Interpretation of Restriction Fragment Length Polymorphism Analysis of Mycobacterium tuberculosis Isolates from a State with a Large Rural Population


Epidemiologic relatedness of Mycobacterium tuberculosis isolates from Arkansas residents diagnosed with tuberculosis in 1992–1993 was assessed using IS6110- and pTBN12-based restriction fragment length polymorphism (RFLP) and epidemiologic investigation. Patients with isolates having similar IS6110 patterns had medical records reviewed and were interviewed to identify epidemiologic links. Complete RFLP analyses were obtained for isolates of 235 patients; 78 (33%) matched the pattern of ≥1 other isolate, forming 24 clusters. Epidemiologic connections were found for 33 (42%) of 78 patients in 11 clusters. Transmission of M. tuberculosis likely occurred many years in the past for 5 patients in 2 clusters. Of clusters based only on IS6110 analyses, those with ≥6 IS6110 copies had both a significantly greater proportion of isolates that matched by pTBN12 analysis and patients with epidemiologic connections, indicating IS6110 patterns with few bands lack strain specificity. Secondary RFLP analysis increased specificity, but most clustered patients still did not appear to be epidemiologically related. RFLP clustering in rural areas may not represent recent transmission.

Restriction fragment length polymorphism (RFLP) of Mycobacterium tuberculosis based on the insertion sequence IS6110 demonstrates relatedness among isolates due to the element’s relative stability within a strain over time yet great variability in its number and position in the genome among different strains [1–5]. RFLP based on this element has been useful in directing and confirming the results of outbreak investigations, as demonstrated in several reports of nosocomial transmission of tuberculosis [6–12]. In addition, the technique has been used successfully to identify cross-contamination of cultures in the clinical laboratory [13].

More recently, RFLP technology has been applied to large cohorts of tuberculosis patients [14–16]. In these studies, patients with M. tuberculosis isolates demonstrating similar RFLP patterns have been considered recently infected and part of the same chain of transmission. In two studies of large, urban and heavily human immunodeficiency virus (HIV)–infected populations in the United States [15–16], the authors concluded that 30%–40% of tuberculosis cases were due to recent transmission, much more than the generally accepted rate of ~10% [17–19]. These estimates have important implications for tuberculosis control programs. They suggest that current methods of case-finding, investigations of contacts of persons with tuberculosis, and provision of preventive therapy to contacts are ineffective in interrupting transmission.

In this study, we analyzed RFLP patterns of M. tuberculosis isolates from tuberculosis patients in Arkansas to determine the usefulness of this technology in studying the epidemiology of tuberculosis transmission in a relatively stable, rural population with a low prevalence of HIV infection. We obtained detailed epidemiologic information for all patients whose isolate had an RFLP pattern that matched the pattern of ≥1 other isolates. In this report, we present the results of RFLP analyses of tuberculosis patients in Arkansas over a 2-year study period (1992 and 1993) and describe the characteristics of these patients and their epidemiologic connections. In addition, we propose important considerations in the interpretation of M. tuberculosis RFLP results.

Methods

Arkansas demographics and study population ascertainment. Arkansas population sizes and demographics were obtained from the 1990 census [20–22]. National and state AIDS and tuberculosis rates were obtained from Centers for Disease Control and Prevention (CDC) surveillance reports [23] and unpublished data. For the purpose of this study, urban areas in Arkansas were defined as metropolitan areas with >150,000 population.
Arkansas Department of Health (Little Rock) and national tuberculosis surveillance records were used to ascertain the study population. Study patients were persons culture positive for *M. tuberculosis*, the first culture having been collected between 1 January 1992 and 31 December 1993.

RFLP analyses. *M. tuberculosis* isolates from culture-positive study patients were obtained from the Arkansas Department of Health and the John L. McClellan Memorial Veterans Administration clinical mycobacteriology laboratories. Clinicians and mycobacteriology laboratories in Arkansas have historically relied on the services of these two laboratories, and past experience indicated that a high percentage of isolates from patients in Arkansas could be obtained from these sources. IS6110 RFLP patterns were determined according to standard procedure as previously described [24]. Similar patterns were identified by visual inspection of autoradiograms. Patterns that contained ≤5 hybridizing fragments, or bands, were considered matched if the patterns were identical; patterns with >6 bands were considered matched if they were identical or differed by the addition, subtraction, or size of a single band. Matched patterns and the patients from whom the isolates were obtained were grouped into clusters.

After the epidemiologic investigations were completed, secondary RFLP analysis became available. For isolates with matching IS6110 RFLP patterns, supplementary analysis based on the *M. tuberculosis* polymorphic GC-rich repeat sequence contained in the recombinant plasmid pTBN12 [25] was performed. The resulting RFLP patterns were visually compared. Isolates with matching IS6110 patterns and identical pTBN12 patterns were grouped into secondary clusters. Patients in secondary clusters were not stratified by the number of bands in pTBN12 patterns because these patterns contain too many small DNA fragments to accurately count bands.

Epidemiologic investigation. For all study patients, information concerning previous tuberculosis disease, risk factors for *M. tuberculosis* infection, and patient demographics was obtained from the Arkansas Department of Health records and the national tuberculosis surveillance database at the CDC. For patients in clusters, additional health department, health care provider, and laboratory records were abstracted. Interviews of clustered patients were also conducted. Experienced interviewers administered a structured, open-ended questionnaire at a location convenient for the patient. The information obtained in interviews included demographics and histories of tuberculous infection and disease, exposure to others with tuberculosis, and exposing others during the patient’s tuberculosis illness. The patients were asked to recall their lifetime history of residence, work, schools, church, and other social gathering places, as well as hospitalizations and stays in congregate living facilities. In addition, patients were asked if they recognized any names from a list of actual and fictitious tuberculosis patients. Names on the list were tuberculosis patients of the same cluster as the interviewee, patients in other clusters, and fictitious persons randomly mixed to produce a list of about 25 names in length. In the case of death of a patient, a close family member was interviewed.

Results

Arkansas Demographics and Study Population

The 1990 Arkansas state population was 2.35 million persons [20]. Two metropolitan areas had >400,000 population (West Memphis–Memphis and Little Rock), and one other area had >150,000 population (Fort Smith); these three areas were considered to be urban. The Arkansas population has been relatively stable. In 1990, 88% of residents ≥5 years old had lived in Arkansas in 1985, 79% lived within the same county, and 54% lived in the same house (total US figures: 70%, 63%, and 42%, respectively) [22]. Arkansas ranked 38th among the 50 states in foreign immigration, with the addition of 12,339 immigrants from abroad between the years 1985 and 1990 [22]. During 1992 and 1993, the annual rate of AIDS in Arkansas was approximately half the national figure (14.1 vs. 29.1/100,000 population) [23].

During the study period (1992 and 1993), there were 352 culture-positive cases of tuberculosis. Sixty-eight (19%) patients lived in one of the three urban metropolitan areas at the time of their diagnosis. Sixty-seven percent were male, and the mean age of patients was 62 years (range, 2–93). While patients outnumbered blacks (234 vs. 111), although the rate of tuberculosis among black persons was greater than for whites (14.6 vs. 5.9/100,000 population, *P* < 0.01). Ten patients (3%) were foreign born. Five patients (1%) were known to be dually infected with *M. tuberculosis* and HIV; however, HIV test results were reported for only 22% of all patients. An episode of previous tuberculosis was reported in 18 patients (5%).

RFLP Analysis

Of the 352 study patients, 250 (71%) had available cultures to allow for IS6110 RFLP analysis.

Unique IS6110 RFLP patterns were found in isolates from 113 patients (45%). Isolates from the remaining 137 patients (55%) matched ≥1 other isolate, forming 35 clusters. RFLP analysis with pTBN12 was possible for 122 (89%) of these 137 isolates in 32 of the IS6110 clusters; the remaining isolates were no longer viable. Forty-four isolates had unique RFLP patterns by pTBN12 analysis, and 78 (64%) matched ≥1 other isolate by both methods, to form 24 secondary clusters with
2–7 patients per cluster. Therefore, of the original 352 patients who were culture positive for \textit{M. tuberculosis}, 235 (67\%) had isolates that underwent complete RFLP analysis, including secondary analysis with pTBN12 if isolates clustered by IS6110 alone. Of these 235 patients, 157 (67\%) had isolates with unique RFLP patterns by either method of analysis, and 78 (33\%) matched \( \geq 1 \) other isolate using both methods. Twenty-four secondary clusters were observed, half of which were comprised of 2 members, and the 2 largest clusters contained 7 members each.

Characteristics of the 235 patients whose isolates underwent complete RFLP analysis and the 117 whose isolates did not were compared. Significantly fewer patients with extra pulmonary tuberculosis than patients with only pulmonary tuberculosis had complete RFLP analysis of their isolates (12/28 [43\%] vs. 223/324 [69\%], \( P = 0.05 \)), as did female patients compared with male patients (67/116 [58\%] vs. 168/236 [71\%], \( P = 0.01 \)). In other respects, there were no significant differences between the 2 groups.

Among patients with isolates having had complete RFLP analysis, the mean age of patients in secondary clusters was 15 years less than patients with unique isolates by either RFLP method (52 ± 16 vs. 67 ± 18, \( P < 0.001 \)). No other significant differences were identified between the groups.

**Investigation of Clustered Patients**

Interviews were conducted with 113 (83\%) of 137 patients in IS6110 clusters. In 22 cases, family members served as a proxy for the patient. Of the 24 patients not interviewed, 15 could not be located, and 9 patients (or their next of kin) refused an interview. Those interviewed included 64 (82\%) of the 78 patients in secondary clusters.

After review and comparison of information from records and interviews among the 78 patients in secondary clusters, epidemiologic connections to other members of each patient’s respective cluster were identified for 33 patients (42\%) in 11 clusters; for 45 patients (58\%), no epidemiologic connections were identified. There was no significant difference in identifying epidemiologic connections between patients interviewed directly or by proxy. Patients with no identified epidemiologic connections were significantly older than patients with identified connections (mean age [years], 58 ± 16 vs. 45 ± 14; \( P < 0.001 \)) (table 1). A significantly greater number of patients who had ever been in a substance abuse treatment center or been imprisoned were identified with epidemiologic connections. There were no significant differences in the number of patients with or without identified epidemiologic links according to other variables in table 1.

**Clustered patients with no identified connections.** Of the 45 patients in secondary clusters with no identified epidemiologic connections, 28 (64\%) never lived in the same county as others in their clusters. A group of 10 particularly illustrative patients were members of a single primary cluster based on an IS6110 RFLP pattern with 4 bands. Unfortunately, the isolates from 1 patient could not be revived to undergo secondary typing with pTBN12. Three other patients in the primary cluster had isolates with unique pTBN12 patterns. The remaining 6 patients in the primary cluster had isolates that formed 3 secondary clusters of 2 patients each by pTBN12 analysis. None of the patients in these clusters ever lived in the same county as the other patient in their respective clusters. No common living, work, school, social, or congregate settings could be identified, and none recognized the names of other members in their respective clusters. All 6 of these patients also had a history of prior tuberculosis infection; 2 patients had a positive tuberculosis skin test (TST) at least 20 years previously, and 4 patients each had a history of extensive exposure as children to a family member or friend with tuberculosis.

**Table 1.** Characteristics of patients with \textit{M. tuberculosis} isolates grouped into secondary clusters on the basis of RFLP analysis.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Yes</th>
<th>No</th>
<th>( P )</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (range)</td>
<td>45 (27–81)</td>
<td>58 (19–82)</td>
<td>(&lt;.001)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>28/34 (82)</td>
<td>31/44 (70)</td>
<td>.23</td>
<td>0.48</td>
</tr>
<tr>
<td>White race</td>
<td>19/34 (56)</td>
<td>27/44 (61)</td>
<td>.62</td>
<td>0.41</td>
</tr>
<tr>
<td>Residence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always rural</td>
<td>9/34 (27)</td>
<td>13/44 (30)</td>
<td>.77</td>
<td>0.41</td>
</tr>
<tr>
<td>Always in Arkansas</td>
<td>9/25 (36)</td>
<td>16/37 (43)</td>
<td>.57</td>
<td>0.36</td>
</tr>
<tr>
<td>Always within bordering counties</td>
<td>5/25 (20)</td>
<td>10/37 (27)</td>
<td>.52</td>
<td>0.33</td>
</tr>
<tr>
<td>Always within 1 county</td>
<td>4/25 (16)</td>
<td>6/37 (16)</td>
<td>.98</td>
<td>0.40</td>
</tr>
<tr>
<td>Ever lived in nursing home</td>
<td>1/28 (4)</td>
<td>1/37 (3)</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Ever lived in shelter</td>
<td>6/29 (21)</td>
<td>5/38 (13)</td>
<td>.51</td>
<td>0.55</td>
</tr>
<tr>
<td>Ever in substance abuse treatment center</td>
<td>8/26 (31)</td>
<td>3/36 (8)</td>
<td>.04</td>
<td>0.73</td>
</tr>
<tr>
<td>Ever health care worker</td>
<td>2/27 (7)</td>
<td>5/36 (14)</td>
<td>.68</td>
<td>0.29</td>
</tr>
<tr>
<td>Ever in military</td>
<td>7/25 (28)</td>
<td>12/37 (32)</td>
<td>.71</td>
<td>0.37</td>
</tr>
<tr>
<td>Ever homeless</td>
<td>4/28 (14)</td>
<td>2/37 (5)</td>
<td>.39</td>
<td>0.66</td>
</tr>
<tr>
<td>Ever in prison or jail</td>
<td>17/28 (61)</td>
<td>3/37/8 (8)</td>
<td>(&lt;.001)</td>
<td>0.85</td>
</tr>
</tbody>
</table>

*NOTE.* Except as indicated, data are no./no. total (%). PPV = positive predictive value.
Clustered patients with identified epidemiologic connections. Of 33 patients in 11 clusters with epidemiologic connections to other patients in their respective clusters, the connection for 10 in 4 clusters was to family members or close friends in their respective clusters. In 1 of these clusters, 1 of 2 family members worked in a bar and was likely the source case for 1 other member of the cluster who frequented the bar. Two other patients in another cluster mentioned patronizing a common bar 10 years previous to the study. Eight patients in 3 clusters had multiple social contacts in the same community, and seven cases in 1 cluster were associated with the same prison. Two patients lived and worked driving poultry trucks for separate companies in the same small town. Less definite epidemiologic connections were also established among 3 patients in the last cluster due to their homeless status, with transmission of tuberculosis possibly occurring in one or more of the homeless shelters. These 3 patients could not be located for interview.

Before the results of RFLP analysis were available, health department investigations of tuberculosis cases had identified all cases among family and friends as related. The relationships among cases associated with the prison were known, with one exception. Two community outbreaks had been well documented, but one instance of community transmission between 2 patients was not recognized. The potential link for the 2 truck drivers and the 3 bar patrons had not been suspected. Transmission among 3 patients in the homeless community had been suspected but not documented. Thus, previously unidentified and unsuspected transmission was first suspected by RFLP analysis for 7 patients. Previously suspected transmission among 3 patients was supported by RFLP analysis.

Clustered patients with prior tuberculous infection. Twenty-four patients (31%) in secondary clusters were identified as having had a previous infection with *M. tuberculosis*. Of the 24, 4 had a history of tuberculosis between 1933 and 1987, 9 had a prior positive TST, and 10 had extensive exposure to tuberculosis at least 6 years prior to their current tuberculosis diagnoses. One patient had a drug-resistant isolate. For 15, no epidemiologic connection to others in their respective clusters were identified. For 9 patients, an epidemiologic connection was identified, and for 5 of them, it was remote, having occurred between 1972 and 1977. Two of these 5 patients were in 1 cluster and had a common exposure to a patient who died of tuberculosis in 1977; 1 of these 2 patients had tuberculosis diagnosed in 1979, and the other had a documented negative TST in 1972 and then a positive TST in 1979; he did not complete a course of preventive therapy at that time. Both had tuberculosis diagnosed in 1992. The last 3 of these 5 patients were among the 7 in 1 cluster associated with the prison system. Of these 3, 1 was incarcerated for 14 months in 1969 and 1970, had a positive TST in 1971, and was diagnosed with tuberculosis in 1972. The second patient had a positive TST in 1972 while incarcerated, and the third patient had a positive TST in 1979. This person was not in prison, but her father and 2 brothers had been incarcerated multiple times since 1970, and since that time, they and 3 other members of the immediate family had been diagnosed with a total of 11 episodes of tuberculosis.

Analysis of Primary Versus Secondary Clusters

Study-patient isolates were analyzed with IS6110 alone (primary cluster analysis) and with the combination of the IS6110 and pTBN12 DNA probes (secondary cluster analysis). Of 250 patients whose isolates underwent IS6110 analysis, 137 grouped into 35 primary clusters ranging in size from 2 to 15 patients. The size of primary clusters tended to be small: 15 clusters contained only 2 patients each, 12 had 3 or 4 patients, 6 had 6 or 7 patients, and 2 had 10 and 15 patients each. Figure 1 shows the distribution of the number of bands in the IS6110 patterns. The distribution was bimodal, with modes at 4 and 12 bands and a nadir at 6 bands.

Clusters were divided according to whether patients within each cluster were epidemiologically connected. In 8 (23%) of 35 clusters, all 27 patients could be epidemiologically linked; in 8 clusters (23%), 20 of 40 patients were linked, and for 19 clusters (54%), no links could be identified among any of 70 patients. Figure 2 depicts the number of bands and patients with and without epidemiologic connections for each cluster.

Analysis by IS6110 band count revealed 12 primary clusters with ≤5 bands. For 43 patients in 9 of these clusters, no epidemiologic connections were identified. All 4 patients in 1 cluster were linked and, in the remaining 2 clusters with ≤5 bands, 4 of 10 patients were connected. For clusters with ≤5 bands, 6–11 bands, and ≥12 bands, there was a significant trend for a greater proportion of patients to have epidemiologic connections identified (14% vs. 40% vs. 60%, respectively; *P*...
< 0.001) (figure 3). Of 80 isolates in primary clusters with \(\geq 6\) IS6110 copies, 21 isolates (26%) in 11 clusters differed from other isolates in their clusters by 1 band. Patients who had isolates that matched exactly the others in their respective clusters were significantly more likely to have an epidemiologic connection identified (67%) than were patients with isolates that differed from others by 1 band (17%, \(P < 0.001\)).

Secondary RFLP analysis was concordant with the results of epidemiologic investigation of primary clusters with the exception of the analysis of isolates that differed by 1 band. A significantly greater proportion of isolates in clusters based on \(\leq 5\) IS6110 bands had unique pTBN12 patterns compared with isolates in clusters based on \(\geq 6\) IS6110 bands (29/51 [57%] vs. 13/69 [19%], \(P < 0.001\)) (table 2). However, of the isolates in primary clusters with \(\geq 6\) IS6110 bands, there was no difference in secondary clustering of those that matched exactly in primary clusters compared with those that differed by 1 band.

Overall, the percentage of patients in clusters having an epidemiologic connection increased from 34% (47/137) to 42% (33/78) when results from both RFLP methods were combined; however, this increase did not reach statistical significance (figure 3). The majority of patients in secondary clusters remained without identifiable epidemiologic connections.

Discussion

Previous studies utilizing \(M.\) \textit{tuberculosis} RFLP analysis in large populations assumed that isolates with matching RFLP patterns were epidemiologically related and represented recent transmission of tuberculosis [15, 16]. Using a secondary typing technique that allowed greater strain specificity, one-third of the patients in this study had \(M.\) \textit{tuberculosis} isolates with matching RFLP patterns, a number similar to results obtained from the above studies in urban areas using IS6110 analysis alone; yet, epidemiologic connections could be identified for less than half of these clustered patients. Epidemiologic connections may have been missed because a number of tuberculosis case-patients either could not be included in the study or were not interviewed. These cases include 33% of culture-positive patients whose isolates were either unavailable for analysis or, for technical reasons, the RFLP patterns could not be produced, and 18% of patients in clusters who would not or could not be interviewed. In general, even detailed interview and record review may not be sufficiently sensitive to identify all potential epidemiologic connections among patients.
In many cases, however, there were multiple findings that strongly suggest that either no connection existed among clustered patients or that transmission occurred many years in the past. First, epidemiologic evidence in some clusters indicated that the patients were so demographically, geographically, and socially dissimilar that common casual contact among all, or to a common source, was unlikely. In addition, approximately one-third of clustered patients had evidence of prior tuberculosis infection. A recent review of nosocomial outbreaks of tuberculosis concluded that *M. tuberculosis* reinfection of immunocompetent patients is very rare [28]. Thus, the probability is small that recent transmission with reinfection is responsible for the clustering of this large number of otherwise healthy patients. For 5 patients with prior tuberculosis infection, we identified old epidemiologic connections, indicating that RFLP clustering may result from transmission years in the past.

Second, the epidemiology of tuberculosis in Arkansas suggests effective control with little ongoing transmission in the population. Arkansas had not experienced an increase in the number of tuberculosis patients between 1985 and 1992, as seen in some parts of the United States, and has few recent immigrants or HIV-infected residents. The average age of tuberculosis patients in Arkansas has been well above the national average, and only a few outbreaks have been documented. Yet one-third of tuberculosis patients in this study had isolates with matching RFLP patterns. Any assumption that matching RFLP patterns identify recently infected persons in Arkansas runs counter to considerable evidence that little transmission of *M. tuberculosis* has been occurring there.

Prior knowledge of the characteristics of a population may allow greater accuracy in the interpretation of results of *M. tuberculosis* RFLP analysis. The predictive value of clustered RFLP patterns to identify epidemiologic connections among the patients involved may vary according to the probability of the recent introduction of a new strain(s) of tuberculosis in the population and to the concentration, mixing, and stability of the population. We observed relatively high predictive values for epidemiologic connections among patients in highly mixed and mobile populations, such as those who had been in prison or stayed in shelters or substance abuse treatment centers, although the site of transmission did not necessarily involve these facilities. The predictive value of RFLP clustering may also be high for large, crowded, shifting populations into which new tuberculosis strains are constantly introduced and mixed, as was found in a recent study of tuberculosis patients in Denver. In that study, 78% of clustered patients had identifiable epidemiologic connections [29]. On the other hand, in stable, dispersed populations without evidence of ongoing transmission, RFLP clustering may have a low predictive value for recent epidemiologic connections, as was observed in this study.

The findings in this study suggest important considerations in the interpretation of RFLP analysis. First, *M. tuberculosis* strains with matching RFLP patterns may not be epidemiologically related or may be epidemiologically related as a result of transmission that occurred many years in the past. The relative probability of the various epidemiologic relationships, whether, recent, remote, or nonexistent, depends on the population studied and the method of RFLP analysis used. Matching IS6110 RFLP patterns with \( \geq 6 \) bands are more predictive of clonality than patterns with fewer bands, which may require the use of alternative or additional DNA probes, such as pTBN12, to differentiate strains. Of interest, in this study, several isolates had matching IS6110 patterns with \( \geq 6 \) bands but unique pTBN12 patterns, a finding that was not related to the addition or movement of 1 band in the matching IS6110 patterns. Accurate strain identification, and the ensuing conclusions about the epidemiologic relatedness of isolates, requires appropriate application of RFLP analysis, possibly using multiple probes, and careful interpretation of results. In addition, the predictive value of an epidemiologic relationship among patients with matching isolates may be low in some populations. The investigation of clustered patients, using conventional epidemiologic methods and taking into account these limitations, should provide useful adjunctive information in determining person-to-person *M. tuberculosis* transmission and thus assist in targeting tuberculosis prevention and control activities in the future.

### Acknowledgments

We are indebted to Don Cunningham for collecting and culturing patient samples and for isolating DNA from isolates; to Arkansas communicable disease nurse specialists Hazel Mabry, Delores Brown, Terri Iacobacci, Billie Mitchell, Frances Lunsford, Donetta McGriff, Rose Lipscomb, Elisia White, Mary Lou Painter, and Jeannie Gillis for persistence in locating and contacting patients; and to Willis Good, Judy Gibson, and James Enns for their skill in conducting lengthy interviews.
References


