JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 2005, p. 870–878 Vol. 43, No. 2
0095-1137/05/$08.00+0 doi:10.1128/JCM.43.2.870–878.2005

Long-Term Molecular Analysis of Tuberculosis Strains in Alabama, a State Characterized by a Largely Indigenous, Low-Risk Population

Mirjam-Colette Kempf,[1]* Nancy E. Dunlap,[1,2] Kerry H. Lok,[1] William H. Benjamin, Jr.,[2,3] Nancy B. Keenan,[4] and Michael E. Kimerling[5,6]

Division of Pulmonary and Critical Care Medicine,[1] Department of Microbiology,[2] Department of Pathology,[3] Division of General Internal Medicine,[6] and Department of Epidemiology,[5] University of Alabama at Birmingham, and Alabama Department of Public Health,[4] Montgomery, Alabama

Received 14 May 2004/Returned for modification 4 August 2004/Accepted 10 September 2004

With a tuberculosis case detection rate of 5.9 per 100,000 population in 2001, Alabama ranked twelfth highest in the United States. However, cases among foreign-born and human immunodeficiency virus-infected individuals remain low in Alabama. To understand the endemic statewide disease pattern, tuberculosis strains were studied for clustering in a long-term population-based study from January 1994 to May 2000. IS*6110* restriction fragment length polymorphism analysis was performed for 1,834 strains. Spoligotyping was used as a secondary typing method for the 37% of isolates displaying a restriction fragment length polymorphism pattern with <6 IS*6110* copies. A total of 721 (41%) patients provided isolates that composed 114 clusters, each containing isolates from 2 to 136 patients, suggesting that recent transmission accounted for 35% of tuberculosis cases. Demographic, behavioral, and clinical characteristics of patients with clustered versus nonclustered isolates stratified by low-copy-number strains (<6 IS*6110* copies) versus high-copy-number strains (≥6 IS*6110* copies) were evaluated. Younger age, black race, a history of alcohol abuse, and homelessness were predictors of clustering of low-copy-number, strains and younger age, urban residency, alcohol abuse, homelessness, noninjection drug use, and a history of incarceration and/or cavitary disease were predictors of clustering of high-copy-number strains. By identifying local characteristics of tuberculosis clustering through molecular fingerprinting, control programs can distribute their limited resources to impact the transmission of tuberculosis in high-risk populations and evaluate strain distribution across geographical areas.

During the last decade, tuberculosis case rates in Alabama have decreased significantly from 12.8 per 100,000 population in 1988 to 5.9 per 100,000 population in 2001 (Program Data, Division of Tuberculosis Control, Alabama Department of Public Health). This success has been achieved through the application of traditional control strategies, i.e., the identification and treatment of active cases, coupled with an aggressive contact investigation program. Due to limitations of traditional investigation tools, genotyping techniques have been used to identify unsuspected links between tuberculosis cases. As such, molecular strain typing of *Mycobacterium tuberculosis* by restriction fragment length polymorphism (RFLP), using the IS*6110* insertion sequence, has become the primary standardized method (33) for investigating tuberculosis outbreaks, as well as for assessing ongoing transmission among various populations and within various communities. Because the differentiation of *M. tuberculosis* strains carrying fewer than six IS*6110* copies is poor, secondary molecular typing techniques are now used.

In population-based studies, cluster analysis, as defined by RFLP analysis in combination with other genotyping techniques, has been used successfully as a tool to determine the extent of, and risk factors for, recent transmission of tuberculosis in predominantly urban areas (1, 3, 5, 30, 37). Few studies have systematically evaluated tuberculosis transmission across wide areas where tuberculosis is endemic (7, 9, 22). However, most studies have not taken into account the fact that short study intervals lead to an underestimation of clustering among certain age groups (36). Consequently, ongoing transmission may have been misclassified as reactivated tuberculosis due to short study intervals. The long-term monitoring of tuberculosis clustering through molecular typing techniques has been accomplished in Alabama, where widespread endemic tuberculosis patterns can overlap with distinct outbreaks and endemic disease transmission patterns are inadequately understood (17, 27).

This long-term population-based study describes the sociodemographic and clinical risk factors for clustered versus nonclustered cases of tuberculosis over more than 6 years (1994 to 2000) in a predominantly rural state and without significant contributions to disease epidemiology from commonly identified high-risk groups, such as foreign-born or human immunodeficiency virus (HIV)-infected persons. A statewide molecular fingerprinting database was used for this analysis. Since it has been concluded previously that low-copy-number strains (LCS) and high-copy-number strains (HCS) represent different lineages of strains (13, 20) and the usefulness of molecular typing of low-copy-number strains has been controversial in predicting recent transmission (4, 12, 23, 26, 31), we stratified

* Corresponding author. Mailing address: Department of Epidemiology, RPHB 217G, 1530 3rd Ave. South, Birmingham, AL 35294-0022. Phone: (205) 934-7186. Fax: (205) 975-3329. E-mail: Mkempf@uab.edu.

† Present address: Genetic Core Facility, Biochemistry and Molecular Division, Department of Nutrition Science School of Health-Related Professions, University of Alabama at Birmingham, Birmingham, AL 35294.

our analysis by clustering low- versus high-copy-number strains. Risk factors and risk estimates for clustering of *M. tuberculosis* strains were identified in order to obtain a better understanding of endemic tuberculosis patterns and transmission factors due to different *M. tuberculosis* strains (low-copy-number strains versus high-copy-number strains). Analysis of possible risk factors attributable to recent transmission should enable tuberculosis control personnel to target intervention strategies to populations at high risk for transmission of tuberculosis and better rationalize the use of limited public health resources in the national effort to eliminate transmission of tuberculosis. The findings and conclusion drawn may be of significance to other rural geographic areas where endemic distribution of tuberculosis has been observed.

MATERIALS AND METHODS

Study population. From 1 January 1994 to 1 May 2000, 2,528 tuberculosis cases were diagnosed in the state of Alabama and reported to the Centers for Disease Control and Prevention (CDC) in Atlanta, Ga. Detection, treatment, and prevention of tuberculosis were directed by the Tuberculosis Control Division of the Alabama Department of Public Health in Montgomery. Tuberculosis patients were diagnosed and classified according to the CDC's clinical case and laboratory criteria (8). Demographic and clinical data were obtained from reports of the state Tuberculosis Control Division to the CDC, which included information on treatment of prior episodes of tuberculosis, bacteriology, and specific behavioral characteristics. All variables were measured in categories except for age, which was measured as a continuous variable.

Of the 2,528 tuberculosis cases, 2,204 (87%) were bacteriologically confirmed as culture positive by the central state mycobacteriology laboratory of Alabama. A culture was classified as laboratory cross-contaminated and excluded from further analysis when all of the following criteria were met: (i) the patient had only one positive culture and that specimen was negative for acid-fast bacilli on microscopy, (ii) the specimen had been processed in the same laboratory within 28 days of another specimen from another patient with positive acid-fast bacilli on microscopy, and (iii) the isolates had the same DNA fingerprint.

Molecular analysis and cluster definition. Using standard IS*6110* RFLP typing, 1,834 (83%) of the 2,204 *M. tuberculosis* isolates were successfully typed (1). Due to cross-contamination or inadequate imaging, 28 samples were excluded, leaving 1,806 RFLP-typed strains for consideration. Computer-assisted analysis of RFLP patterns was done by use of a Whole-Band Analyzer, version 3.3 (Bio Image, Inc., Ann Arbor, Mich.). Because the differentiation of *M. tuberculosis* strains carrying few IS*6110* copies is poor (35), 623 (93%) of 670 patient isolates harboring fewer than six IS*6110* copies were subjected to further analysis by spoligotyping (n = 1,759 strains available for complete cluster analysis). Since secondary typing was done retrospectively, DNA was not available for all 670 isolates harboring fewer than six IS*6110* copies.

Clusters were defined as groups of patients (at least two cases in the 6-year period) having isolates with identical RFLP patterns, that is, the same number of IS*6110* copies at identical band positions. Isolates with fewer than six IS*6110* copies had to have matching IS*6110* images, as well as identical spoligotyping patterns, in order to be considered a cluster. Nonclustered cases were defined as patients with isolates having no matching RFLP/spoligotyping pattern identified during the 6-year study period. For the present study, two persons who were blind to the patient demographic data individually confirmed IS*6110* patterns.

Recent transmission was determined by using the correction factor method (the "$n - 1$" method). It is assumed that a cluster size of n contains "$n - 1$" individuals with recent infections and one individual with an old infection (30).

Statistical analysis. Data for 1,759 samples were analyzed by using the SAS software package (version 8.1; SAS Institute, Cary, N.C.). Chi-square tests were performed to test for an association between clustering and various predictor variables. Since clustering has been shown to be associated with age (14), age-associated odds ratios (ORs) were calculated to adjust for possible confounding. The Student t test was used to compare the mean ages of clustered and nonclustered patients. Variables showing univariate significance levels of ≤ 0.2 were considered for entry in a logistic regression model. Models were adjusted for missing data by using indicator terms for the presence or absence of missing data (29). Variables that were significantly associated with clustering ($P < 0.05$) after multivariate analysis were included in a final logistic regression model. Comparisons between clustered and nonclustered cases were expressed as ORs with 95% confidence intervals (CIs), whereas adjusted ORs were calculated from regression coefficients and 95% CIs from standard errors of the respective regression coefficients. *M. tuberculosis* strains harboring fewer than six IS*6110* copies were analyzed separately from *M. tuberculosis* strains harboring six or more IS*6110* copies. Mapping of tuberculosis cases was done by using the Microsoft MapPoint software package.

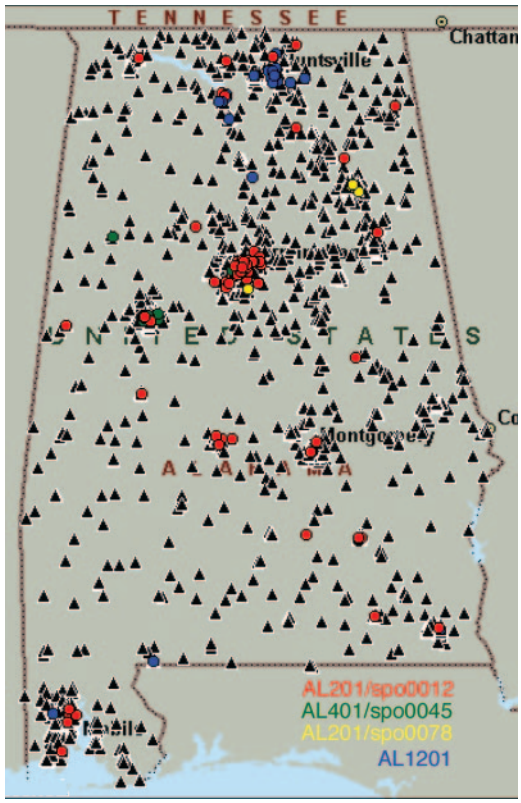

FIG. 1. Map of the state of Alabama showing residence locations of individuals who were diagnosed with tuberculosis from 1 January 1994 to 1 May 2000. Cases previously described in outbreaks of tuberculosis are shown as dots; the remaining cases are shown as triangles.

Consent. Verbal informed consent was obtained from patients by the Tuberculosis Control personnel according to the experimentation guidelines of the U.S. Department of Health and Human Services. Information was used to care for the patients' medical needs during the treatment phase of their illness. Prior to the investigators obtaining information for aggregate data analysis, personal identifiers were stripped from the database.

RESULTS

Endemic and epidemic tuberculosis cases in Alabama (1994 to 2000). Distribution of all 2,528 reported tuberculosis cases is shown in Fig. 1, including four of the previously identified outbreaks (16, 27) (E. Khan, M. E. Kimerling, K. H. Lok, W. H. Benjamin, J. J. Lynch, P. Phillips, and N. E. Dunlap, Am. Thorac. Soc. Meet., San Diego, Calif., abstr. A305, 1999). Although microepidemics are concentrated in urban areas of the state, cases are evenly distributed across the entire state. The geographic overlap of endemic and epidemic cases makes it difficult to distinguish the two modes of transmission from each other without molecular characterization of the strains.

Demographic characterization of the study population. In total, 1,759 molecularly characterized patient isolates were included in the final univariate and multivariate analyses. The

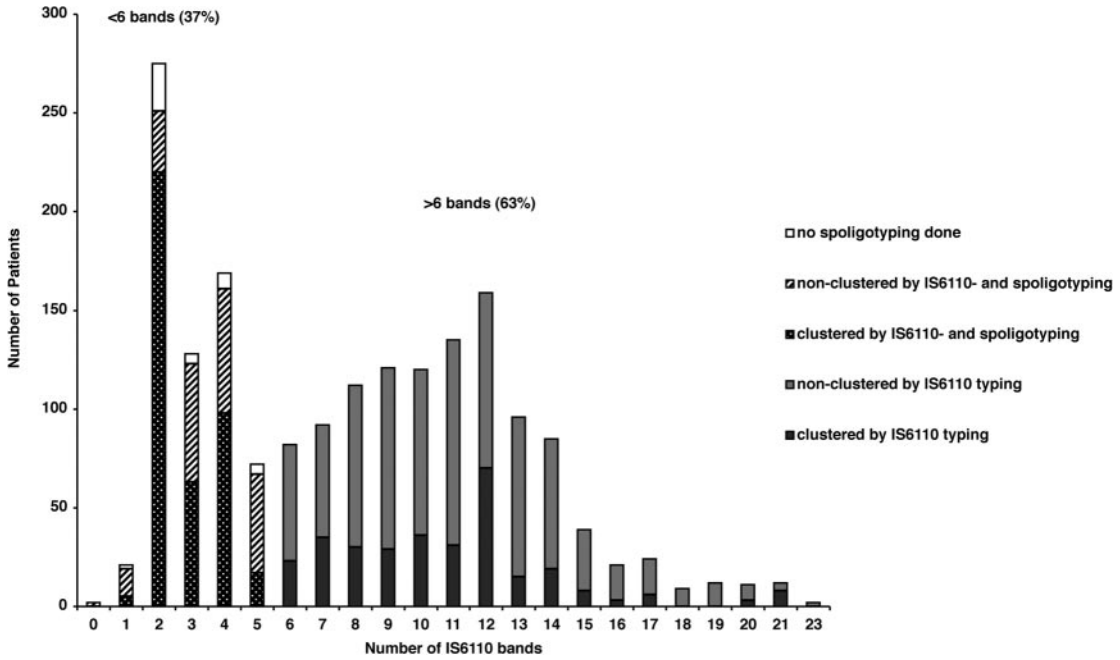

FIG. 2. Patient distribution according to number of bands based on RFLP IS*6110* patterns of tuberculosis patients in Alabama. A total of 1,806 tuberculosis patients were analyzed by using IS*6110* RFLPs to differentiate between clustered (■) and nonclustered (▨) cases. Strains harboring fewer than six IS*6110* copies were subjected to secondary fingerprinting by spoligotyping. □, patient samples that were not subjected to spoligotyping.

study population with molecular data available did not significantly differ with regard to demographics from the total population of 2,528 tuberculosis patients diagnosed in Alabama during the 6-year study period. The majority of the sample population (1,427 [81%] subjects) lived within their respective city limits at the time of their diagnosis. The predominant gender was male (1,225 [69%] subjects). Furthermore, 767 (44%) of the patients diagnosed with tuberculosis were white, 934 (53%) were black, and the remaining 62 (3%) were either American Indian/Alaskan Native or Asian/Pacific Islander. At the time of treatment, the mean age of the patients was 55 ± 20 years.

Characteristics of clusters based on IS*6110* profiles. Of the 2,528 patients who were identified with tuberculosis in Alabama, 1,759 corresponding clinical samples were typed by either RFLP typing alone (*M. tuberculosis* strains harboring six or more IS*6110* insertion sequences, i.e., HCS) or RFLP typing with subsequent spoligotyping (*M. tuberculosis* strains harboring fewer than six IS*6110* insertion sequences, i.e., LCS) (Fig. 2). RFLP typing revealed various numbers of hybridizing bands (0 to 23 bands), corresponding to 1,136 cases of HCS and 623 cases of LCS. All together, 318 of 1,136 (28%) HCS and 403 of 623 (65%) LCS clinical specimens were classified as clustered, representing 33 clusters and 81 clusters, respectively. If we assume that each cluster originated in one index case, recent transmission accounted for 35% of tuberculosis cases (607 patients).

Within a cluster the number of patients varied from 2 to 136, whereas 48 (42%) of the clusters consisted of only two patients. However, almost half (49%) of the clustered cases were grouped into large clusters (≥10 cases); four of the biggest clusters have been described previously as distinct outbreaks (16, 27) (Khan et al., Am. Thorac. Soc. Meet.) (Fig. 3).

Risk factors for clustering among high-copy-number and low-copy-number tuberculosis strains. To distinguish between HCS and LCS and risk factors for transmission within the patient cohorts, separate analyses were performed for each cohort (see Tables 1 to 4). Demographic characterization of the HCS cohort showed that 69% of the population were male, 49% were black, and 48% were white and that the majority (79%) of the subjects lived within city limits at the time of diagnosis. By comparison, 71% of the LCS cohort were male, 60% were black, 36% were white, and 85% lived within city limits. In both groups, younger age was significantly associated with clustering; however, patients with HCS were older (mean age, 54 ± 19 years) than patients with LCS (mean age, 48 ± 19 years). Gender was not associated with clustering in either of the cohorts, although black race was associated with clustering within the LCS cohort (OR, 2.3; 95% CI, 1.6 to 3.2) but not within the HCS cohort. Conversely, clustering was associated with residency within city limits among the HCS cohort (OR, 2.0; 95% CI, 1.4 to 2.8) but not among the LCS cohort.

Investigation of specific behaviors among tuberculosis patients revealed several risk factors that were associated with clustering. Among the HCS cohort, the risk factors for clustering were noninjection drug use within the past year prior to diagnosis (OR, 1.9; 95% CI, 1.1 to 3.3), residence in a correctional facility within the past year prior to diagnosis (OR, 2.9; 95% CI, 1.3 to 6.6), excessive alcohol use within the past year (OR, 2.4; 95% CI, 1.8 to 3.3), and homelessness within the past year (OR, 4.4; 95% CI, 2.6 to 7.7). Analysis of risk factors for clustering among the LCS cohort identified only excessive al-

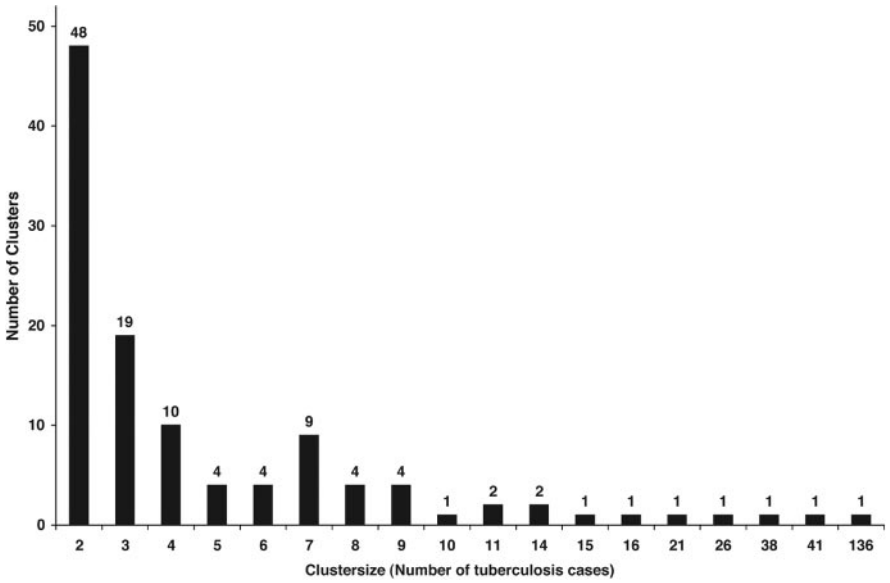

FIG. 3. Number of clusters according to cluster size. Clusters were defined by IS*6110* RFLP typing. For *M. tuberculosis* strains with fewer than six bands, additional secondary typing (spoligotyping) was used to define clusters.

cohol use (OR, 2.0; 95% CI, 1.4 to 2.9) and homelessness (OR, 2.3; 95% CI, 1.2 to 4.3) within the year prior to diagnosis as risk factors. HIV seropositivity, injection drug use, and residence in a long-term care facility were not risk factors for clustering in either group. Nevertheless, the overall proportions of risky behaviors were low and did not account for more than 10% of the study population, except for alcohol abuse (HCS [27%] versus LCS [35%]) and noninjection drug use among the LCS cohort (12%) (Tables 3 and 4).

Analysis of both HCS and LCS clustering revealed that clinical parameters such as a positive sputum smear and a history of previous tuberculosis were not associated with clustering (Tables 3 and 4). However, 1,042 (91%) patients with HCS isolates were identified with abnormal X rays, and within this subgroup of patients, cavitary disease was associated with clustering (OR, 1.5; 95% CI, 1.1 to 2.0). Overall, the clinical presentations and histories of HCS and LCS groups were similar except for a slightly higher proportion of sputum smear-positive cases (48% versus 44%) and presentation of cavitary disease (25% versus 21%) among the LCS cohort.

Age-adjusted analysis of the study variables among patients with HCS eliminated some of the associations between risk factors and clustering; nonetheless, urban residence (age-adjusted OR, 1.9; 95% CI, 1.3 to 2.7), alcohol abuse (age-adjusted OR, 2.2; 95% CI, 1.6 to 3.0), and homelessness (age-adjusted OR, 4.1; 95% CI, 2.3 to 7.1) remained strong risk factors for clustering. Among the LCS cohort the only variable that was not confounded by age appeared to be race. Being

TABLE 1. Characteristics of clustered and nonclustered tuberculosis patients in Alabama (1994 to 2000) who were infected with *M. tuberculosis* strains harboring ≥6 IS*6110* insertion sequences (HCS cohort)

Characteristic	No. of patients with characteristic or mean age (yr) ± SD	% of patients in cluster	*P*[a]	OR (95% CI)[b]	
				Crude	Age adjusted
Type			0.0007		
Clustered	54 ± 19				
Nonclustered	59 ± 21				
Gender					
Male	779	29		1.2 (0.9–1.6)	1.2 (0.9–1.6)
Female	357	25		1	1
Race					
Black	556	31		1.3 (1.0–1.6)	1.2 (0.9–1.5)
White	540	26		1	1
Other	40	5		0.1 (0.0–0.6)	0.1 (0.0–0.5)
Residence within city limits					
Yes	895	31		2.0 (1.4–2.8)	1.9 (1.3–2.7)
No	241	18		1	1

[a] As determined by the Student *t* test.
[b] As determined by the chi-square test.

TABLE 2. Characteristics of clustered and nonclustered tuberculosis patients in Alabama (1994 to 2000) who were infected with *M. tuberculosis* strains harboring <6 IS*6110* insertion sequences (LCS cohort)

Characteristic	No. of patients with characteristic or mean age (yr) ± SD	% of patients in cluster	P^a	OR (95% CI)[b]	
				Crude	Age adjusted
Type			<0.0001		
Clustered	48 ± 19				
Nonclustered	59 ± 20				
Gender					
Male	442	67		1.5 (1.0–2.1)	1.5 (1.0–2.2)
Female	181	58		1	1
Race					
Black	375	73		2.3 (1.6–3.2)	1.7 (1.2–2.5)
White	226	54		1	1
Other	22	32		0.4 (0.2–1.0)	0.3 (0.1–0.7)
Residence within city limits					
Yes	531	65		1.2 (0.8–1.9)	0.8 (0.5–1.3)
No	92	61		1	1

[a] As determined by the Student *t* test.
[b] As determined by the chi-square test.

black was still a strong risk factor for clustering within this group (age-adjusted OR, 1.8; 95% CI, 1.3 to 2.6).

Multivariate analysis revealed the same independent associations between variables and clustering as seen in the age-adjusted analysis. In addition to these variables, age was inversely related with clustering among isolates with fewer than six IS*6110* copies after adjustment for confounding in a logistic regression model (adjusted OR, 0.97; 95% CI, 0.96 to 0.98).

DISCUSSION

In this population-based study, 1,759 patient isolates of *M. tuberculosis* were analyzed during a 6-year period, representing 70% of all culture-confirmed tuberculosis cases in Alabama. A total of 35% of the detected tuberculosis cases in our study could be attributed to recent transmission, a statistic similar to that noted in studies in Arkansas, The Netherlands, and Switzerland (7, 22, 24). However, it has been argued that molecular finger-typing cannot be used in rural areas to differentiate recent transmission from remote transmission, since epidemiological linkages were not evident between clustered cases (7). Although the present study did not link traditional contact investigation with molecular fingerprinting techniques, molecular typing can be useful in investigating transmission patterns in rural settings, particularly when extended time periods are utilized. Not only were half of our clustered cases grouped into large clusters (≥10 patients), but even under optimal conditions traditional contact tracing may not reveal all contacts (37). It is plausible, therefore, that the time/duration of the analysis period provides more insight into transmission patterns than a rural setting versus a nonrural setting. Our study shows that if tuberculosis epidemiology is followed for a sufficiently long time in a rural population, then recent transmission can be observed and linked to the clinical epidemiology of infection.

Nevertheless, caution when evaluating isolates with fewer than six IS*6110* patterns is advised. It has been shown that matching IS*6110* RFLP patterns with ≥6 bands is more predictive for clonality than patterns with fewer bands, which require additional DNA probes to differentiate strains (11, 23, 31). Previous studies concluded that low- and high-copy-number strains represent different lineages of strains (13, 20). Therefore, we stratified our analysis into HCS and LCS groups to account for the prevalence of different strains in different risk groups and for accuracy of cluster determination.

A higher percentage of LCS specimens were clustered than those with at least six bands (HCS), 65% versus 28%, which corroborates recent findings by Ellis et al. (19). These findings can be partially explained by large clusters with more cases in the LCS group (one cluster with 136 patients) but also could be due to an overestimation of clustering and therefore overestimation of active transmission in this group. Although secondary typing has led to better characterization of IS*6110* low-copy-number isolates, further discrimination of clusters with so few bands could be required.

Nonetheless, the prominence of a two-band RFLP pattern in the state is striking. This group of *M. tuberculosis* isolates, called JH2 or 00016 (National Tuberculosis Genotyping Fingerprint Pattern), was reported with a final frequency of 5% among seven sentinel surveillance sites in the United States (27). It is suggested to be an older, more stable IS*6110* pattern within the United States with links to the African continent (24, 27). The strong association found in the present study with regard to black race and clustering among the LCS cohort supports these findings.

Seven statistically significant predictors for clustering were identified among the HCS cohort: younger age, residence within city limits, noninjection drug use, incarceration during the past year, excessive alcohol abuse, homelessness, and cavitary disease. Among the LCS group, younger age, black race, excessive alcohol abuse, and homelessness were predictive of clustering. Recent transmission appears to be predominant among middle-aged African-Americans engaging in a risk be-

TABLE 3. Risk factors for clustering of tuberculosis isolates in Alabama (1994 to 2000) from patients that were infected with *M. tuberculosis* strains harboring six or more IS*6110* insertion sequences (HCS cohort)

Demographic or clinical parameter	No. of patients with characteristic[a]	% of patients in cluster	OR (95% CI)[b]	
			Crude	Age adjusted
Demographic parameters				
Injection drug use				
Yes	11	45	2.2 (0.7–7.3)	1.9 (0.6–6.3)
No	1,003	28	1	1
Noninjection drug use				
Yes	54	41	1.9 (1.1–3.3)	1.6 (0.9–2.8)
No	952	27	1	1
Alcohol abuse				
Yes	270	41	2.4 (1.8–3.3)	2.2 (1.6–3.0)
No	733	23	1	1
HIV positive				
Yes	51	31	1.0 (0.6–1.9)	0.9 (0.5–1.7)
No	743	31	1	1
Homeless				
Yes	57	61	4.4 (2.6–7.7)	4.0 (2.3–7.0)
No	1,070	26	1	1
Long-term care facility resident				
Yes	27	22	0.7 (0.3–1.8)	0.9 (0.4–2.4)
No	1,108	28	1	1
Correction facility resident				
Yes	23	52	2.9 (1.3–6.6)	2.3 (1.0–5.4)
No	1,113	27	1	1
Clinical parameters				
Abnormal X ray				
Yes	1,038	28	1.0 (0.6–1.7)	1.1 (0.7–1.9)
No	75	28	1	1
Abnormal X ray/cavitary disease				
Yes	244	34	1.5 (1.1–2.0)	1.4 (1.0–1.9)
No	798	26	1	1
Sputum smear positive				
Yes	501	31	1.2 (0.9–1.6)	1.2 (0.9–1.5)
No	491	27	1	1
Previously tuberculosis positive				
Yes	55	31	1.2 (0.6–2.1)	1.2 (0.6–2.0)
No	1,081	28	1	1

[a] The totals are less than 1,136 due to missing data.
[b] As determined by the chi-square test.

havior, such as excessive alcohol consumption, or who are associated with risk groups, such as the homeless.

In general, the negative correlation between age and clustering in this population-based study, as well as in similar studies from the United States (8, 28, 30), Cuba (14), and Denmark (38), support the widely held view that the clustering of DNA fingerprints does reflect recent transmission. These studies show that the percentage of cases clustered decreased with age, indicating that the proportion of disease attributable to endogenous reactivation increases with age. We also found that certain demographic variables influenced the clustering of tuberculosis cases. The association of black race and recent transmission of tuberculosis among the LCS group again underscores the race disparity of infectious disease and tuberculosis transmission, as has been shown by Borgdorff et al. (6). This is of particular concern, since only 25% of Alabamians reported in the 1990 population census were of black race. Furthermore, a higher rate of tuberculosis transmission was observed in urban environments, probably due to higher transmission rates among alcohol and/or drug abusers and homeless populations. These observations are similar to findings reported in The Netherlands, New York, and California (3, 25, 32, 34). However, since Alabama is a mixed rural-urban setting, with tuberculosis distributed throughout the state (17), linkage between urban residence and clustering helps to more clearly identify ongoing transmission geographically.

TABLE 4. Risk factors for clustering of tuberculosis cases in Alabama (1994 to 2000) from patients that were infected with *M. tuberculosis* strains harboring fewer than six IS*6110* insertion sequences (LCS cohort)

Demographic or clinical parameter	No. of patients with characteristic[a]	% of patients in cluster	OR (95% CI)[b]	
			Crude	Age adjusted
Demographic parameters				
Injection drug use				
Yes	18	78	2.0 (0.7–6.3)	1.4 (0.4–4.4)
No	531	63	1	1
Noninjection drug use				
Yes	66	73	1.6 (0.9–2.8)	1.0 (0.5–1.8)
No	483	63	1	1
Alcohol abuse				
Yes	196	74	2.0 (1.4–2.9)	1.4 (1.0–2.2)
No	362	59	1	1
HIV positive				
Yes	36	72	1.2 (0.6–2.6)	0.9 (0.4–1.9)
No	427	68	1	1
Homeless				
Yes	63	79	2.3 (1.2–4.3)	1.6 (0.9–3.2)
No	555	63	1	1
Long-term care facility resident				
Yes	17	65	1.0 (0.4–2.7)	1.7 (0.6–4.9)
No	606	65	1	1
Correction facility resident				
Yes	18	83	2.8 (0.8–9.8)	2.1 (0.6–7.4)
No	605	64	1	1
Clinical parameters				
Abnormal X ray				
Yes	573	65	1.7 (0.9–3.3)	1.9 (0.9–3.8)
No	36	53	1	1
Abnormal X ray/cavitary disease				
Yes	153	71	1.4 (1.0–2.1)	1.1 (0.7–1.7)
No	420	63	1	1
Sputum smear positive				
Yes	298	69	1.3 (0.9–1.9)	1.1 (0.8–1.6)
No	247	63	1	1
Previously tuberculosis positive				
Yes	38	71	1.4 (0.7–2.8)	1.3 (0.6–2.7)
No	585	64	1	1

[a] Some totals are less than 623 due to missing data.
[b] As determined by the chi-square test.

In this analysis, homelessness and alcohol abuse were found to be the major risk factors for clustering. Similar findings have been described in San Francisco, Calif.; Los Angeles, Calif.; St. Louis, Mo.; and Hamburg, Germany (3, 10, 15, 28), where substance abuse and homelessness were among the most significant risk factors contributing to the transmission of tuberculosis. In contrast to studies in San Francisco and New York City (1, 30), HIV was not found to be associated with clustering. The most likely explanation is the higher prevalence of HIV in the population monitored in New York and San Francisco and prior transmission in institutional settings such as hospitals and jails (18, 21). However, negative associations between clustering and HIV were also found in St. Louis and Los Angeles (3, 28).

Clinical parameters that were analyzed, such as abnormal chest radiographs, positive sputum smear status, and a history of previous tuberculosis, were not associated with clustering. Only cases with cavitary disease were more likely to be clustered than noncavitary ones, which can be explained by the fact that cavitary disease represents the main source of transmission and/or infection and thus can generate clusters, which has been clearly shown in our setting (2).

Tuberculosis remains endemic in Alabama, with relatively few cases occurring among foreign-born individuals (3% in

1994 and 7% in 2000) (Program Data, Division of Tuberculosis Control, Alabama Department of Public Health). Intensive control efforts have resulted in significant declines in rates of disease. The annual case rate dropped by more than half from 12.8 to 5.9 between 1988 and 2001. However, in order to eradicate tuberculosis in Alabama, a better understanding of the dynamics of transmission and dissemination of *M. tuberculosis* is required, especially if ongoing transmission linked to local conditions and behaviors are overlooked or underestimated. Findings from long-term, population-based molecular epidemiologic studies, such as the one presented here, can be a powerful tool to improve tuberculosis control programs locally by focusing control measures on persons at risk for recently acquired infection, as identified by ongoing molecular epidemiologic surveillance. The results of the present study demonstrate that clustering of tuberculosis cases, as a measurement of the likelihood of recent disease transmission in Alabama, is clearly associated with different social factors and risk behaviors. Homelessness, alcohol abuse, younger age, black race, and residence in an urban setting are risk factors for clustering and indicate that the emphasis of tuberculosis intervention programs must be directed toward these population groups. The development of simple algorithms including the described risk factors could help to prospectively predict recent transmission in different subpopulations, particularly when combined with traditional contact investigation methods.

ACKNOWLEDGMENTS

This study was supported through a contract with the Alabama Department of Public Health.

We thank Donna Mulcahy, Nancy Robinson, and Virginia Pruitt from the State Mycobacteria Laboratory for processing and providing culture slants; Stephen W. Duncan for data management assistance; Karen B. Fowler for data analysis assistance; and deNay Kirkpatrick for assistance in the preparation of the manuscript.

REFERENCES

1. **Alland, D., G. E. Kalkut, A. R. Moss, R. A. McAdam, J. A. Hahn, W. Bosworth, E. Drucker, and B. R. Bloom.** 1994. Transmission of tuberculosis in New York City: an analysis by DNA fingerprinting and conventional epidemiologic methods. N. Engl. J. Med. **330:**1710–1716.
2. **Bailey, W. C., L. B. Gerald, M. E. Kimerling, D. Redden, N. Brook, F. Bruce, S. Tang, S. Duncan, C. M. Brooks, and N. E. Dunlap.** 2002. Predictive model to identify positive tuberculosis skin test results during contact investigations. JAMA **287:**996–1002.
3. **Barnes, P. F., Z. Yang, S. Preston-Martin, J. M. Pogoda, B. E. Jones, M. Otaya, K. D. Eisenach, L. Knowles, S. Harvey, and M. D. Cave.** 1997. Patterns of tuberculosis transmission in central Los Angeles. JAMA **278:** 1159–1163.
4. **Bauer, J., A. B. Andersen, K. Kremer, and H. Miorner.** 1999. Usefulness of spoligotyping to discriminate IS*6110* low-copy-number *Mycobacterium tuberculosis* complex strains cultured in Denmark. J. Clin. Microbiol. **37:**2602–2606.
5. **Bishai, W. R., N. M. Graham, S. Harrington, D. S. Pope, N. Hooper, J. Astemborski, L. Sheely, D. Vlahov, G. E. Glass, and R. E. Chaisson.** 1998. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. JAMA **280:**1679–1684.
6. **Borgdorff, M. W., M. A. Behr, N. J. Nagelkerke, P. C. Hopewell, and P. M. Small.** 2000. Transmission of tuberculosis in San Francisco and its association with immigration and ethnicity. Int. J. Tuberc. Lung Dis. **4:**287–294.
7. **Braden, C. R., G. L. Templeton, M. D. Cave, S. Valway, I. M. Onorato, K. G. Castro, D. Moers, Z. Yang, W. W. Stead, and J. H. Bates.** 1997. Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. J. Infect. Dis. **175:**1446–1452.
8. **Centers for Disease Control and Prevention.** Case definitions for public health surveillance. 1990. Morb. Mortal. Wkly. Rep. **39:**41–42.
9. **Chevrel-Dellagi, D., A. Abderrahman, R. Haltiti, H. Koubaji, B. Gicquel, and K. Dellagi.** 1993. Large-scale DNA fingerprinting of *Mycobacterium tuberculosis* strains as a tool for epidemiological studies of tuberculosis. J. Clin. Microbiol. **31:**2446–2450.
10. **Chin, D. P., K. DeRiemer, P. M. Small, A. P. de Leon, R. Steinhart, G. F. Schecter, C. L. Daley, A. R. Moss, E. A. Paz, R. M. Jasmer, C. B. Agasino, and P. C. Hopewell.** 1998. Differences in contributing factors to tuberculosis incidence in U.S.-born and foreign-born persons. Am. J. Respir. Crit. Care Med. **158:**1797–1803.
11. **Cronin, W. A., J. E. Golub, M. J. Lathan, L. N. Mukasa, N. Hooper, J. H. Razeq, N. G. Baruch, D. Mulcahy, W. H. Benjamin, L. S. Magder, G. T. Strickland, and W. R. Bishai.** 2002. Molecular epidemiology of tuberculosis in a low- to moderate-incidence state: are contact investigations enough? Emerg. Infect. Dis. **8:**1271–1279.
12. **Cronin, W. A., J. E. Golub, L. S. Magder, N. G. Baruch, M. J. Lathan, L. N. Mukasa, N. Hooper, J. H. Razeq, D. Mulcahy, W. H. Benjamin, and W. R. Bishai.** 2001. Epidemiologic usefulness of spoligotyping for secondary typing of *Mycobacterium tuberculosis* isolates with low copy numbers of IS*6110*. J. Clin. Microbiol. **39:**3709–3711.
13. **Dale, J. W., H. Al-Ghusein, S. Al-Hashmi, P. Butcher, A. L. Dickens, F. Drobniewski, K. J. Forbes, S. H. Gillespie, D. Lamprecht, T. D. McHugh, R. Pitman, N. Rastogi, A. T. Smith, C. Sola, and H. Yesilkaya.** 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110*. J. Bacteriol. **185:**2555–2562.
14. **Diaz, R., R. Gomez, E. Restrepo, R. Rumbaut, J. Sevy-Court, J. A. Valdivia, and D. van Soolingen.** 2001. Transmission of tuberculosis in Havana, Cuba: a molecular epidemiological study by IS*6110* restriction fragment length polymorphism typing. Mem. Inst. Oswaldo Cruz **96:**437–443.
15. **Diel, R., S. Schneider, K. Meywald-Walter, C. M. Ruf, S. Rusch-Gerdes, and S. Niemann.** 2002. Epidemiology of tuberculosis in Hamburg, Germany: long-term population-based analysis applying classical and molecular epidemiological techniques. J. Clin. Microbiol. **40:**532–539.
16. **Dobbs, K. G., K. H. Lok, F. Bruce, D. Mulcahy, W. H. Benjamin, and N. E. Dunlap.** 2001. Value of *Mycobacterium tuberculosis* fingerprinting as a tool in a rural state surveillance program. Chest **120:**1877–1882.
17. **Dunlap, N. E.** 2000. The use of RFLP as a tool for tuberculosis control: utility or futility? Int. J. Tuberc. Lung Dis. **4:**S134–S138.
18. **Edlin, B. R., J. I. Tokars, M. H. Grieco, J. T. Crawford, J. Williams, E. M. Sordillo, K. R. Ong, J. O. Kilburn, S. W. Dooley, K. G. Castro, et al.** 1992. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. N. Engl. J. Med. **326:**1514–1521.
19. **Ellis, B. A., J. T. Crawford, C. R. Braden, S. J. McNabb, M. Moore, and S. Kammerer.** 2002. Molecular epidemiology of tuberculosis in a sentinel surveillance population. Emerg. Infect. Dis. **8:**1197–1209.
20. **Fomukong, N., M. Beggs, H. el Hajj, G. Templeton, K. Eisenach, and M. D. Cave.** 1997. Differences in the prevalence of IS6110 insertion sites in *Mycobacterium tuberculosis* strains: low and high copy number of IS*6110*. Tuberc. Lung Dis. **78:**109–116.
21. **Frieden, T. R., P. I. Fujiwara, R. M. Washko, and M. A. Hamburg.** 1995. Tuberculosis in New York City: turning the tide. N. Engl. J. Med. **333:**229–233.
22. **Genewein, A., A. Telenti, C. Bernasconi, C. Mordasini, S. Weiss, A. M. Maurer, H. L. Rieder, K. Schopfer, and T. Bodmer.** 1993. Molecular approach to identifying route of transmission of tuberculosis in the community. Lancet **342:**841–844.
23. **Goyal, M., N. A. Saunders, J. D. van Embden, D. B. Young, and R. J. Shaw.** 1997. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. J. Clin. Microbiol. **35:**647–651.
24. **Hermans, P. W., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, D. Frommel, et al.** 1995. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J. Infect. Dis. **171:**1504–1513.
25. **Jasmer, R. M., J. A. Hahn, P. M. Small, C. L. Daley, M. A. Behr, A. R. Moss, J. M. Creasman, G. F. Schecter, E. A. Paz, and P. C. Hopewell.** 1999. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. Ann. Intern. Med. **130:**971–978.
26. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607–2618.
27. **Lok, K. H., W. H. Benjamin, Jr., M. E. Kimerling, V. Pruitt, D. Mulcahy, N. Robinson, N. B. Keenan, and N. E. Dunlap.** 2002. Molecular typing of *Mycobacterium tuberculosis* strains with a common two-band IS*6110* pattern. Emerg. Infect. Dis. **8:**1303–1305.
28. **McConkey, S. J., M. Williams, D. Weiss, H. Adams, M. D. Cave, Z. Yang, T. Lindner, and T. C. Bailey.** 2002. Prospective use of molecular typing of *Mycobacterium tuberculosis* by use of restriction fragment-length polymorphism in a public tuberculosis-control program. Clin. Infect. Dis. **34:**612–619.
29. **Miettinen, O. S.** 1985. Theoretical epidemiology. John Wiley & Sons, Inc., New York, N.Y.
30. **Small, P. M., P. C. Hopewell, S. P. Singh, A. Paz, J. Parsonnet, D. C. Ruston,**

G. F. Schecter, C. L. Daley, and G. K. Schoolnik. 1994. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N. Engl. J. Med. **330:**1703–1709.
31. **Soini, H., X. Pan, A. Amin, E. A. Graviss, A. Siddiqui, and J. M. Musser.** 2000. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J. Clin. Microbiol. **38:**669–676.
32. **Tornieporth, N. G., Y. Ptachewich, N. Poltoratskaia, B. S. Ravi, M. Katapadi, J. J. Berger, M. Dahdouh, S. Segal-Maurer, A. Glatt, R. Adamis, C. Lerner, D. Armstrong, M. Weiner, R. D'Amato, T. Kiehn, S. Lavie, M. Y. Stoeckle, and L. W. Riley.** 1997. Tuberculosis among foreign-born persons in New York City, 1992–1994: implications for tuberculosis control. Int. J. Tuberc. Lung Dis. **1:**528–535.
33. **van Embden, J. D., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnick, et al.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:** 406–409.
34. **van Soolingen, D., M. W. Borgdorff, P. E. de Haas, M. M. Sebek, J. Veen, M. Dessens, K. Kremer, and J. D. van Embden.** 1999. Molecular epidemiology of tuberculosis in The Netherlands: a nationwide study from 1993 through 1997. J. Infect. Dis. **180:**726–736.
35. **van Soolingen, D., P. E. de Haas, P. W. Hermans, P. M. Groenen, and J. D. van Embden.** 1993. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **31:**1987–1995.
36. **Vynnycky, E., N. Nagelkerke, M. W. Borgdorff, D. van Soolingen, J. D. van Embden, and P. E. Fine.** 2001. The effect of age and study duration on the relationship between "clustering" of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. Epidemiol. Infect. **126:**43–62.
37. **Weis, S. E., J. M. Pogoda, Z. Yang, M. D. Cave, C. Wallace, M. Kelley, and P. F. Barnes.** 2002. Transmission dynamics of tuberculosis in Tarrant county, Texas. Am. J. Respir. Crit. Care Med. **166:**36–42.
38. **Yang, Z. H., P. E. de Haas, C. H. Wachmann, D. van Soolingen, J. D. van Embden, and A. B. Andersen.** 1995. Molecular epidemiology of tuberculosis in Denmark in 1992. J. Clin. Microbiol. **33:**2077–2081.