Progression to Active Tuberculosis, but Not Transmission, Varies by *Mycobacterium tuberculosis* Lineage in The Gambia

Bouke C. de Jong,1,2,a Philip C. Hill,1 Alex Aiken,1 Timothy Awine,1 Martin Antonio,1 Ifedayo M. Adetifa,1 Dolly J. Jackson-Sillah,1 Annette Fox,1 Kathryn DeRiemer,3 Sebastien Gagneux,4 Martien W. Borgdorff,5,6 Keith P. W. J. McAdam,1 Tumani Corrah,1 Peter M. Small,4 and Richard A. Adegbola1

1Bacterial Diseases Programme, Medical Research Council Laboratories, Banjul, The Gambia; 2Stanford University, Stanford, and 3University of California, Davis; 4Institute for Systems Biology, Seattle, Washington; 5University of Amsterdam, Amsterdam, and 6Royal Netherlands Tuberculosis Association, The Hague, The Netherlands

**Background.** There is considerable variability in the outcome of *Mycobacterium tuberculosis* infection. We hypothesized that *Mycobacterium africanum* was less likely than *M. tuberculosis* to transmit and progress to tuberculosis disease.

**Methods.** In a cohort study of patients with tuberculosis and their household contacts in The Gambia, we categorized 1808 HIV-negative tuberculosis contacts according to exposure to *M. tuberculosis* or *M. africanum*. Positive skin test results indicated transmission, and development of tuberculosis during 2 years of follow-up indicated progression to disease.

**Results.** Transmission rates were similar, but rates of progression to disease were significantly lower in contacts exposed to *M. africanum* than in those exposed to *M. tuberculosis* (1.0% vs. 2.9%; hazard ratio [HR], 3.1 [95% confidence interval {CI}, 1.1– 8.7]). Within *M. tuberculosis* sensu stricto, contacts exposed to a Beijing family strain were most likely to progress to disease (5.6%; HR relative to *M. africanum*, 6.7 [95% CI, 2.0 –22]).

**Conclusions.** *M. africanum* and *M. tuberculosis* transmit equally well to household contacts, but contacts exposed to *M. africanum* are less likely to progress to tuberculosis disease than those exposed to *M. tuberculosis*. The variable rate of progression by lineage suggests that tuberculosis variability matters in clinical settings and should be accounted for in studies evaluating tuberculosis vaccines and treatment regimens for latent tuberculosis infection.

Tuberculosis remains a significant public health problem, particularly in settings with limited resources [1]. The considerable variability in the outcome of *Mycobacterium tuberculosis* infection has usually been attributed to host and environmental factors [2, 3]; indeed, host immune suppression, such as that caused by HIV infection, is the strongest known risk factor for the development of active tuberculosis [4]. However, pathogen-related factors may also play a role [5]. Although distinct genotypes have been identified in the *Mycobacterium tuberculosis* complex, it is unclear whether these translate into phenotypic differences in humans [6].

Significant strain differences between *Mycobacterium bovis* and *M. tuberculosis* were identified early in the history of mycobacteriology [7], and several experimental studies have shown that *M. tuberculosis* strains can differ in immunogenicity and virulence in animal models [8]. Recent advances in genotyping now allow more detailed analyses of the contribution of bacterial factors to the variability in transmission and rates of progression to tuberculosis disease in its natural human host [9]. In population-based studies, these techniques have identified clustered isolates of *M. tuberculosis* (i.e., with identical genotypes) [10, 11]. Although clustering is suggestive of recent transmission, these studies were unable to distinguish between rates of transmission and rates of
progression to tuberculosis disease. A different study described an outbreak strain associated with high rates of tuberculin skin test (TST) conversion in contacts compared with historical control subjects [12]. In the context of a household study of tuberculosis in The Gambia, where *Mycobacterium africanum* is endemic, we assessed the likelihood of transmission and of progression to tuberculosis disease according to mycobacterial lineage. We specifically tested 2 hypotheses: that *M. africanum* would be less likely than *M. tuberculosis* to be transmitted and that it would be less likely to cause disease.

**METHODS**

In the Tuberculosis Case-Contact Study, we followed up 317 adult tuberculosis index case patients (sputum smear positive) and 2381 of their household contacts (figure 1). Participants were recruited between September 2002 and September 2004 and were followed up for 2 years. Household members were eligible for inclusion in the study if they had been sleeping in the same compound (walled group of houses) as the index case patient during the period of illness with tuberculosis. Household contacts underwent a TST (PPD R23 2TU; Staten Serum Institute), performed using the Mantoux technique. Those with negative TST results (induration of <10 mm) underwent another test after 3 months. There was no practice of treating asymptomatic TST-positive persons for latent infection in The Gambia.

**Follow-up.** Five follow-up visits were made to each of the 317 households, at 3, 6, 12, 18, and 24 months after enrollment. Any participant who reported tuberculosis symptoms at these visits was encouraged to present to the Medical Research Council (MRC) tuberculosis clinic and had free access to treatment for any illness during this period. At each household visit, we reevaluated each individual for symptoms of tuberculosis. Any patients with symptoms of pulmonary disease underwent chest radiography and sputum analysis (3 samples) for acid-fast bacilli (AFB; smear analysis and culture). If tuberculosis disease was bacteriologically confirmed or clinically suspected in smear-negative or extrapulmonary cases, patients were referred to the Gambian National Tuberculosis Treatment Program for the standard 6-month course of tuberculosis treatment. A diagnosis of tuberculosis disease among household contacts during the 2-year follow-up period was used as the main outcome variable for the analysis of progression to tuberculosis by mycobacterial lineage.

**Case definition.** All contacts with symptoms consistent with tuberculosis (fever, night sweats, and persistent cough) or with a positive TST result at enrollment or at the 3-month follow-up visit were offered a chest radiograph and 3 sputum tests if they had a productive cough. On the basis of the results of the chest radiograph, sputum smear, and culture and/or their response to a trial tuberculosis treatment course with the standard 4-drug regimen, contacts were classified as nondiseased or diseased (secondary case patients; see definitions in the section on statistical analysis). A positive TST result was not an essential element of the case definition.

In addition to the identification of secondary case patients presenting at the MRC clinic and during follow-up visits, the names and ages of all tuberculosis case patients treated at the government health clinics during the course of the study were recorded. Those who matched contacts participating in our study (on the basis of age-category matching within 5 years of the stated age on the government record) were visited again to confirm whether they had received tuberculosis treatment. Those who confirmed treatment were asked to undergo chest radiography to look for evidence of tuberculosis. The radiographs were reviewed by 2 physicians experienced in infectious diseases and a pediatrician if the participant was a child. After review, a consensus opinion was formed on the presence or absence of tuberculosis. Contacts with symptoms compatible with tuberculosis but with insufficient information to support a diagnosis of tuberculosis for this study were classified as unconfirmed secondary case patients.

**Laboratory procedures.** Sputum from index case patients and symptomatic contacts was examined for AFB by the auramine and Ziehl-Nielsen methods. Decontaminated sputum was cultured both in liquid medium (BACTEC 9000; Becton Dickinson) and on Lowenstein-Jensen slopes, prepared as described elsewhere [13]. Symptomatic contacts who presented directly to the Gambian government tuberculosis clinic had sputum examined for AFB but not stored for culture and genotyping.
HIV testing was restricted to those who either wanted to know their result or were selected for full immunological testing by skin test and enzyme-linked immunospot (ELISPOT) assay. Those individuals who tested positive for HIV-1 and/or HIV-2 were referred to the on-site MRC HIV clinic, where they were eligible for antiretroviral therapy.

**Genotyping methods.** Clinical isolates were characterized molecularly by means of spoligotype analysis [14], resulting in a binary pattern based on the presence or absence of 43 spacers. In addition, we performed polymerase chain reactions for large-sequence polymorphisms [15] based on phylogenetically informative regions of difference (RDs). This resulted in classification in 1 of 12 lineages, including the subspecies *M. africanum* (RD702) and 11 lineages within *M. tuberculosis* sensu stricto. The main lineages within *M. tuberculosis* sensu stricto included the Beijing family (RD105), lineage RD174 with a deletion in the DosR regulon, and lineages RD182 and RD219.

**Genotypic definition.** Recently, molecular techniques have identified 2 lineages within the true *M. africanum* type I: West African type 1, which is phylogenetically closer to *M. tuberculosis* and is found predominantly around the gulf of Guinea, and West African type 2, which is phylogenetically closer to *M. bovis* and is confined to the western parts of West Africa [9, 16]. In this article, *M. africanum* refers to *M. africanum* type I, West African type 2.

Identical spoligotype patterns (genotypes) in isolates from an index case patient and a diseased contact from the same household were referred to as concordant genotypes, and different patterns were referred to as “discordant genotypes.” Isolates sharing the same spoligotype pattern except for 1 of the 43 spacers were classified as concordant, because these patterns may reflect evidence of direct transmission between household members.

**Ethical approval.** The study was approved by the Gambian government/MRC ethics committee and the Stanford University Institutional Review Board. All participants provided informed consent before enrollment.

**Definitions.** Contacts were classified as prevalent TST positive if they had positive TST results (≥10 mm) at recruitment and as incident TST positive if they had negative TST results (<10 mm) at recruitment and positive results at the 3-month follow-up visit, with induration increased by ≥6 mm since recruitment. Contacts in whom tuberculosis disease developed (secondary case patients) were classified as having coprevalent tuberculosis if their tuberculosis was diagnosed at recruitment or within 3 months after recruitment and as having incident tuberculosis if their tuberculosis was diagnosed ≥3 months after recruitment.

**Statistical analysis.** Analysis was done stratifying by subspecies (*M. africanum* vs. *M. tuberculosis*) and by lineage within the subspecies, where lineages with information on <70 contacts were grouped as other. In the absence of complete mycobacterial genotype data for the diseased contacts, we did not correct the analysis of the secondary case patients by lineage for the likelihood of infection with a discordant genotype. Prevalent positive TST results in household contacts and incident TST results at 3 months were used as outcome variables for transmissibility of the different mycobacterial lineages. After confirmation that HIV was an independent predictor of lower TST indurations and of increased progression to disease, analyses of both transmission and progression were limited to known HIV-uninfected contacts. We used the χ2 test to test for differences in TST positivity and progression to disease between contacts exposed to the different mycobacterial lineages. We calculated odds ratios and their corresponding 95% confidence intervals (CIs) for positive TST results, using logistic regression and controlling for household clustering. Possible confounders, such as age, presence of a scar from bacille Calmette-Guérin (BCG) vaccination, and sleeping proximity, were added to the model. We used Cox regression models to estimate the hazard ratios (HRs) and 95% CIs for secondary case patients among the contacts of index case patients with *M. tuberculosis* versus *M. africanum* infection (controlling for household clustering) after confirming compliance with the proportional hazards assumption.

We performed several different sensitivity analyses to assess potential biases. A degree of misclassification bias occurred, because not all secondary case patients were infected via their index case patient. To determine whether misclassification bias was important, we removed from the analysis the secondary case patients known to be infected with a discordant genotype. The potential effect of unconfirmed tuberculosis disease was assessed by repeating the analysis with unconfirmed secondary case patients classified as either nondiseased contacts or secondary case patients. The potential effect of including coprevalent secondary case patients in the analysis of progression was assessed by recalculating the HR for incident secondary case patients only.

All analyses were conducted using Stata software (version 9; StataCorp).

**RESULTS**

**Study participants.** In a tuberculosis case-contact study, cultured isolates were successfully obtained from 301 index case patients (95%), of whom 291 (97%) had interpretable genotyping patterns. These 291 index case patients had 2174 household contacts, of whom 1853 (85%) underwent HIV testing, with 45 (2.4%) positive results; the 1808 HIV-uninfected contacts were eligible and were included in this study. At recruitment, the TST was administered to 1727 contacts (96%), of whom 708 (41%) tested positive. Of the 1808 HIV-uninfected contacts, 48% were male, and 49% had a scar from BCG vaccination; their median age was 16 years (interquartile range, 8–25 years).

**Mycobacterial genotypes.** The 291 isolates included 12 different lineages within the *M. tuberculosis* complex: 110 isolates of
M. africanum (1 lineage) and 181 isolates of M. tuberculosis sensu stricto (11 lineages). Detailed genotypic findings are available from the authors upon request.

**Follow-up.** The HIV-uninfected contacts had 3301 person-years of follow-up in total. This yielded 19 contacts with coprevalent tuberculosis (presenting within 3 months of enrollment) and 20 contacts with incident tuberculosis (in whom symptoms consistent with tuberculosis developed >3 months after enrollment), for a total of 39 secondary case patients. The incident tuberculosis case patients are described in detail elsewhere [17]. Their median age was 17 years (range, 2–70 years), and 22 (56%) were male. Six contacts had extrapulmonary tuberculosis, and 33 had pulmonary tuberculosis. Four secondary case patients were identified through the records of the government tuberculosis clinic, with x-ray findings suggestive of (treated) tuberculosis. Eight contacts had symptoms but insufficient information to support a diagnosis of tuberculosis, including 3 who had received tuberculosis treatment but had normal chest x-ray findings afterward; they were excluded from the analysis of the progression to disease.

Paired genotyped isolates from index and secondary case patients were available for 18 of the 39 secondary case patients. Eleven of these 18 paired isolates had identical genotypes; 1 isolate from a secondary case patient lacked 1 additional spacer (of 43) compared with the isolate from the index case patient, and 6 had different genotypes.

**Transmission by mycobacterial lineage.** The prevalence of household contacts with a positive TST result at enrollment was not significantly different between contacts exposed to M. africanum and those exposed to M. tuberculosis (39% vs. 42%; P = .14) or between the lineages within M. tuberculosis (table 1). Similarly, the incidence of TST conversion did not differ between lineages or between M. africanum and M. tuberculosis (25% vs. 26%; P = .70). Adding known predictors of TST positivity to the logistic regression model, such as age and sleeping proximity to the index case patient, did not significantly change these results.

**Progression to tuberculosis disease by mycobacterial lineage.** Comparing M. tuberculosis sensu stricto (all 11 lineages combined) with M. africanum, the rate of progression to tuberculosis was 2.9% in household contacts of index case patients with M. tuberculosis sensu stricto infection, compared with 1.0% in those with M. africanum infection (HR, 3.1 [95% CI, 1.1–8.7]). Figure 2 depicts the results using survival curves, showing significant divergence over time (P = .036).

Table 2 shows the likelihood of progression to disease by lineage (P = .033 across lineages). Household contacts for whom the index case patient had tuberculosis caused by a Beijing family isolate of M. tuberculosis were more likely to develop tuberculosis disease (5.6%) than contacts of index patients with M. africanum infection (1.0%) (HR, 6.7 [95% CI, 2.0–22]) (table 2), but they were not more likely to develop disease than other lineages within M. tuberculosis sensu stricto. Contacts of an index patient with the lineage within M. tuberculosis sensu stricto defined by RD174, a deletion in part of the DosR regulon, were the second most likely to progress to disease (3.9% in 2 years; HR, 4.4 [95% CI, 1.4–14]).

Progression to tuberculosis in contacts with prevalent or incident positive TST results occurred in 4.9% of contacts of index case patients with M. tuberculosis sensu stricto infection versus 2.2% of those with M. africanum infection (P = .046). When the analyses were repeated with exclusion of the secondary case patients infected with isolates that differed by at least 1 spacer from the isolate for the index patient (n = 7), the association between genotype and the number of secondary case patients was stronger (HR for M. tuberculosis–exposed relative to M. africanum–exposed households, 4.0 [95% CI 1.2–13]; P = .026). Similarly, when the analyses were repeated excluding the coprevalent case patients (n = 19), incident tuberculosis was lower in contacts exposed to M. africanum than in those exposed

Table 1. Tuberculin skin test (TST) results in HIV-uninfected contacts, by lineage within the *Mycobacterium tuberculosis* complex.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Total contacts, no.</th>
<th>TST-positive contacts, no. (%)</th>
<th>OR (95% CI)</th>
<th>Total contacts, no.</th>
<th>TST-positive contacts, no. (%)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD702 (M. africanum)</td>
<td>653</td>
<td>253 (39)</td>
<td>1</td>
<td>280</td>
<td>69 (25)</td>
<td>1</td>
</tr>
<tr>
<td>RD105 (Beijing)</td>
<td>72</td>
<td>27 (38)</td>
<td>0.90 (0.33–2.4)</td>
<td>24</td>
<td>6 (25)</td>
<td>1.3 (0.31–5.7)</td>
</tr>
<tr>
<td>RD174 (DosR)</td>
<td>200</td>
<td>84 (42)</td>
<td>1.0 (0.55–1.9)</td>
<td>91</td>
<td>26 (29)</td>
<td>1.2 (0.49–2.7)</td>
</tr>
<tr>
<td>RD182</td>
<td>311</td>
<td>136 (44)</td>
<td>1.3 (0.77–2.3)</td>
<td>127</td>
<td>32 (25)</td>
<td>1.1 (0.48–2.4)</td>
</tr>
<tr>
<td>RD219</td>
<td>84</td>
<td>39 (46)</td>
<td>1.4 (0.58–3.3)</td>
<td>33</td>
<td>5 (15)</td>
<td>0.39 (0.09–1.7)</td>
</tr>
<tr>
<td>Other</td>
<td>407</td>
<td>169 (42)</td>
<td>1.2 (0.76–2.0)</td>
<td>188</td>
<td>51 (27)</td>
<td>1.2 (0.59–2.3)</td>
</tr>
<tr>
<td>Total</td>
<td>1727</td>
<td>708 (41)</td>
<td>...</td>
<td>743</td>
<td>189 (25)</td>
<td>...</td>
</tr>
</tbody>
</table>

**NOTE.** CI, confidence interval; OR, odds ratio; RD, region of difference (major phylogenetic groups are identified by major large-sequence polymorphisms or deletions).

a TST positivity at enrollment.

b TST positivity at 3 months in those who tested TST negative at baseline, with conversion defined as a ≥6-mm increase in induration.

Comparing M. tuberculosis sensu stricto (all 11 lineages combined) with M. africanum, the rate of progression to tuberculosis was 2.9% in household contacts of index case patients with M. tuberculosis sensu stricto infection, compared with 1.0% in those with M. africanum infection (HR, 3.1 [95% CI, 1.1–8.7]). Figure 2 depicts the results using survival curves, showing significant divergence over time (P = .036).
to *M. tuberculosis* (HR, 5.5 [95% CI, 1.3–23]; *P* = .020). Finally, when we included contacts who had an unconfirmed tuberculous diagnosis as case patients (*n* = 8), the incidence of tuberculosis was lower in contacts exposed to *M. africanum* than in those exposed to *M. tuberculosis* (HR, 2.7 [95% CI, 1.2–6.5]; *P* = .022).

**DISCUSSION**

Although transmission of the *M. tuberculosis* complex from patients with tuberculosis to their household contacts did not vary by strain lineage in The Gambia, the percentage of contacts with incident tuberculosis during the 2-year follow-up varied 5-fold between contacts exposed to the different *M. tuberculosis* lineages. These percentages ranged from 1.0% of contacts in households exposed to *M. africanum* to 5.6% of those exposed to the Beijing family within *M. tuberculosis* sensu stricto (*P* = .033). The different rate of progression to active tuberculosis between lineages provides strong evidence that pathogen-specific factors are important in the observed variability in outcomes of tuberculosis infection and indicates that the bacterial determinants of transmission and initial infection are distinct from the determinants of progression to disease.

The finding that fewer HIV-negative contacts of *M. africanum* index case patients have incident tuberculosis is consistent with other characteristics of this lineage. We previously reported that *M. africanum*-infected case patients and their contacts were less likely to mount an interferon-γ response against early secreted antigenic target (ESAT)-6 than *M. tuberculosis*-infected case patients and contacts [18]. The gene *esat-6*, located in RD1 [16], is a known virulence factor [19] and is thought to play a role in the cell-to-cell spread of *M. tuberculosis* [20]. The sequence of RD1

![Figure 2. Survival curves showing survival free of tuberculosis after exposure to *Mycobacterium africanum* vs. *Mycobacterium tuberculosis*, on the basis of a Cox regression model.](https://academic.oup.com/jid/article-abstract/198/7/1037/2191885)

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Contacts, no.</th>
<th>Secondary case patients, no. (%)</th>
<th>HR (95% CI) a</th>
<th>HR (95% CI) b</th>
<th>HR (95% CI) c</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD702 (<em>Mycobacterium africanum</em>)</td>
<td>681</td>
<td>7 (1.0)</td>
<td>1 (reference)</td>
<td>1 (reference)</td>
<td>1 (reference)</td>
</tr>
<tr>
<td>RD105 (Beijing)</td>
<td>72</td>
<td>4 (5.6)</td>
<td>6.7 (2.0–22)</td>
<td>7.7 (1.9–31)</td>
<td>16 (2.8–89)</td>
</tr>
<tr>
<td>RD174 (DosR)</td>
<td>204</td>
<td>8 (3.9)</td>
<td>4.4 (1.4–14)</td>
<td>5.8 (1.5–23)</td>
<td>10 (2.2–45)</td>
</tr>
<tr>
<td>RD182</td>
<td>317</td>
<td>9 (2.8)</td>
<td>2.8 (0.84–9.5)</td>
<td>3.6 (0.92–15)</td>
<td>4.2 (0.77–23)</td>
</tr>
<tr>
<td>RD219</td>
<td>86</td>
<td>1 (1.2)</td>
<td>1.4 (0.16–12)</td>
<td>NO</td>
<td>4.2 (0.42–43)</td>
</tr>
<tr>
<td>Other</td>
<td>440</td>
<td>10 (2.3)</td>
<td>2.4 (0.74–7.6)</td>
<td>3.5 (0.95–13)</td>
<td>3.2 (0.60–17)</td>
</tr>
<tr>
<td>Total</td>
<td>1800</td>
<td>39 (2.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.** CI, confidence interval; HR, hazard ratio; NO, no observations; RD, region of difference (major phylogenetic groups are identified by major large-sequence polymorphisms or deletions).

a With a random effects model, to account for household clustering.

b Excluding secondary case patients whose isolate differed by at least 1 spacer from the index case patient’s isolate.

c Excluding coprevalent case patients (i.e., contacts with tuberculosis diagnosed within 3 months of enrollment).
of *M. africanum* showed a frameshift mutation in gene Rv3879c compared with the same region in *M. tuberculosis* and *M. bovis*. As a result, Rv3879c is a pseudogene in *M. africanum* [18]. A transposon mutant of the Rv3879c homologue in *Mycobacterium marinum*, a mycobacterium species that also contains RD1, showed undetectable levels of ESAT-6 in the culture filtrate but normal levels of ESAT-6 in cell lysate [21]. When Rv3879c is nonfunctional, ESAT-6 secretion may be impaired, and this would explain the attenuated ESAT-6 response induced by *M. africanum*. Therefore, although reduced secretion of ESAT-6 may impair the cell-to-cell spread of the bacterium and possibly explain its lower progression to disease, that reduced secretion could also facilitate immune evasion, balancing selective pressures on the pathogen’s evolution. These findings suggest a role for RD1 in within-host but not between-host interactions.

Although the above provides a potential molecular explanation (i.e., a proximal cause) for why *M. africanum* infection exhibits slower disease progression than *M. tuberculosis* infection, it does not address the ultimate evolutionary reason for this difference. The development of active disease is a *sine qua non* for its lower progression to disease, that reduced secretion could also facilitate immune evasion, balancing selective pressures on the pathogen’s evolution. These findings suggest a role for RD1 in within-host but not between-host interactions.

In summary, our findings confirm the importance of mycobacterial factors in the variability in progression to tuberculosis disease in humans. The contribution of pathogen factors to the outcome of mycobacterial infections has implications for understanding tuberculosis epidemics and the development and assessment of new interventions, specifically vaccines and treatments for latent tuberculosis infection. Further studies in other settings can identify specific lineages with high and low predilection for progression to disease, and detailed molecular analyses...
may provide fresh insights for the development of new interventions.

**Acknowledgments**

We thank the tuberculosis workers in The Gambia for their excellent work, the study participants for their time and cooperation, and Hilton Whittle, Paulus de Jong, Julie Parsonnet, and Joel Ernst for critical reading of the manuscript.

**References**