banding patterns were shared by ≥2 patients (i.e., clusters). Fourteen clusters were found in 76 strains. Clusters occurred in 2–28 strains (figure 2A; table 1).

![Figure 1. DNA restriction fragment length polymorphism (RFLP) patterns of 21 *M. tuberculosis* isolates from Ethiopia randomly chosen from 158 isolates. Chromosomal DNA was digested with restriction enzyme PvuII. 245-bp polymerase chain reaction-amplified fragment of IS6110 was used as DNA probe. Predominant RFLP type E1 is in lanes 2, 6, and 9; type E2 is in lanes 5, 17, and 18. Numbers at left indicate sizes (in kb pairs) of standard DNA fragments. Lane 22, DNA from *M. tuberculosis* reference strain Mt14323.](image-url)
The two largest clusters were found in 16 and 28 isolates. The former had a single IS6110 element (type E1; figure 1, lanes 2, 6, 9), and the latter carried 3 IS6110 copies (type E2; figure 1, lanes 5, 17, 18). In addition to the clusters of strains with identical RFLP patterns, various groups with highly related IS6110 banding patterns could be distinguished (figure 2A).

To visualize more objectively the relatedness between the banding patterns of all Ethiopian isolates, a similarity matrix was generated. This matrix shows the degree of relatedness of each IS6110 banding pattern with any other in the collection (figure 2A). All clustered strains are shown as black triangles just below the diagonal, and all families of related strains are shown in triangular groupings in gray values. About two-thirds of the isolates belonged to clusters of identical strains or to related families of strains that shared more than two-thirds of their IS6110-containing PvuII restriction fragments.

Among the 158 Ethiopian Mycobacterium tuberculosis isolates, 24 were from patients in Sidamo Province. Twenty distinct RFLP types were observed among the Sidamo strains, and 15 had unique banding patterns. Five of the patterns were also observed among the 134 isolates from Addis Ababa. These results indicate that the DNA types and the degree of RFLP among the Sidamo strains did not differ significantly from the Addis Ababa strains.

DNA polymorphism in Ethiopian strains of predominant IS6110 DNA types E1 and E2 determined by PGRS and DR markers. Previous studies showed that Mycobacterium tuberculosis strains carrying one or several IS6110 copies are often difficult to differentiate by IS6110 RFLP analysis because of a site-specific preference of insertion of the IS element [14, 21, 22]. Therefore, we attempted to further differentiate types E1 and E2 using the genetic markers PGRS and DR. Sixteen Mycobacterium tuberculosis IS6110 type E1 strains were investigated and differentiated into 12 PGRS and 8 DR RFLP types (figure 3; table 1). By combining the PGRS and the DR results, 13 distinct PGRS-DR types could be distinguished (table 1). Among the 16 type E1 strains, 4 were from tuberculosis patients in Sidamo Province. Three of the 4 strains shared
PGRS-DR patterns with clinical isolates from Addis Ababa (data not shown).

The degree of PGRS- and DR-based RFLP among the type E2 strains was significantly less compared with polymorphism among type E1 strains. Only 5 PGRS types and 3 DR types were found among the 17 IS6110 type E2 strains investigated (figure 3, table 1). When PGRS and DR data were combined, 7 distinct PGRS-DR types were identified. Ten belonged to a single, predominant PGRS-DR type. These data indicate that 60% of type E2 strains are indistinguishable by one or a combination of the three different genetic markers. From these data, we estimate that ~10% of Ethiopian M. tuberculosis strains belong to a single clone. In addition, by PGRS and DR analysis, we investigated 5 strains that shared 2 bands with type E2 strains. Although 1 or 2 differed in mobility, the PGRS-DR fingerprint of 2 of the strains was identical to the predominant PGRS-DR type (data not shown). Computer-assisted comparison of the PGRS and DR RFLP patterns of IS type E1 and E2 strains showed that only 2 strains in these groups shared PGRS or DR patterns. The exceptions were strains of the predominant PGRS-DR type.

Analysis of the population structure of Tunisian and Dutch M. tuberculosis strains. The 190 M. tuberculosis isolates from Tunisia were analyzed by IS6110 RFLP. Results are shown in figure 2B and table 1. In total, 117 different patterns were found, and 111 (58%) of the isolates had patterns shared by ≥2 patients. Most (62%) Tunisian isolates could be grouped into three large families with closely related IS6110 DNA banding patterns (shown by the similarity matrix in figure 2B). The strains within each family contained >10 IS6110 copies, and the members of any of the three families differed in 1 or a few bands. This suggests that the majority of the Tunisian strains descended from three recently expanded clones.

We analyzed 153 isolates from the Netherlands (figure 2C, table 1). The degree of IS6110 DNA polymorphism among this set of strains was much greater than among the M. tuberculosis strains originating from Ethiopia and Tunisia. Two families of distantly related strains, comprising 23 and 14 strains, respectively, were distinguished. These accounted for 24% of all strains (figure 2C). Furthermore, only 29 (19%) IS6110 patterns were shared by ≥2 patients, suggesting that the fraction of recently acquired infections in the Netherlands is significantly lower than in Tunisia and Ethiopia. Case history investigation of contact tracing of the patients with shared IS6110 patterns revealed that the majority of the clustered strains represented recently acquired infections.

Comparison of M. tuberculosis RFLP patterns from Ethiopia, Tunisia, and the Netherlands. A computer analysis of the IS6110 banding patterns was done for all 501 M. tuberculosis isolates from the three countries. Surprisingly, no Ethiopian strain shared a pattern with any Tunisian isolate; however, 2 Ethiopian strains shared patterns with Dutch strains (DNA types E2 and E3, see figure 2A), and 1 Tunisian isolate was identical to 1 Dutch isolate (DNA type T1, see fig-
Figure 3. Restriction fragment length polymorphism (RFLP) analysis of *M. tuberculosis* strain types E2 (lanes 1–5) and E1 (lanes 6–21) using repetitive polymorphic GC-rich sequence probe (A) and direct repeat probe (B). Both panels show RFLP patterns of *Alul*-digested DNAs. Numbers at left, sizes of standard DNA fragments (in kb pairs).

Because of this striking result, we investigated whether strains isolated in Ethiopia and the Netherlands with related but nonidentical RFLP types might have a common geographic origin. For this purpose, we constructed a dendrogram for all Ethiopian and Dutch strains. We arbitrarily chose a group of related strains with >10 IS6110 copies that shared ≥60% of the bands. This group comprised 13 Dutch (branching from node 1, figure 2C) and 19 Ethiopian strains (branching from node 1, figure 2A). Figure 4 shows the banding patterns of these strains, the dendrogram, and the nationalities of the patients. All patients from Ethiopia were born in Ethiopia. Six of the 13 Dutch strains were isolated...
from patients born in Eritrea or Somalia, whereas the remaining 7 strains were mainly from immigrants from a variety of countries, including Germany, Iraq, Pakistan, and Tanzania. The RFLP patterns of Dutch isolates from patients born in Eritrea or Somalia generally shared more bands with Ethiopian isolates than did isolates from other countries (figure 4).

To further substantiate the association between DNA pattern and country of origin, we analyzed 13 *M. tuberculosis* isolates from Eritrean refugees now living near Rotterdam who were thought to be involved in a tuberculosis outbreak in 1992–1994. Two of these isolates were among the 153 Dutch isolates studied. As shown in figure 5, the IS6110 patterns of 11 of the 13 isolates were identical (type N1), indicating transmission of a single strain to 11 refugees over 2 years. Type N1 is related to the group branching from node 1 in figure 2A and C. These data confirm the hypothesis of public health authorities in Rotterdam (H. W. M. Baars, Municipal Health Service, Rotterdam, Netherlands, personal communication) that an outbreak occurred among refugees and that the index case imported the strain from his homeland. One of the other 2 refugee isolates displayed 2 IS6110 bands, and the RFLP type (DNA E4, see figure 2) was identical to 1 of the 158 Ethiopian isolates. The 2 bands of type E4 comigrated with 2 of the 3 bands of the predominant RFLP type E2 found in Ethiopia. By screening RFLP analyses of 1100 Dutch *M. tuberculosis* strains isolated in 1993–1994, we found 2 more strains with IS6110 type N1. It is possible that the latter 2 cases were due to transmission of the bacterium from the Eritrean refugees, as these strains were isolated >1 year after the first isolate with the pattern and because the patients also lived near Rotterdam.

**Discussion**

Three recent population-based studies showed that patients with *M. tuberculosis* strains with identical RFLP patterns (clusters) are likely to have recently become infected
in a single band. Nevertheless, our results clearly demonstrate that analysis of related strains results in significant epidemiologic relationships. Therefore, we conclude that UPGMA-based clustering is practical.

Comparison of RFLP patterns of \( M. \) \textit{tuberculosis} from three countries suggests that the Tunisian isolates are the most homogeneous: 62% belonged to 3 genetically related groupings, whereas 52% of the strains from Ethiopia were in 4 families of related strains. About half of the Ethiopian isolates in these families were \( \text{IS}6110 \) types E1 or E2 and carried only 1 or 3 \( \text{IS}6110 \) elements, respectively. Differentiation of such strains is poor because of preferred sites of integration of the insertion element in the mycobacterial genome [14, 21, 22, 24]. By use of the repetitive genetic elements PGRS and DR, we found a high degree of genetic diversity among E1 type strains but less diversity among E2 strains. When the genetic heterogeneity of types E1 and E2 are taken into account, the real number of clustered strains from Ethiopia is at most 36%, a significantly lower value than for strains from Tunisia. This result is somewhat surprising because one might expect a higher transmission rate in Ethiopia where there is a higher incidence of tuberculosis (300,000 in Ethiopia) than in Tunisia (30,000). Factors may contribute to this paradoxical result.

First, migration may play a larger role among Ethiopians than among Tunisians, thus contributing to the import of strains from a wider geographic area and to the genetic diversity of \( M. \) \textit{tuberculosis}.

Second, the transmission rates in Tunisia differ greatly from province to province. Chevrel-Dellagi et al. [25] found that virtually all clustered strains originated in a single province, Menzel Bourgiba, and suggested that the control programs to prevent transmission of tuberculosis differed in the four provinces studied.

Third, bacille Calmette-Guérin (BCG) vaccination may play a role in the population dynamics of \( M. \) \textit{tuberculosis}. In contrast to Ethiopia, BCG vaccination in Tunisia is common: For several decades, >90% of infants have been vaccinated. Assuming that BCG protects against certain genetic variants of \( M. \) \textit{tuberculosis} more than others, one may expect that BCG vaccination over extended periods leads to the selection of variants that are most resistant to a BCG immune response. It is tempting to speculate that the BCG vaccination in Tunisia led to the selection of strains belonging to only a few groupings or even to the expansion of single clones that escaped a BCG-imposed selection. Such a selection mechanism could at least partly explain the abundance of \( M. \) \textit{tuberculosis} strains belonging to the three predominant families in Tunisia. In Ethiopia, no such strong selection would be imposed on the \( M. \) \textit{tuberculosis} population because of the low degree (<20%) of BCG vaccination of infants. In the Netherlands, mass BCG vaccination has never been practiced.

This study on the population structure of \( M. \) \textit{tuberculosis}
in countries with a high incidence of tuberculosis shows that although the degree of RFLP in such countries is less than in a low-incidence country, such as the Netherlands, the polymorphism is large enough to allow epidemiologic studies by use of RFLP analysis.

Computer analysis of all 501 strains investigated showed that a limited number of families of large numbers of \textit{M. tuberculosis} strains circulate in Tunisia and Ethiopia and that the banding patterns have signatures that are country-specific. Apparently, \textit{M. tuberculosis} strains of related families have been circulating for long periods, likely for centuries, in these regions and little transmission from other regions has taken place. No RFLP patterns of strains from Tunisia matched those of strains from Ethiopia. A few matching patterns were found among isolates in the Netherlands. Of these, all that matched \( \geq 85\% \) of the patterns of the Ethiopian strains were found in persons who recently moved to the Netherlands. In addition, the single Dutch isolate with a Tunisian banding pattern was likely transmitted to a Dutch native on one of his frequent visits to northern Africa and Tunisia. Additional evidence for such a connection was provided by RFLP patterns of isolates from Eritrean refugees living in the Netherlands and from a blind selection of strains isolated in the Netherlands that had banding patterns similar to those of a group of strains from Ethiopia. Therefore, our preliminary observations [12] on the association between IS6110 RFLP type and geographic origin were strongly confirmed by this study. We have identified many other region-specific \textit{M. tuberculosis} types among \textit{M. tuberculosis} collections from Asia, South America, the Middle East, Central Africa, and Greenland (unpublished data). As shown in this study, the existence of region-specific \textit{M. tuberculosis} strains allows the tracing of tuberculosis across international borders or continents.

In 1993, international consensus was reached on a standardized method of \textit{M. tuberculosis} RFLP analysis [20] with a goal of comparing \textit{M. tuberculosis} patterns produced in different laboratories. In the current study, DNA analyses were done in three countries, and the results show that interlaboratory comparison is possible, thus opening the way to international epidemiologic studies of tuberculosis. Recently, a European Community-supported consortium of US and European laboratories was formed with the goal of performing such studies (unpublished data). A crucial factor for the success of such studies is the creation of a reliable data base of \textit{M. tuberculosis} RFLP analyses. This study shows that standardized DNA analyses and appropriate software will enable analysis and management of large data sets so that epidemiologic study of tuberculosis on a global scale is within reach. Such studies may finally answer long-standing important questions about differences in pathogenicity of \textit{M. tuberculosis} and specific interactions of this reemerging pathogen with its host, including the connection with human immunodeficiency virus infection [26].

Acknowledgments

We thank Leen Rigouts for critical comments; Bahri Hedhoud, Judith Abraham, Gitachew Kessella, Yohannes Negussie, Alemneh Wolde, Kassahun Temesgen, Seyoum Tatschitik, Mikedesse Gebreyehu, Senteayehu Abicho, and Bernt Lindgren for clinical specimens; Kristin Kremer, Joke Pijsenburg, Fatima Jed, and Tridia van der Laan for technical assistance; and Harm van Kranen and Nicole keteman for secretarial assistance.

References


