Molecular Epidemiology of *Mycobacterium tuberculosis* among South African Gold Miners

Barun Mathema[1,2], James J. Lewis[3], Jeremy Connors[2], Violet N. Chihota[4,5], Elena Shashkina[2], Minty van der Meulen[4], Edward A. Graviss[6], Ngan P. Ha[6], Barry N. Kreiswirth[2], Alison D. Grant[3], Katherine L. Fielding[3], Susan E. Dorman[7], and Gavin J. Churchyard[3,4,5]

[1]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York; [2]Tuberculosis Center, Public Health Research Institute, Newark, New Jersey; [3]London School of Hygiene and Tropical Medicine, London, United Kingdom; [4]Aurum Institute, Johannesburg, South Africa, and [5]School of Public Health, University of Witwatersrand, Johannesburg, South Africa; [6]Houston Methodist Research Institute, Houston, Texas; and [7]Johns Hopkins University School of Medicine, Baltimore, Maryland

Abstract

Rationale: HIV-associated tuberculosis remains a major health problem among the gold-mining workforce in South Africa. We postulate that high levels of recent transmission, indicated by strain clustering, are fueling the tuberculosis epidemic among gold miners.

Objectives: To combine molecular and epidemiologic data to describe *Mycobacterium tuberculosis* genetic diversity, estimate levels of transmission, and examine risk factors for clustering.

Methods: We conducted a cross-sectional study of culture-positive *M. tuberculosis* isolates in 15 gold mine shafts across three provinces in South Africa. All isolates were subject IS*6110*-based restriction fragment length polymorphisms, and we performed spoligotyping analysis and combined it with basic demographic and clinical information.

Measurements and Main Results: Of the 1,602 *M. tuberculosis* patient isolates, 1,240 (78%) had genotyping data available for analysis. A highly diverse bacillary population was identified, comprising a total of 730 discrete genotypes. Four genotypic families (Latin American Mediterranean spoligotype family; W-Beijing; AH or X; and T1–T4) accounted for over 50% of all strains. Overall, 45% (560/1,240) of strains were genotypically clustered. The minimum estimate for recent transmission ($n-1$ method) was 32% (range, 27–34%). There were no individual-level risk factors for clustering, apart from borderline evidence for being non–South African and having self-reported HIV infection.

Conclusions: The high *M. tuberculosis* genetic diversity and lack of risk factors for clustering are indicative of a universal risk for disease among gold miners and likely mixing with nonmining populations. Our results underscore the urgent need to intensify interventions to interrupt transmission across the entire gold-mining workforce in South Africa.

Keywords: gold mines, molecular epidemiology, South Africa, tuberculosis

An increase in tuberculosis (TB) cases coincident with the HIV epidemic has been reported worldwide. The increase in TB notification rates has been most marked in sub-Saharan Africa, where 75% of coinfected individuals live and where HIV-associated TB remains a major health problem (1, 2). The situation is particularly dire among the gold-mining workforce

in South Africa; in this population, the TB rate surpassed 4,000/100,000/yr in 1999, partly due to the high prevalence of HIV coinfection (which was 29% in 2001) and silicosis (3–6). Although the contribution of these factors to TB case rates is clear, the TB transmission dynamics in this large and highly mobile population have not been well described.

Molecular methods with traditional epidemiologic data have expanded the ability to explore TB epidemics. In low-incidence settings, molecular epidemiology has been used to investigate TB outbreaks and laboratory cross-contamination (7–9). In endemic settings, community-based studies have revealed high strain diversity, resolved TB recurrence into exogenous reinfection or endogenous reactivation, and, in some instances, identified clinical characteristics associated with specific *Mycobacterium tuberculosis* strains (10–15). In addition, in some studies, researchers have estimated the level of, and factors associated with, recent transmission by using genotypic clustering as a proxy for recent transmission (11–13). Identifying correlates of clustering can help to identify at-risk groups to facilitate targeted interventions.

The Thibela TB Study investigators evaluated the impact of community-wide isoniazid preventive therapy (IPT) in 15 groups of gold-mining shafts in three provinces of South Africa (16). During the Thibela TB Study, sputum specimens were collected from consenting adults with suspected pulmonary TB. Cultured *M. tuberculosis* isolates were subjected to molecular analysis. To better understand the TB epidemiology in this population, we combined molecular data with concomitantly collected epidemiologic data to describe the genetic diversity of the bacillary population, estimate recent transmission, and examine individual and ecological factors associated with strain clustering in the context of the current control programs.

Methods

Study Population
The initiation of study activities at the 15 mines from three provinces in South Africa (Gauteng, North West, and Free State) was phased in over 33 months, and the duration of recruitment varied, largely depending on population size. Participants in this substudy were consecutive samples of adults with suspected pulmonary TB, based on clinical and/or radiologic findings, who were identified in participating mine shafts between June 2006 and February 2010. The participants were identified as follows: (*1*) self-presentation at mine health services, (*2*) referral to mine health services following routine radiologic chest screening, and (*3*) screening for active TB prior to or during IPT offered as part of the Thibela TB Study (16, 17). Individuals who self-reported a prior history of TB were excluded if they had been identified by the mine health services before November 2008. Those identified from November 2008 onward through screening prior to IPT were included, regardless of TB history. Each participant provided one spot sputum specimen. Information on demographics, self-reported HIV status, and TB history were collected in a participant interview at the time of sputum collection (17). The ethics boards of the University of the KwaZulu-Natal, the London School of Hygiene and Tropical Medicine, and the Public Health Research Institute (New Jersey Medical School, Newark, NJ) approved this study.

Laboratory Procedures
Mycobacteriologic tests, as described previously (18), were performed at the National Health Laboratory Services laboratory in Braamfontein, Johannesburg, South Africa. Positive cultures were confirmed as *M. tuberculosis* using an anti-MPB64 monoclonal antibody assay (TAUNS Laboratories, Numazu, Japan). Isoniazid and rifampicin susceptibility testing was performed on *M. tuberculosis* isolates using the BACTEC MGIT culture system (BD Biosciences, San Jose, CA). Primary cultures of *M. tuberculosis* were initially stored at 2–8°C and later subcultured into 7H9 liquid medium supplemented with oleic acid, albumin, dextrose, and catalase and 15% glycerol for storage at −80°C.

M. tuberculosis stocks were subcultured on Löwenstein-Jensen slants and grown at 37°C for 3–5 weeks. IS*6110*-based restriction fragment length polymorphism (RFLP) analysis was performed on each isolate as described elsewhere (19). RFLP patterns were analyzed using Bio Image software (Bio Image Systems, Jackson, MI). *M. tuberculosis* isolates with identical DNA hybridization patterns were considered to be the same and assigned according to a previously described nomenclature system (9). A strain family was a group that exhibited similar but nonidentical IS*6110* patterns suggestive of relatedness by descent (e.g., Beijing). Spoligotyping was performed to further define each isolate and classify it according to the fourth international spoligotyping database (20, 21). In addition, *M. tuberculosis* strains were assigned to discrete, synonymous single-nucleotide polymorphism (sSNP)–based phylogenetic lineages (I–VIII and II.A) on the basis of RFLP patterns previously described (22).

Definitions
M. tuberculosis isolates were characterized using IS*6110*-RFLP and spoligotyping. The Thibela TB Study baseline survey data suggested that mixing between miners in different provinces was minimal (16). Hence, it was assumed that two isolates sharing the same RFLP spoligotype pattern most likely represented recent transmission if the isolates came from two participants working in the same province. An orphan strain was defined as an isolate with an RFLP spoligotype pattern that occurred in only one patient in one province within the study dataset. Strain clusters were defined as more than one occurrence of a specific strain (RFLP spoligotype) in different individuals from the same province during the study period.

The parameter "recent transmission" was estimated using the $n - 1$ method: recent transmission = (number of clustered patients − number of clusters)/total number of patients (23). In ecological analyses at the mine shaft–group level, the proportion of clustered strains was compared with "sampling duration," "coverage," and mine health service TB case notification rates within each group of mine shafts. The sampling duration was the number of months between enrollment dates for the first and last study participants in each mine shaft group. Two definitions of "coverage" were used: (*1*) the proportion of *M. tuberculosis* culture–positive isolates that had genotyping data and so were included in the clustering analysis (Definition 1) and (*2*) the ratio of the number of *M. tuberculosis* strains with genotyping data to the number of TB cases notified by mine health services over the same time period corresponding to the

sampling duration (Definition 2). Baseline case notification rates as a predictor for proportion of clustered strains in the ecological analysis were measured over the calendar year prior to the Thibela TB Study baseline survey.

Data Analysis

Statistical analyses were conducted with Stata 12.1 software (StataCorp, College Station, TX). Risk factors for clustering at the individual level were assessed by univariate logistic regression analysis. Correlation of risk factors for strain clustering at the ecological level of the 15 groups of mine shafts in the Thibela TB Study was assessed with Pearson's coefficients. Molecular data were not available for all participants, so associations between a participant's characteristics and whether these data were available for analysis were assessed with Chi-squared tests. Associations between participant characteristics and strain families were also assessed with Chi-squared tests.

Results

Over the time period of this study, there were a total of 5,513 TB cases identified by mine health services. During the same period, we identified 1,602 participants who were *M. tuberculosis* culture–positive; of these, 1,240 had RFLP and spoligotype data (77.4%) and so were suitable for genotypic analysis (Figure 1). Reflecting the gold-mining workforce, these 1,240 strains were isolated predominantly (97.3%) from men (Table 1). The participants' median age was 45 years (interquartile range, 39 to 49 years), and their median tenure in the workforce was 21 years (interquartile range, 14 to 28 years). A prior history of TB was reported by 14.7% of participants. HIV status was self-reported by 450 (36.3%) of 1,240 participants, of whom 196 (43.6%) were HIV-positive. Sixty-two (5.2%) of 1,240 participants were resistant to isoniazid, 7 (0.6%) were resistant to rifampin, and 46 (4.1%) were multidrug-resistant (Table 1). Individuals whose RFLP spoligotype data were unavailable were more likely to have been recruited before 2008 and hence less likely to have a prior history of TB, in line with the study enrollment criteria (*see* Table E1 in the online supplement). They were also less likely to have a cough, more likely to live in
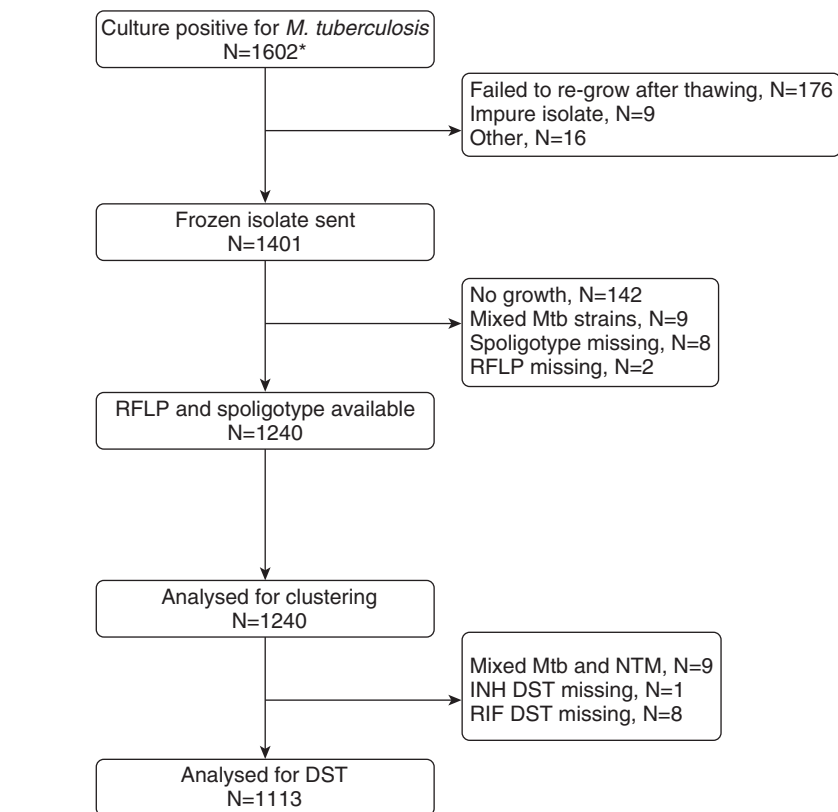

Figure 1. Specimen flowchart. DST = drug susceptibility testing; INH = Isoniazid; Mtb = *Mycobacterium tuberculosis*; NTM = nontuberculosis mycobacteria; RFLP = restriction fragment length polymorphism; RIF = rifampin. *During the time period during which the substudy was operating, there were a total of 5,513 TB cases identified.

a hostel, and more likely to be working in the North West province (Table E1).

Genetic Diversity of *M. tuberculosis* Isolates

The 1,240 isolates came from participating mines in three provinces: 415 from Gauteng (33.5%), 391 from North West (31.5%), and 434 (35.0%) from the Free State (Table 1). All nine discrete sSNP phylogenetic lineages were represented in the bacillary population, with the three most common groups being VI (41.9%), II (13.9%), and V (10.6%). Using IS6110-RFLP spoligotype classification, a total of 730 discrete genotypes were identified. When work province was added to the definition of clustering, an increase to 845 discrete genotypes ("strains") was realized (Table 2). The strains were grouped into 62 IS6110-RFLP families, with an additional 11 orphan IS6110-RFLP patterns. The four most common families, designated by the letter corresponding to their IS6110-RFLP

(spoligotype), were CC (Latin American Mediterranean [LAM]), W-Beijing (Beijing), AH (X family), and BM (T1–T4). These families comprised 263 (21.2%), 168 (13.6%), 129 (10.4%), and 91 (7.3%) isolates, respectively, and together accounted for over half (52.5%) of all isolates. The diversity within each of the major genotypic families varied considerably (Figure 2). There were 170 variants in the CC family, 69 in W-Beijing, 49 in BM, and 34 in AH. Within these genotypic families, the extent of diversity (i.e., the fraction of the number of genotypes to the total number of strains within the family) was highest within the CC (170 of 263 [0.65]) and BM (49 of 91 [0.54]) families, followed by W-Beijing (69 of 168 [0.41]) and AH (34 of 129 [0.26]).

Estimation of Transmission

Genotypic strain clustering was used as a proxy for recent transmission. Of 1,240 characterized isolates, 560 (45.2%) were represented in 165 clusters (Table 2). The

Table 1. Risk factors for being in a cluster, as determined by RFLP, spoligotype, and province

	Overall (N = 1,240)		Clustered (n = 560)		Odds ratio (95% CI)	P-value (likelihood ratio test)
	n	Column %	n	Row %		
Sex						
Male	1,201	97.3%	542	45.1%	1	0.970
Female	33	2.7%	15	45.5%	1.01 (0.51, 2.03)	
Age group, yr						
18–34	160	13.0%	72	45.0%	1	0.998
35–44	451	36.5%	202	44.8%	0.99 (0.69, 1.42)	
45–54	526	42.6%	237	45.1%	1.00 (0.70, 1.43)	
55+	98	7.9%	45	45.9%	1.04 (0.63, 1.72)	
Years in industry						
0–9	198	16.3%	89	44.9%	1	0.699
10–19	306	25.2%	146	47.7%	1.12 (0.78, 1.60)	
20–29	448	36.8%	194	43.3%	0.94 (0.67, 1.31)	
30+	264	21.7%	119	45.1%	1.01 (0.69, 1.46)	
Country of origin						
South Africa	659	53.3%	276	41.9%	1	0.091
Lesotho	425	34.4%	205	48.2%	1.29 (1.01, 1.65)	
Mozambique	102	8.2%	53	52.0%	1.50 (0.99, 2.28)	
Other	51	4.1%	24	47.1%	1.23 (0.70, 2.18)	
Residence						
Hostel	743	60.1%	334	45.0%	1	0.923
Other	493	39.9%	223	45.2%	1.01 (0.80, 1.27)	
Previous history of TB						
No	1,056	85.3%	472	44.7%	1	0.437
Yes	182	14.7%	87	47.8%	1.13 (0.83, 1.55)	
HIV status (by self-report)						
Negative	254	56.4%	104	40.9%	1	0.089
Positive	196	43.6%	96	49.0%	1.38 (0.95, 2.02)	
Smear result						
Negative	620	50.0%	273	44.0%	1	0.424
Positive	620	50.0%	287	46.3%	1.10 (0.88, 1.37)	
Cough						
No	327	26.8%	138	42.2%	1	0.440
Yes for ≤ 2 wk	168	13.7%	79	47.0%	1.22 (0.84, 1.77)	
Yes for > 2 wk	727	59.5%	335	46.1%	1.17 (0.90, 1.52)	
Hospitalized*						
No	423	78.6%	193	45.6%	1	0.807
Yes	115	21.4%	51	44.3%	0.95 (0.63, 1.44)	
Province						
Gauteng	415	33.5%	197	47.5%	1	0.040
North-West	391	31.5%	156	39.9%	0.73 (0.56, 0.97)	
Free State	434	35.0%	207	47.7%	1.01 (0.77, 1.32)	
DST						
INH sus, RIF sus	998	89.7%	457	45.8%	1	0.878
INH res, RIF sus	62	5.6%	28	45.2%	0.97 (0.58, 1.63)	
INH sus, RIF res	7	0.6%	4	57.1%	1.58 (0.35, 7.09)	
INH res, RIF res	46	4.1%	23	50.0%	1.18 (0.66, 2.14)	

*Inpatient at a hospital or health clinic within the preceding 2 years.
Definition of abbreviations: DST = drug susceptibility tests; INH = isoniazid; res = resistant; RFLP = restriction fragment length polymorphism; RIF = rifampin; sus = susceptible; TB = tuberculosis.

median cluster size was two isolates, and there were five clusters, each of which comprised more than 10 isolates (17–23 isolates each). Three of these clusters were from one IS*6110*-RFLP spoligotype classification that was represented in all three provinces, and two were from one IS*6110*-RFLP spoligotype classification that was represented in two provinces (with seven isolates with this classification in the third province). Recent transmission for the overall population was estimated excluding or including the province of work in the definition of clustering. The estimate excluding the province of work was 31.9% ([range estimate, 560–165]/1,240), and estimate including the province of work was 41.1% ([range estimate, 677–167]/1,240) (Table 2). Strain clustering in each province ranged from 39.9% to 47.7%, and estimates of recent transmission varied from 27.4% to 34.1% (Table 2).

Risk Factors for Clustering

Individual level. Univariate logistic regression analyses were used to examine risk factors for being in a cluster (Table 1). There was some evidence that miners from the North West province were less likely to have clustered strains than miners from the other two provinces (odds ratio [OR], 0.73; 95% confidence interval [CI], 0.56–0.97 [compared with Gauteng province]). There was an increased odds of strain clustering

Table 2. Strain, IS6110-RFLP families and cluster results by province*

	Gauteng (n = 415)	North West (n = 391)	Free State (n = 434)	Overall, including province in strain definition (n = 1,240)	Overall, excluding province from strain definition (n = 1,240)
Variants	275	284	286	845	730
IS6110-RFLP families	46	45	45	62	62
Novel IS6110-RFLP patterns	5	4	2	11	11
Clustered	197 (47.5%)	156 (39.9%)	207(47.7%)	560 (45.2%)	677 (54.6%)
Clusters	57	49	59	165	167
Recent transmission,[†] %	33.7%	27.4%	34.1%	31.9%	41.1%

*"Strain" defined based on same IS6110-RFLP, spoligotype, and province.
[†]Estimated using the n − 1 method.
Definition of abbreviation: RFLP = restriction fragment length polymorphism.

for non–South Africans compared with South Africans (OR, 1.32; 95% CI, 1.06–1.66), as well as borderline evidence for those who self-reported having HIV, compared with not having HIV, infection among those who knew and were willing to report their HIV status (OR, 1.38; 95% CI, 0.95–2.02).

Mine shafts group level. Within the 15 groups of mine shafts in the Thibela TB
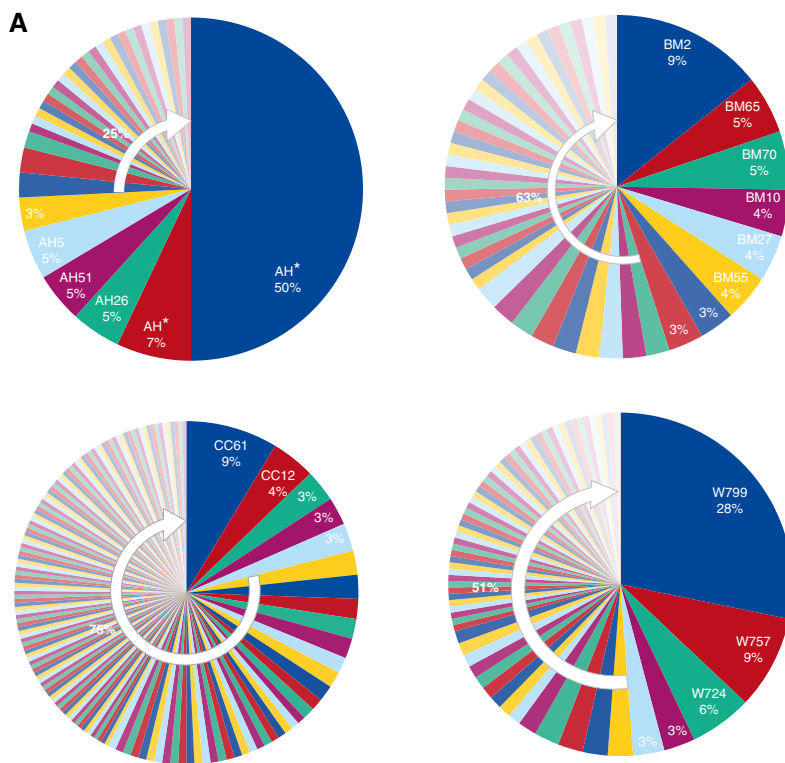
Study, the proportion of strains that were clustered varied between 27.3% and 63.6% (Figure 3). There was borderline evidence that the proportion of strains clustered in each group of mine shafts was correlated with increasing levels of sampling coverage (Definition 1: $R^2 = 0.25$, $P = 0.060$; Definition 2: $R^2 = 0.20$, $P = 0.095$) (Figure 3) and with decreasing duration of sampling ($R^2 = 0.20$, $P = 0.097$). There was no evidence of a correlation between clustering and case notification rates ($R^2 = 0.01$, $P = 0.688$). Within the 15 groups of mine shafts, estimates of recent transmission varied between 0% and 33.1% (median, 19.2%).

Strain Family Epidemiology

Isolates belonging to the four most common strain families were more likely to be from participants who reported fewer years working in the industry (Table E2). Strain family was also associated with the participant's country of origin and province. Of note, 24 (51.1%) of 47 of strain W799 were isolated from miners from Lesotho. Drug resistance was associated with strain family, with resistance to both isoniazid and rifampicin being most common in the AH strain family. Clustering was also more common among strain families AH (X family, low-copy clade) and W-Beijing than in other strain families ($P < 0.001$).

Discussion

We used molecular epidemiology to determine the genetic diversity of circulating *M. tuberculosis* strains and to provide estimates of recent transmission and risk factors for clustering among gold miners in South Africa. We found a high level of

Figure 2. Genotypic diversity of *M. tuberculosis* within the four most prevalent IS6110-based restriction fragment length polymorphism (RFLP) and spoligotype strain families. (A) Pie chart representing the distribution of specific genotypes within each of the four strain families. (B) Table illustrating the extent of diversity within each strain family. *AH strains with two distinct spoligotypes.
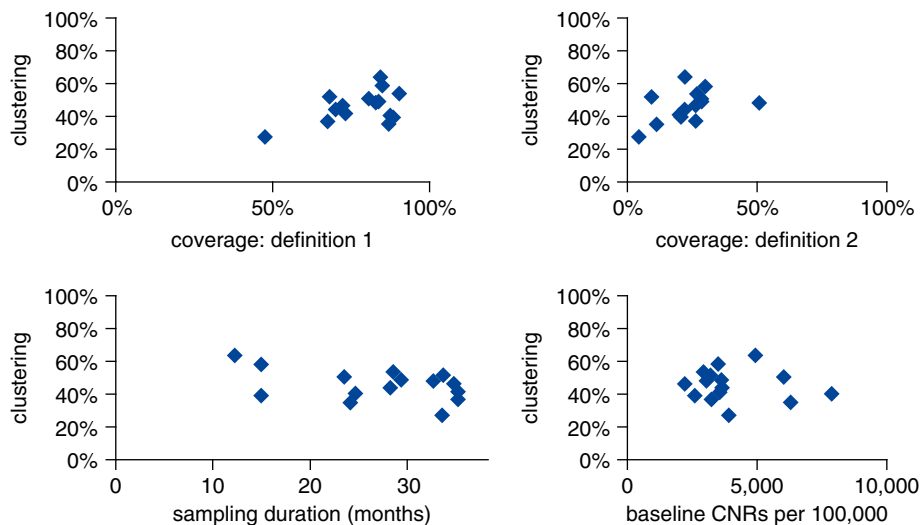
	AH	BM	CC	W-Beijing
Genotypes	34	49	170	69
Patient isolates	129	91	263	168
Diversity Index	0.26	0.54	0.65	0.41

Figure 3. Ecological analysis of factors associated with clustering in the 15 groups of mine shafts in the Thibela TB trial. Coverage: Definition 1 = isolates included in clustering analysis as a percentage of all isolates whose culture was positive for *M. tuberculosis*. Coverage: Definition 2 = ratio of number of isolates included in clustering analysis to the number of cases identified for TB over the same time period. CNR = case notification rate.

genetic diversity within the bacillary population. Our molecular data suggest that the minimum estimated proportion of TB due to recent transmission was approximately one-third, considerably lower than previous reports (11, 12, 24, 25). Despite the extremely high TB incidence in the study (3), we did not detect strong factors associated with genetic clustering at the individual or ecological level.

Researchers in molecular epidemiology studies in Southern Africa have reported on the genetic diversity of circulating *M. tuberculosis* strains (11, 12, 26–29). All major *M. tuberculosis* phylogenetic lineages were represented in these studies. The CC (LAM), W-Beijing (Beijing), and AH (X) families predominated, consistent with previous reports (12, 27–29). In low-incidence settings, where TB is often due to reactivation from a remote infection, genetic diversity of *M. tuberculosis* isolates typically reflects the extent of ethnic heterogeneity and the mobility of the population. In the present study, we found a higher level of strain diversity (845 discrete genotypes) than previously reported in other South African populations (11, 12, 30, 31). This high diversity supports our assessment that two strains with identical genotypes are likely related. Unlike other studies of many South African communities in which TB epidemiology has been described (12, 15), we focused

on a relatively more mobile population. The gold-mining workforce represents individuals from different regions of South Africa, as well as considerable numbers from neighboring countries such as Lesotho and Mozambique (32). Thus, our bacillary population likely represents a mixture of strains actively circulating within the mining communities and, to a lesser extent, those imported by the migrant labor force. For instance, over half of a large W-Beijing strain (strain W799) cluster (n = 47) was recovered from gold miners originating from Lesotho. Although it is possible that strain W799 was transmitted within the mining community, these strains were recovered from miners working from different mining shafts and/or regions and likely represent a closer network, either within the mining system or in the originating communities. In addition, high genetic variation within major strain families (Figure 2) and its association with fewer years in the industry (Table E2) strengthen the argument that these strains are endemic in South Africa and southern Africa and likely are being imported into the study region. It is worth noting that researchers in studies from different regions in South Africa have noted variation in the predominant genotype families seen in our study (25, 27, 28, 33).

The rate of genotypic biomarker change (e.g., IS*6110*) is often proportionate

to the extent of bacillary replication (34). Therefore, similar nonidentical genotypes typically accumulate in a community over time, as strains transmit from patient to patient. We found considerable genetic variation within the four major genotype families that collectively accounted for over 50% of the isolates (Figure 3). The most diverse CC and BM families, and, to a lesser extent, the W-Beijing family, are likely endemic in southern Africa (12, 35, 36). The W-Beijing family has been reported to be present in nearly all locales in southern Africa and is thought to have been imported within the past century (37, 38). In contrast, the AH family showed the least diversity, which is partly explained by their characteristically low copy number of IS*6110* elements. Our use of RFLP spoligotyping helped to provide some resolution. The AH strains have been reported to have caused large drug-resistant outbreaks in the Western Cape of South Africa (29, 39), a finding consistent with ours (Table E2).

Molecular clustering has been used to estimate the levels of recent transmission in different epidemiologic settings (7, 11, 23, 40). The validity of such estimates is heavily dependent on the sampling proportion of total TB cases, the duration of sampling (41, 42), and the discriminatory power of genotyping method(s) (43). In previous studies done in South Africa, researchers have reported a proportion of clustered genotypes in the range of 70% (11, 12, 30, 31) and estimated that 50–60% of cases were due to recent infection. We found considerably less clustering (~45%) and recent infection (~32%) among our gold-mining study populations. In contrast to other previously described communities that are relatively static, miners originate from and visit many regions and nearby countries. Additionally, our results are in contrast to estimates of 70–99% of active disease being due to reinfection, as indicated by mathematical models of the same gold-mining populations reported here (44). However, caution must be used when directly comparing estimates of recent transmission based on identical molecular data versus fitted estimates of disease due to reinfection from mathematical models. In addition, estimates of recent transmission are restricted to only the study population and catchment area and do not preclude disease due to recent transmission from other

mining regions or disease not related to occupational exposure (i.e., from nearby communities). Unlike other studies that typically report the fraction of strains analyzed from total culture-positive cases, we note that, despite capturing 78% of all culture-positive *M. tuberculosis* strains, our sampling fraction of total TB cases examined from those that were identified during the study was low (Figure 3). Given the relatively low sampling fraction and population heterogeneity, our estimates of recent transmission are likely conservatively low.

Identifying risk factors associated with recent transmission is a priority for TB control programs. As such, patient factors associated with molecular clustering have yielded insight into transmission at the individual (e.g., HIV infection, substance abuse) and community (e.g., household, hospital) levels (9, 23, 45). In our study, we examined individual-level factors associated with clustering, using basic clinical, demographic, geographic, and employment history data. Other than mining region, no clear risk factors were identified. Our results contrast with those from a molecular epidemiologic study conducted more than 10 years ago in four South African gold mines where treatment failure was associated with genotypic clustering, whereas multidrug resistance, despite being associated with treatment failure, was negatively associated with clustering (11). The study differences likely reflect the consequence of epidemic levels of TB incidence coupled with greater access to antiretroviral therapy.

The lack of readily identifiable risk factors for clustering can be explained in part by the high force of infection (i.e., the proportion of susceptible individuals who have become infected in a specified period). A high force of infection in these populations would result in considerable rates of primary and secondary TB infection and, given the high HIV prevalence among gold miners (~30%), could explain the high disease incidence. As such, the force of infection is influenced by the prevalence of untreated infectious TB, host susceptibility, and extent of social mixing patterns. Although the exact nature and extent of social interactions is hard to quantify, our molecular clustering data indicate that mixing among miners with nonminers from nearby communities and home visits (e.g., Eastern Cape, Lesotho)

are likely considerable (46). The impact of these interactions and the circular migration of the study population on *M. tuberculosis* genetic diversity may provide a foundation for explaining the epidemics seen in these gold-mining settings. There was weak evidence that coverage (sampling level and duration) at the mine shaft level was associated with the proportion of strain clustering. Taken together, our results suggest that, in an environment with a high force of infection and mobile patient populations, reactivation disease (particularly among HIV-negative individuals), recent infection and reinfection are likely occurring at high rates, thus obscuring the relative contribution of recently transmitted TB within this setting. The confluence of factors dampening detection of recent transmission in this study may explain in part the null results of the cluster-randomized IPT Thibela trial (16). Our study also highlights possible shortcomings of current tools that focus on individuals with disease used to estimate recent transmission of infection in hyperepidemic and/or hyperendemic settings of transient populations.

Our study has a number of limitations. First, patient epidemiologic data linking cases that appeared clonally related were not available. The lack of patient-linked data limits the inference that clustered cases represent direct patient-to-patient linkage or recent spread. These data were available for a study by Godfrey-Faussett and coworkers among South African gold miners, but those researchers were not able to identify hot spots for transmission and direct patient-to-patient linkage was established among only 19% of cases within molecular clusters (11). Second, despite capturing 78% of all *M. tuberculosis* isolates from the study population, our overall sampling of identified TB cases (1,240 of 5,513) was 22.5% (Figure 3). In addition, not all mining companies routinely investigated all miners with suspected pulmonary TB on the basis of sputum cultures. At one mining company that routinely investigated all persons with suspected TB on the basis of cultures, 4,268 TB cases between 2002 and 2008 were diagnosed, of which 2,250 (52.7%) were culture-positive for *M. tuberculosis* (47). The low proportion of TB episodes with strains available for analysis would tend to underestimate clustering,

recent transmission, and associated risk factors. However, our good sampling of infectious cases (i.e., culture-positive) likely reflects the overall characteristics of circulating strains. Third, we had limited data on participants' HIV status and on the movement of miners within and outside mining regions. Although HIV coinfection has a profound impact on the natural course of TB, previous studies have not found HIV to be associated with molecular clustering (11, 12, 45).

Despite implementation of TB control measures among South African gold miners, TB case identification rates remain very high. Although the elevated risk for disease among gold miners may point to the need for more intensified interventions across the entire community, a better understanding of the dynamics of transmission is of paramount necessity in designing control programs (48). However, it is clear from this study that detailing transmission pathways within environments with very high force of infection is challenging. Such environments may require additional tools with more granular resolution, such as whole-genome analysis. Such data, coupled with traditional epidemiology and social network analysis (26, 49), may highlight hot spots of TB transmission as portals for strain introduction and dissemination, thereby providing much-needed strategies for improved intervention. Finally, our results underscore the need for strengthening of active case identification and ensuring the successful treatment and cure of all patients with TB. ■

References

1 World Health Organization (WHO). Global tuberculosis control: a short update to the 2009 report. Geneva: WHO; 2009.

2 Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med* 2003;163: 1009–1021.

3 Lewis JJ, Charalambous S, Day JH, Fielding KL, Grant AD, Hayes RJ, Corbett EL, Churchyard GJ. HIV infection does not affect active case finding of tuberculosis in South African gold miners. *Am J Respir Crit Care Med* 2009;180:1271–1278.

4 Corbett EL, Churchyard GJ, Clayton TC, Williams BG, Mulder D, Hayes RJ, De Cock KM. HIV infection and silicosis: the impact of two potent risk factors on the incidence of mycobacterial disease in South African miners. *AIDS* 2000;14:2759–2768.

5 Corbett EL, Churchyard GJ, Charalambos S, Samb B, Moloi V, Clayton TC, Grant AD, Murray J, Hayes RJ, De Cock KM. Morbidity and mortality in South African gold miners: impact of untreated disease due to human immunodeficiency virus. *Clin Infect Dis* 2002;34: 1251–1258.

6 Glynn JR, Murray J, Bester A, Nelson G, Shearer S, Sonnenberg P. Effects of duration of HIV infection and secondary tuberculosis transmission on tuberculosis incidence in the South African gold mines. *AIDS* 2008;22:1859–1867.

7 Bifani PJ, Mathema B, Liu Z, Moghazeh SL, Shopsin B, Tempalski B, Driscol J, Frothingham R, Musser JM, Alcabes P, *et al*. Identification of a W variant outbreak of *Mycobacterium tuberculosis* via population-based molecular epidemiology. *JAMA* 1999;282: 2321–2327.

8 Braden CR, Templeton GL, Stead WW, Bates JH, Cave MD, Valway SE. Retrospective detection of laboratory cross-contamination of *Mycobacterium tuberculosis* cultures with use of DNA fingerprint analysis. *Clin Infect Dis* 1997;24:35–40.

9 Mathema B, Bifani PJ, Driscoll J, Steinlein L, Kurepina N, Moghazeh SL, Shashkina E, Marras SA, Campbell S, Mangura B, *et al*. Identification and evolution of an IS*6110* low-copy-number *Mycobacterium tuberculosis* cluster. *J Infect Dis* 2002;185:641–649.

10 Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, Stepniewska K, Huyen MN, Bang ND, Loc TH, *et al*. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* 2008;4: e1000034.

11 Godfrey-Faussett P, Sonnenberg P, Shearer SC, Bruce MC, Mee C, Morris L, Murray J. Tuberculosis control and molecular epidemiology in a South African gold-mining community. *Lancet* 2000;356: 1066–1071.

12 Middelkoop K, Bekker LG, Mathema B, Shashkina E, Kurepina N, Whitelaw A, Fallows D, Morrow C, Kreiswirth B, Kaplan G, *et al*. Molecular epidemiology of *Mycobacterium tuberculosis* in a South African community with high HIV prevalence. *J Infect Dis* 2009;200: 1207–1211.

13 Sonnenberg P, Murray J, Glynn JR, Shearer S, Kambashi B, Godfrey-Faussett P. HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. *Lancet* 2001;358:1687–1693.

14 Streicher EM, Warren RM, Kewley C, Simpson J, Rastogi N, Sola C, van der Spuy GD, van Helden PD, Victor TC. Genotypic and phenotypic characterization of drug-resistant *Mycobacterium tuberculosis* isolates from rural districts of the Western Cape Province of South Africa. *J Clin Microbiol* 2004;42:891–894.

15 van Rie A, Warren R, Richardson M, Victor TC, Gie RP, Enarson DA, Beyers N, van Helden PD. Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* 1999; 341:1174–1179.

16 Churchyard GJ, Fielding KL, Lewis JJ, Coetzee L, Corbett EL, Godfrey-Faussett P, Hayes RJ, Chaisson RE, Grant AD; Thibela TB Study Team. A trial of mass isoniazid preventive therapy for tuberculosis control. *N Engl J Med* 2014;370:301–310.

17 Chihota VN, Grant AD, Fielding K, Ndibongo B, van Zyl A, Muirhead D, Churchyard GJ. Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. *Int J Tuberc Lung Dis* 2010;14:1024–1031.

18 Kent PT, Kubica GP. Public health mycobacteriology: a guide for the level III laboratory. Atlanta, GA: Centers for Disease Control and Prevention; 1985.

19 van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, *et al*. Strain identification of *Mycobacterium tuberculosis* DNA by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;31:406–409. (see comments).

20 Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, Allix C, Aristimuño L, Arora J, Baumanis V, *et al*. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 2006;6:23.

21 Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, *et al*. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997; 35:907–914.

22 Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, Graviss EA, Musser JM. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J Infect Dis* 2006;193:121–128.

23 Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, Schecter GF, Daley CL, Schoolnik GK. The epidemiology of tuberculosis in San Francisco: a population-based study using conventional and molecular methods. *N Engl J Med* 1994;330: 1703–1709. (see comments).

24 Richardson M, van Lill SW, van der Spuy GD, Munch Z, Booysen CN, Beyers N, van Helden PD, Warren RM. Historic and recent events contribute to the disease dynamics of Beijing-like *Mycobacterium tuberculosis* isolates in a high incidence region. *Int J Tuberc Lung Dis* 2002;6:1001–1011.

25 Chihota V, Apers L, Mungofa S, Kasongo W, Nyoni IM, Tembwe R, Mbulo G, Tembo M, Streicher EM, van der Spuy GD, *et al*. Predominance of a single genotype of *Mycobacterium tuberculosis* in regions of Southern Africa. *Int J Tuberc Lung Dis* 2007;11:311–318.

26 Gandhi NR, Weissman D, Moodley P, Ramathal M, Elson I, Kreiswirth BN, Mathema B, Shashkina E, Rothenberg R, Moll AP, *et al*. Nosocomial transmission of extensively drug-resistant tuberculosis in a rural hospital in South Africa. *J Infect Dis* 2013;207:9–17.

27 Mlambo CK, Warren RM, Poswa X, Victor TC, Duse AG, Marais E. Genotypic diversity of extensively drug-resistant tuberculosis (XDR-TB) in South Africa. *Int J Tuberc Lung Dis* 2008;12:99–104.

28 Nicol MP, Sola C, February B, Rastogi N, Steyn L, Wilkinson RJ. Distribution of strain families of *Mycobacterium tuberculosis* causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa. *J Clin Microbiol* 2005;43:5779–5781.

29 Warren RM, Victor TC, Streicher EM, Richardson M, van der Spuy GD, Johnson R, Chihota VN, Locht C, Supply P, van Helden PD. Clonal expansion of a globally disseminated lineage of *Mycobacterium tuberculosis* with low IS*6110* copy numbers. *J Clin Microbiol* 2004; 42:5774–5782.

30 Glynn JR, Crampin AC, Traore H, Chaguluka S, Mwafulirwa DT, Alghamdi S, Ngwira BM, Yates MD, Drobniewski FD, Fine PE. Determinants of cluster size in large, population-based molecular epidemiology study of tuberculosis, northern Malawi. *Emerg Infect Dis* 2008;14:1060–1066.

31 Verver S, Warren RM, Munch Z, Vynnycky E, van Helden PD, Richardson M, van der Spuy GD, Enarson DA, Borgdorff MW, Behr MA, *et al*. Transmission of tuberculosis in a high incidence urban community in South Africa. *Int J Epidemiol* 2004;33:351–357.

32 Rees D, Murray J, Nelson G, Sonnenberg P. Oscillating migration and the epidemics of silicosis, tuberculosis, and HIV infection in South African gold miners. *Am J Ind Med* 2010;53:398–404.

33 Pillay M, Sturm AW. Evolution of the extensively drug-resistant F15/ LAM4/KZN strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. *Clin Infect Dis* 2007;45:1409–1414.

34 McEvoy CR, Falmer AA, Gey van Pittius NC, Victor TC, van Helden PD, Warren RM. The role of IS*6110* in the evolution of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 2007;87:393–404.

35 van der Spuy GD, Kremer K, Ndabambi SL, Beyers N, Dunbar R, Marais BJ, van Helden PD, Warren RM. Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. *Tuberculosis (Edinb)* 2009;89:120–125.

36 Victor TC, de Haas PE, Jordaan AM, van der Spuy GD, Richardson M, van Soolingen D, van Helden PD, Warren R. Molecular characteristics and global spread of *Mycobacterium tuberculosis* with a Western Cape F11 genotype. *J Clin Microbiol* 2004;42: 769–772.

37 Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol* 2002;10:45–52.

38 Cowley D, Govender D, February B, Wolfe M, Steyn L, Evans J, Wilkinson RJ, Nicol MP. Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. *Clin Infect Dis* 2008;47:1252–1259.

39 Victor TC, Streicher EM, Kewley C, Jordaan AM, van der Spuy GD, Bosman M, Louw H, Murray M, Young D, van Helden PD, *et al*. Spread of an emerging *Mycobacterium tuberculosis* drug-resistant strain in the Western Cape of South Africa. *Int J Tuberc Lung Dis* 2007;11:195–201.

40 Hu Y, Mathema B, Jiang W, Kreiswirth B, Wang W, Xu B. Transmission pattern of drug-resistant tuberculosis and its implication for tuberculosis control in eastern rural China. *PLoS ONE* 2011;6: e19548.

41 Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. *Am J Epidemiol* 2002;155:565–571.

42 Vynnycky E, Nagelkerke N, Borgdorff MW, van Soolingen D, van Embden JD, Fine PE. 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* 2001;126:43–62.

43 Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev* 2006;19:658–685.

44 Chihota VN, Popane F, Churchyard GJ, Lewis JJ, Fielding KL, Vynnycky E, White RG, Grant AD; Thibela TB Study Team. Community-wide isoniazid preventive therapy among gold miners in South Africa: the Thibela TB study. Presented at the Third South African Tuberculosis Conference. Durban, South Africa; June 12, 2012, Durban, South Africa [accessed 2014 Dec 5]. Available from: http://www.auruminstitute.org/thibela-TB/Thibela-SA-TB-conf-12-06-12.pdf

45 Driver CR, Kreiswirth B, Macaraig M, Clark C, Munsiff SS, Driscoll J, Zhao B. Molecular epidemiology of tuberculosis after declining incidence, New York City, 2001-2003. *Epidemiol Infect* 2007;135: 634–643.

46 Stuckler D, Basu S, McKee M, Lurie M. Mining and risk of tuberculosis in sub-Saharan Africa. *Am J Public Health* 2011;101:524–530.

47 van Halsema CL, Fielding KL, Chihota VN, Lewis JJ, Churchyard GJ, Grant AD. Trends in drug-resistant tuberculosis in a gold-mining workforce in South Africa, 2002-2008. *Int J Tuberc Lung Dis* 2012;16: 967–973.

48 Nardell E, Churchyard G. What is thwarting tuberculosis prevention in high-burden settings? *N Engl J Med* 2011;365:79–81.

49 Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, *et al*. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011;364:730–739.