Molecular Epidemiology of *Mycobacterium tuberculosis* in a South African Community with High HIV Prevalence

Keren Middelkoop,1,2 Linda-Gail Bekker,1,2 Barun Mathema,1 Elena Shashkina,1 Natalia Kurepina,1 Andrew Whitelaw,3,4 Dorothy Fallow,3 Carl Morrow,1 Barry Kreiswirth,4 Gilla Kaplan,3 and Robin Wood5

1Desmond Tutu HIV Centre, Institute of Infectious Diseases and Molecular Medicine, Department of Medicine, and 5Division of Medical Microbiology, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, and 6National Health Laboratory Service, Johannesburg, South Africa; 2Laboratory of Mycobacterial Immunity and Pathogenesis and 3Tuberculosis Center, Public Health Research Institute, University of Medicine and Dentistry of New Jersey, Newark, New Jersey

To explore the relationship between human immunodeficiency virus (HIV) and *Mycobacterium tuberculosis* genotypes, we performed IS6110-based restriction fragment-length polymorphism analysis on *M. tuberculosis* culture specimens from patients with smear-positive tuberculosis in a periurban community in South Africa from 2001 through 2005. Among 151 isolates, 95 strains were identified within 26 families, with 54% clustering. HIV status was associated with *W*-Beijing strains (*P* = .009) but not with clustering per se. The high frequency of clustering suggests ongoing transmission in both HIV-negative and HIV-positive individuals in this community. The strong association between *W*-Beijing and HIV infection may have important implications for tuberculosis control.

Tuberculosis remains a major cause of morbidity and mortality worldwide and in Africa human immunodeficiency virus (HIV) is fueling the epidemic [1]. South Africa, which bears 28% of the global burden of HIV-related tuberculosis, is undergoing rapid urbanization [2], with immigrants to the cities concentrating in poor, crowded periurban townships where both HIV prevalence and tuberculosis incidence rates are high [3]. We have previously described high HIV prevalence in adults (23%) [4] and rapidly escalating tuberculosis notification rates [3] in a periurban township in Cape Town, South Africa. Despite a well-implemented national tuberculosis control program (based on the World Health Organization’s DOTS [directly observed therapy, short course] strategy [5]), the number of tuberculosis cases has escalated at the single community clinic that manages all of this township’s resident patients with tuberculosis. This geographically well-defined community with ~15,000 people of low socioeconomic status living in overcrowded, largely informal dwellings is ideally suited for tuberculosis transmission studies.

The advent of such molecular epidemiological tools as IS6110-based restriction fragment–length polymorphism (RFLP) genotypic analysis of *Mycobacterium tuberculosis* has expanded our ability to investigate and understand tuberculosis [6]. Different *M. tuberculosis* strains have been associated with diverse levels of virulence, immunological responses, epidemic potential, and even drug resistance [6]. However, the relationship between specific host characteristics (such as HIV infection) and *M. tuberculosis* genotypes is poorly understood. To explore the association between HIV infection and circulating *M. tuberculosis* strains, we performed RFLP analysis of *M. tuberculosis* isolates cultured from individuals with smear-positive pulmonary tuberculosis in the above-mentioned study community from 2001 through 2005.

**Methods.** Sputum specimens from patients with acid-fast bacilli smear–positive tuberculosis residing in the study community were collected for genotype analysis from 2001 through 2005. Sputum specimens were obtained from patients in accordance with National Tuberculosis Control Program guidelines [7] and were labeled as study specimens at the clinical site for identification by laboratory personnel. The age, sex, details of clinical diagnosis, clinical outcome, and HIV status of each patient were collected from the clinic’s tuberculosis register and patient folders. The study was approved by the ethics review boards of the University of Cape Town and of the University of Medicine and Dentistry of New Jersey (UMDNJ), and participants provided informed consent.

Sputum specimens were assessed for the presence of *M. tuberculosis* bacilli by means of fluorescent (auramine) microscopy. Isoniazid and rifampicin susceptibility testing was performed on samples from retreatment patients and from patients whose samples still tested positive for acid-fast bacilli after 2
months of treatment [7]. Susceptibility to rifampicin was tested at a concentration of 1.0 μg/mL, and susceptibility to isoniazid was tested at concentrations of 0.1 and 0.2 μg/mL on mycobacterial growth indicator tubes and on Middlebrook 7H11 agar, respectively. Sputum samples that tested positive for acid-fast bacilli were cultured on Lowenstein-Jensen slants at the Tuberculosis Laboratory at the Institute of Infectious Diseases and Molecular Medicine at the University of Cape Town. Isolates that tested positive for *M. tuberculosis* by culture were inoculated in duplicate into 7H9 liquid medium supplemented with oleic acid, albumin, dextrose, and catalase (OADC) and 15% glycerol, then stored at −70°C.

Frozen duplicate culture stock was shipped to the Public Health Research Institute (PHRI) Tuberculosis Center at UMDNJ. Culture stocks were subcultured on Lowenstein-Jensen slants, and DNA was extracted from each isolate. IS6110-based RFLP analysis was performed as described elsewhere [8]. RFLP patterns were analyzed using Bio Image pattern matching software (Bio Image). *M. tuberculosis* isolates that had DNA fingerprints with an identical hybridization banding pattern were considered to be the same strain and assigned a strain code following a nomenclature system that has been described elsewhere [9]. *M. tuberculosis* strains were assigned to 1 of 9 (I–VIII and IIA) discrete synonymous single-nucleotide polymorphism (SNP)–based phylogenetic lineages (synonymous SNP clusters) that were inferred on the basis of RFLP patterns and of previous analysis, reported elsewhere [6], of *M. tuberculosis* clinical isolates at PHRI. In addition, strains that exhibited similar IS6110 hybridization profiles (≥65% similarity), suggesting common recent ancestry, were collectively grouped into genotype families (eg, W-Beijing, CC, and BM) [6]. Strain clusters were defined as 1 occurrence of a specific strain during the study period. Strain patterns that were represented only once in the PHRI database and did not qualify for a family assignment were considered unique and given a default assignment (001).

Data were analyzed using Stata software (version 10.0; StataCorp). Bivariate analyses used the Student *t*, χ², and Fisher exact tests, as appropriate. A Wilcoxon rank sum test was used for comparison of median age between different groups. Patients’ ages were categorized by decade (15–19, 20–29, 30–39, 40–49, 50–59, and ≥60 years of age) for analysis of the dis-

---

**Figure 1.** *Mycobacterium tuberculosis* strains from the study community in a phylogenetic framework. RFLP, restriction fragment–length polymorphism; SNP, single-nucleotide polymorphism.
Figure 2. Distribution of the clusters, by cluster size, date of diagnosis, human immunodeficiency virus (HIV) status, and Mycobacterium tuberculosis strain family.

Results. Over the 5-year study period, 467 patients in the study community were diagnosed with sputum smear–positive pulmonary tuberculosis. The study laboratory received sputum specimens from 282 patients (60% of smear-positive patients) over this period. Of the 282 specimens received from these patients, 53% (n = 149) were successfully cultured for RFLP analysis, 22% (n = 61) lost viability during shipping or storage, 19% (n = 55) failed to culture M. tuberculosis, and 6% (n = 55) were contaminated. Two patients had dual infections with 2 different strains, and therefore a total of 151 M. tuberculosis isolates were included in this analysis.

No statistically significant differences were found between patients with and those without RFLP data (including patients for whom we did not receive specimens) in terms of age (P = .11), sex (P = .63), tuberculosis category (ie, new or retreatment cases; P = .22), test results for multidrug-resistant (MDR) tuberculosis (resistant to at least isoniazid and rifampicin; P = .15) or HIV status (among those tested; P = .58). However, patients for whom we did not obtain RFLP data had a higher death rate than did those for whom we did obtain RFLP data (P < .001).

The 149 patients with RFLP results ranged in age from 14 to 67 years (median age, 36 years), and 62% (n = 92) were men. In total, 81% (n = 121) of the patients were tested for HIV, and 54% (n = 65) of those 121 patients tested were infected with HIV. Four of the 149 patients (3%) had confirmed MDR tuberculosis. There were no tuberculosis-related deaths in this cohort.

A total of 95 different M. tuberculosis strains were identified (including 4 unique isolates), which are presented in a phylogenetic framework in Figure 1. Eight of the 9 recognized synonymous SNP clusters [10] were present in the community. The synonymous SNP cluster VI comprised 49% (n = 74) of the patients. Genetic variability within this group was high, with 57 strains detected in 74 patients. The synonymous SNP cluster II was the second largest group, with 26% (n = 39) of strains.

Twenty-six different M. tuberculosis genotype families were identified; the 4 largest families were W-Beijing (accounting for 26% of all M. tuberculosis strains in the community), CC (25%), AH (11%), and BM (7%). Bivariate analysis revealed no statistically significant association between the 4 dominant families and sex (P = .56), age category (P = .31), tuberculosis category (P = .73), or outcomes of tuberculosis treatment (P = .53). There was no change in the distribution of the main strain families across the 5-year period of data collection (P = .54).

Multivariate analysis adjusting for age, sex, tuberculosis category, and outcome yielded no association between HIV status and CC strain (odds ratio [OR], 1.23 [95% confidence interval...
clusters were diagnosed on average 157 days apart (95% CI, 0.48–3.17), AH strain (OR, 0.68 [95% CI, 0.21–2.25]), or BM strain (OR, 0.26 [95% CI, 0.04–1.49]). However, patients with the W-Beijing strain were significantly more likely to be HIV positive (OR, 3.65 [95% CI, 1.37–9.71]).

Of the patients with MDR tuberculosis, 2 were infected with W-Beijing strains, 1 was infected with a CC strain, and 1 was infected with an H strain. There was no statistical association between MDR tuberculosis and strain (P = .67) or HIV status (P = .87).

In this study, 54% (n = 81) of isolates occurred in strain clusters ranging in size from 2 to 14 patients, with 27% of the clustered strains occurring in pairs. Figure 2 demonstrates the distribution of the clustered strains over time by HIV status. In multivariate analysis, there was no association between clustering and HIV status (OR, 1.20 [95% CI, 0.55–2.65]). Paired clusters were diagnosed on average 157 days apart (<6 months), compared with an average of 321 days (>10 months) between occurrences of cases in the larger clusters (P = .013). Smaller, temporally associated clusters were noted within the larger clusters of the W451 and AH strains. With the exception of 2 cases in the AH cluster, none of the clustered cases occurred on the same residential plot in the township studied.

Discussion. The key finding of this study is the association between W-Beijing, one of the largest identified M. tuberculosis strain families, and HIV infection. This association persisted after controlling for a number of clinical factors and could be due to either an increased pathogenicity or virulence of the strain or an increased susceptibility of HIV-infected patients to these strains. W-Beijing M. tuberculosis strains have shown marked virulence in animal models of infection [6] and it has been suggested that certain sublineages of W-Beijing may have increased transmissibility and/or pathogenicity [11]. However, to our knowledge this is the first population-based study to show an association between W-Beijing and HIV infection. Further study of the biology of the W-Beijing strains and of their interaction with the HIV-infected host may help to explain the increased susceptibility of HIV-infected patients to tuberculosis and may also indicate novel ways to either protect or more effectively treat HIV-infected patients coinfected with M. tuberculosis. Other studies have found W-Beijing strains to be associated with MDR tuberculosis [12]. Therefore, the association with HIV-infected patients may have serious implications for the spread of MDR tuberculosis. The increased mortality in those without RFLP analysis may reflect a lower sputum retrieval rate at the local hospital, where tuberculosis was initially diagnosed in sicker patients.

This study has demonstrated a broad diversity of different M. tuberculosis strains, consistent with findings of other studies in sub-Saharan Africa [13, 14]. The high degree of genotypic diversity within the CC strains (related to F11/LAM) [13] may indicate that they are endemic in this population. The W-Beijing family also shows a high degree of diversity (16 variants in 39 patients), although to a lesser degree than the CC family, suggesting that these strains may be emerging and diversifying in the community. However, chromosomal location of IS6110 insertions may affect the movement of the insertion sequence elements, and therefore genetic diversity as determined by IS6110 RFLP may be independent of strain endemicity.

Another finding was the high rate of strain clustering. Clustering is not necessarily synonymous with recent transmission; evidence of geographical linkage, temporal association, and social contacts may be needed to support the suggestion of transmission. IS6110 fingerprinting is one of the most discriminatory typing techniques for isolates with IS6110 bands (eg, CC and W-Beijing), although it is less discriminatory for strains with fewer bands (eg, AH) [6], and may underestimate subclusters within the AH family. In this study approximately half of the strains were clustered, and there were close temporal associations, especially among the paired clusters. A proportion of disease may therefore be due to new infections. Because no association was found between HIV infection and clustering, new infections may be occurring in both HIV-negative and HIV-positive patients. Given the incomplete sampling of the study population, we probably underestimated the number of circulating strains, number of clusters, and sizes of clusters [15].

The spatial analysis demonstrated that temporally linked clusters did not occur on the same residential plots, suggestive of transmission outside of households. Temporally related clustering was also identified within larger clusters, such as of the W451 and AH strains. The strong temporal relationship of tuberculosis in HIV-positive patients within the W451 strain cluster (4 patients in 2003 and 6 patients in 2005) (Figure 2) may reflect nosocomial transmission at the single community clinic. Traditional epidemiological studies, including social interaction studies, are required to further delineate transmission.

In conclusion, we have shown that a wide diversity of strains exists in this periurban community in South Africa. Strain clustering, suggestive of ongoing transmission, was common in both HIV-positive and HIV-negative adults. The W-Beijing family was associated with HIV infection, a new finding that requires confirmation and explanation.

References

4. Wood R, Middelkoop K, Myer L, et al. Undiagnosed tuberculosis in...


