



Original article

Factors affecting tuberculosis strain success over 10 years in a high TB- and HIV-burdened community

Keren Middelkoop,^{1,2*} Linda-Gail Bekker,^{1,2} Barun Mathema,^{3,4} Landon Myer,⁵ Elena Shashkina,⁴ Andrew Whitelaw,^{6,7} Natalia Kurepina,⁴ Gilla Kaplan,⁸ Barry Kreiswirth⁴ and Robin Wood^{1,2}

¹Desmond Tutu HIV Centre, Institute of Infectious Disease & Molecular Medicine, University of Cape Town, Cape Town, South Africa, ²Department of Medicine, University of Cape Town, Cape Town, South Africa, ³Department of Epidemiology, Mailman School of Public Health, Columbia University, NY, New York, USA, ⁴Public Health Research Institute, Tuberculosis Center, New Jersey Medical School, Rutgers, State University of New Jersey, Newark, USA, ⁵Division of Epidemiology & Biostatistics, School of Public Health & Family Medicine, University of Cape Town, Cape Town, South Africa, ⁶National Health Laboratory Service, Cape Town, South Africa, ⁷Division of Medical Microbiology, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa and ⁸Laboratory of Mycobacterial Immunity and Pathogenesis, Public Health Research Institute, New Jersey Medical School, Rutgers, The State University of New Jersey, Newark, USA

*Corresponding author. Desmond Tutu HIV Centre, Institute of Infectious Disease & Molecular Medicine, University of Cape Town, PO Box 13801, Mowbray Observatory, Cape Town 7705, South Africa. E-mail: keren.middelkoop@hiv-research.org.za

Accepted 5 February 2014

Abstract

Background: Factors associated with *Mycobacterium tuberculosis* (*Mtb*) strain success over time in high burdened communities are unknown.

Methods: *Mtb* isolates collected over 10 years from sputum-positive tuberculosis (TB) patients resident in the study site underwent IS6110-based restriction fragment length polymorphism analysis. Clinical, demographic and social data were extracted from clinic records and interviewer-administered questionnaires. Strains were defined as persistently successful, transiently successful or unsuccessful based on the average number of cases per year and their continued presence over time.

Results: Genotyping data were available on 789 TB cases. Of the 311 distinct *Mtb* strains (≥ 6 bands) identified, 247 were categorized as unsuccessful strains, 12 transiently successful and 10 persistently successful strains. Strain success was not associated with age, gender, antiretroviral use or social factors. Persistently successful strains were less likely to be drug-resistant compared with transiently successful strains [odds ratio (OR): 0.13; 95% confidence interval (CI): 0.04 – 0.5]. Persistently successful strains were positively associated with host HIV-infection compared with unsuccessful strains, but this finding was not robust in sensitivity analyses.

Conclusions: Pathogen characteristics appear to play a greater role in *Mtb* strain success compared with social or host factors. This study supports the need for further investigations into the role of pathogen characteristics in strain success.

Key words: Tuberculosis, pulmonary, epidemiology, molecular, polymorphism, restriction fragment length

Key Messages

- Wide diversity of *Mtb* strain was observed in this high burdened HIV and TB community during a 10-year period.
- Strain success was not associated with social factors.
- Transiently successful *Mtb* strains were associated with drug resistance.
- Persistently successful *Mtb* strains were weakly associated with HIV infection.
- Pathogen characteristics appear to play a greater role in *Mtb* strain success compared with social or host factors.

Introduction

Tuberculosis (TB) remains a significant cause of morbidity and mortality in sub-Saharan Africa, and current control strategies are failing to contain TB epidemics in countries with high HIV prevalence¹. In order to inform effective adjunctive control strategies, good longitudinal data are required to provide insights into time trends of TB epidemics² and to understand the factors driving endemic TB. Molecular epidemiological tools have expanded our ability to explore TB epidemics. In endemic settings, community-based studies have reported high diversity of strains^{3,4} and identified clinical characteristics associated with *Mycobacterium tuberculosis* (*Mtb*) strain clustering^{5,6} and cluster size.⁷ However, there are few data describing the host, environmental and bacterial factors affecting strain competition and survival.⁸ Such data may inform the development of adjunctive control strategies. We therefore analysed *Mtb* strain diversity over a 10 year period in a South African community with high TB and HIV prevalence, and assessed the host, social and pathogen characteristics associated with success of *Mtb* strains.

Methods

Study community

The study community has been described previously.^{9,10} Briefly, it is a geographically well-defined South African peri-urban township with a population of approximately 18 000 people and an HIV prevalence of 23–25%.^{9,11} TB care, provided by a single primary care clinic, follows the South African National TB Control Program guidelines¹² which utilize passive case detection for HIV-positive and HIV-negative cases with World

Health Organization-recommended directly observed treatment short courses. Despite an apparently well-functioning TB programme¹³ and high coverage antiretroviral treatment (ART) provision, TB notification rates remain extraordinarily high at ~2000/100 000.¹⁰

Patient population

Demographic and clinical data, including HIV and ART status, were extracted from the clinic TB register and patient folders of all resident TB patients in the community, notified from 2001 to 2010. Social data including housing and household crowding were obtained by interviewer-administered questionnaire completed by TB patients. The study was approved by the Human Research Ethics Committee of the University of Cape Town and the Institutional Review Board of the University of Medicine and Dentistry of New Jersey (UMDNJ). Patients provided written informed consent.

Laboratory procedures

Sputum specimens were obtained from TB suspects from 2001 through 2010 in accordance with the national TB programme diagnostic protocol.¹² All sputum specimens were assessed for the presence of acid-fast bacilli (AFB) by fluorescent (auramine) microscopy after concentration by centrifugation. From 2001 to end 2005, AFB-positive sputum samples were cultured on Lowenstein-Jensen (LJ) medium, as were smear-negative specimens from patients in whom culture was clinically indicated (e.g. re-treatment patients), and isoniazid and rifampicin susceptibility testing was performed on isolates from re-treatment patients or patients remaining AFB-positive after 2 months of

treatment.¹² From 2006 to end 2010, culture and isoniazid and rifampicin susceptibility testing was performed on all specimens, and second-line susceptibility testing was performed on all multi-drug resistant TB (MDR-TB, resistant to at least isoniazid and rifampicin) isolates. Drug susceptibility testing was performed using the MGIT culture system (Becton Dickinson, Sparks, MD). All isolates were inoculated in duplicate into 7H9 liquid medium supplemented with oleic acid, albumin, dextrose and catalase (OADC) and 15% glycerol at the research laboratory at the Institute of Infectious Disease and Molecular Medicine at the University of Cape Town, and stored at -70°C .

Genotyping analysis

Frozen culture stocks were shipped to the Public Health Research Institute Tuberculosis Center at UMDNJ. Stocks were sub-cultured on LJ medium and DNA extracted from each isolate. IS6110-based restriction fragment length polymorphism (RFLP) analysis was performed on each isolate as described elsewhere.¹⁴ RFLP patterns were analysed using BioImage pattern-matching software (BioImage, MI). *Mtb* isolates with an identical DNA hybridization banding pattern were considered to be the same strain and assigned a strain code following the previously described nomenclature system.¹⁵ In addition, *Mtb* strains were assigned to one of nine (I–VIII and II.A) discrete synonymous single nucleotide polymorphism (sSNP)-based phylogenetic lineages, as determined by RFLP patterns and previous analysis of *Mtb* clinical isolates.¹⁶

Definitions

MDR-TB was defined as resistance to isoniazid and rifampicin, with or without resistance to other TB drugs. Any drug-resistant TB (DR-TB) included any combination of mono- or dual-drug resistance and/or MDR-TB.

A unique or singleton strain was an isolate with an RFLP pattern that occurred in only one patient within the study dataset. Strain clusters were defined as more than one occurrence of a specific strain in different individuals during the study period. Re-treatment TB resulting from relapse of the same strain in the same individual was excluded from cluster analyses. Strain family was a group of strains that exhibited similar, but not identical, IS6110 hybridization profiles, suggesting relatedness by descent (e.g. W-Beijing family).¹⁷

For clustered strains, cluster density was calculated as cluster size (i.e. number of isolates of a specific strain) divided by the duration of time that strain was present in the community. Persistently successful strains were defined as strains that had been present for at least 2 consecutive

years and were still present in 2010 (persistence), and had ≥ 2 cases/year (cluster density). Transiently successful strains were strains that did not necessarily persist to 2010 but occurred for at least 2 consecutive years with a cluster density of ≥ 3 cases/year in the consecutive period. Unsuccessful strains were clustered strains that did not fall into these two groups, and singleton strains. Clustered strains occurring for the first time in 2009/2010 as well as singleton strains from 2010 were excluded from analysis, as their longevity or possibility of clustering in the community remains undetermined. A correction factor to adjust for the increasing sputum sampling rates across the study period was used to correct the strain cluster density (number of isolates of a strain in year k /sampling proportion in year k). A corrected assessment of successful strains was calculated using this factor. Sensitivity analyses were performed, with persistent success defined at cluster densities of ≥ 1 cases/year and of ≥ 3 cases/year, and transient success defined as strains that occurred for at least 3 consecutive years with a cluster density of ≥ 2 cases/year in that period. Sensitivity analyses were also performed restricted to 2006–10: the period of more complete sampling.

Data analysis

Data were analysed using STATA 10.0 (StataCorp, College Station, TX). Bivariate analyses employed Student's t -, Wilcoxon rank sum and chi-square tests, as appropriate. Multivariate logistic regression models were developed to examine factors associated with persistently successful, transiently successful and unsuccessful strains. All statistical tests were two-sided at $\alpha = 0.05$.

Results

From 2001 to 2010, there were 2237 TB cases, of which 1836 were pulmonary TB cases. Of these, 1353 were smear and/or culture-positive. *Mtb* cultures were obtained from 1011 (75%) sputum-positive cases and 789 (78% of collected specimens) had RFLP patterns available. The sputum sampling rate increased from 27% in 2001 to 92% in 2010 (Table 1).

Reasons for missing RFLP data are shown in Figure 1. Patients with RFLP patterns did not differ from those without RFLP patterns, with regard to age ($P = 0.31$), gender ($P = 0.59$) or HIV status ($P = 0.52$). The proportion of patients on ART increased over successive calendar years, but there was no independent association between ART use and genotype data availability ($P = 0.78$). We were less likely to obtain RFLP data for TB patients who died ($P = 0.01$), and

Table 1. Number of cases, sampling and RFLP completed rates by year over the study period

Year	Notified TB cases	Pulmonary TB cases	Smear and/or cult.-positive TB cases	Sputum specimen collected	RFLP genotyping completed
2001	131	86	67	20 (30%)	18 (27%)
2002	142	93	81	44 (54%)	32 (40%)
2003	169	125	97	63 (65%)	41 (42%)
2004	195	139	114	63 (55%)	34 (30%)
2005	275	237	189	97 (51%)	34 (18%)
2006	251	203	162	128 (79%)	109 (67%)
2007	256	215	162	148 (91%)	116 (72%)
2008	258	210	163	151 (93%)	137 (84%)
2009	301	283	179	164 (92%)	146 (82%)
2010	259	245	139	133 (96%)	122 (92%)
TOTAL	2237	1836	1353	1011 (75%)	789 (58%)

more likely to obtain RFLP data for patients with MDR-TB ($P=0.02$). There were no differences in the proportion of new and re-treatment cases ($P=0.53$) or in other TB outcomes between patients with and without RFLP data.

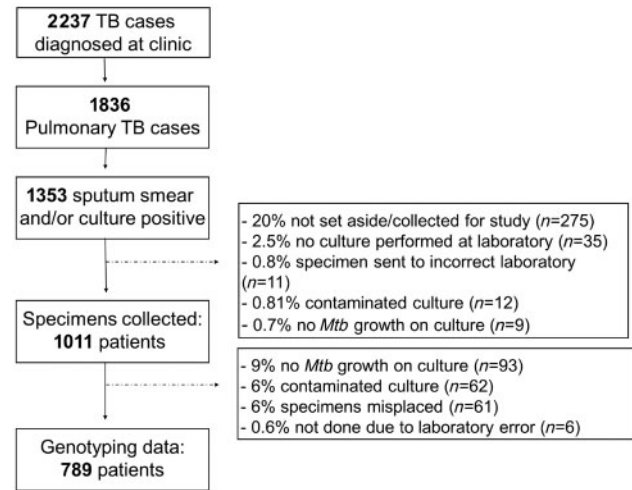
The 789 patients with RFLP results ranged in age from 1 to 77 years [median age: 33 years, interquartile range (IQR): 27–40], 18 (2%) of the patients were children and 462 (59%) were male. A total of 25 (3%) patients had MDR-TB and no cases of extensively drug-resistant TB were identified. Treatment outcomes were as follows: 611 (78%) completed TB treatment, 66 (8%) were transferred out of the community before treatment completion, 66 (8%) interrupted treatment, 31 (4%) died, 11 (1%) failed treatment, 2 were still receiving MDR-TB treatment and the MDR treatment outcome status for a further 2 patients was unknown at the end of the study.

Of 721 (91%) patients tested for HIV, 64% (462/721) were HIV-positive and 105 (23%) were receiving ART at TB diagnosis.

Whereas RFLP data were available for 789 TB cases, in total 798 RFLP patterns were available, as 10 patients had a dual infection; for one patient the second strain was not fully genotyped (variant of W-Beijing). Of the 798 isolates, 235 (30%) were singleton strains, 545 (68%) were clustered and 18 (2%) were reactivation of a previous strain in re-treatment patients.

M. tuberculosis strain diversity

Figure 2 depicts the 798 isolates present from 2001 to 2010 in the nine sSNP-discrete phylogenetic lineages. Strains from all nine sSNP clusters were identified in the community, indicating high diversity. sSNP clusters II and VI were the largest clusters with 230 and 360 isolates, respectively. *Mtb*

**Figure 1.** Consort diagram of TB cohort in the study community from 2001 to 2010

lineages 1 to 4, based on an alternative classification system, were all represented.¹⁹ Based on IS6110 RFLP analysis, the 798 isolates were divided into 59 strain families and 330 strains. The dominant strain families are W-Beijing (29%), CC-related (27%), AH (8%) and BM (5%). RFLP analysis poorly differentiates strains with more than six bands, such as the AH *Mtb* strain. Therefore 19 strains with more than six bands were excluded from further analysis (80 isolates).

Strain success

Among the 311 strains analysed, there were 87 strain clusters in this community, ranging in size from 2 to 85 isolates in a cluster (median cluster size: 3; IQR: 2–5). Of the 87 clustered strains, 13 appeared in the community for the first time in 2009/2010. Of the 74 eligible clusters (Figure 3a), 26 (35%) clustered strains were still present in 2010. The median cluster density of the 74 clusters was 1.2 cases/year (IQR: 0.7–2.2) (Figure 3c), and the median of the corrected cluster density was 1.4 cases/year (IQR: 0.9–2.6).

Of these 74 clustered strains, there were 10 persistently successful strains when the corrected cluster density was taken into consideration (Figure 3c). There were 12 transiently successful strains; unsuccessful strains comprised 193 singletons and 52 clustered strains (Figure 4). Four of the persistently successful strains belonged to the W-Beijing family of strains, and four belonged to the CC-related family.

Persistently successful vs unsuccessful strains

In bivariate analysis, persistently successful strains were positively associated with HIV-infection compared with unsuccessful strains [OR: 1.5; 95% confidence interval (CI): 1.0–2.2]. Success was not associated with sputum smear-positivity ($P=0.83$) nor with grades of smear

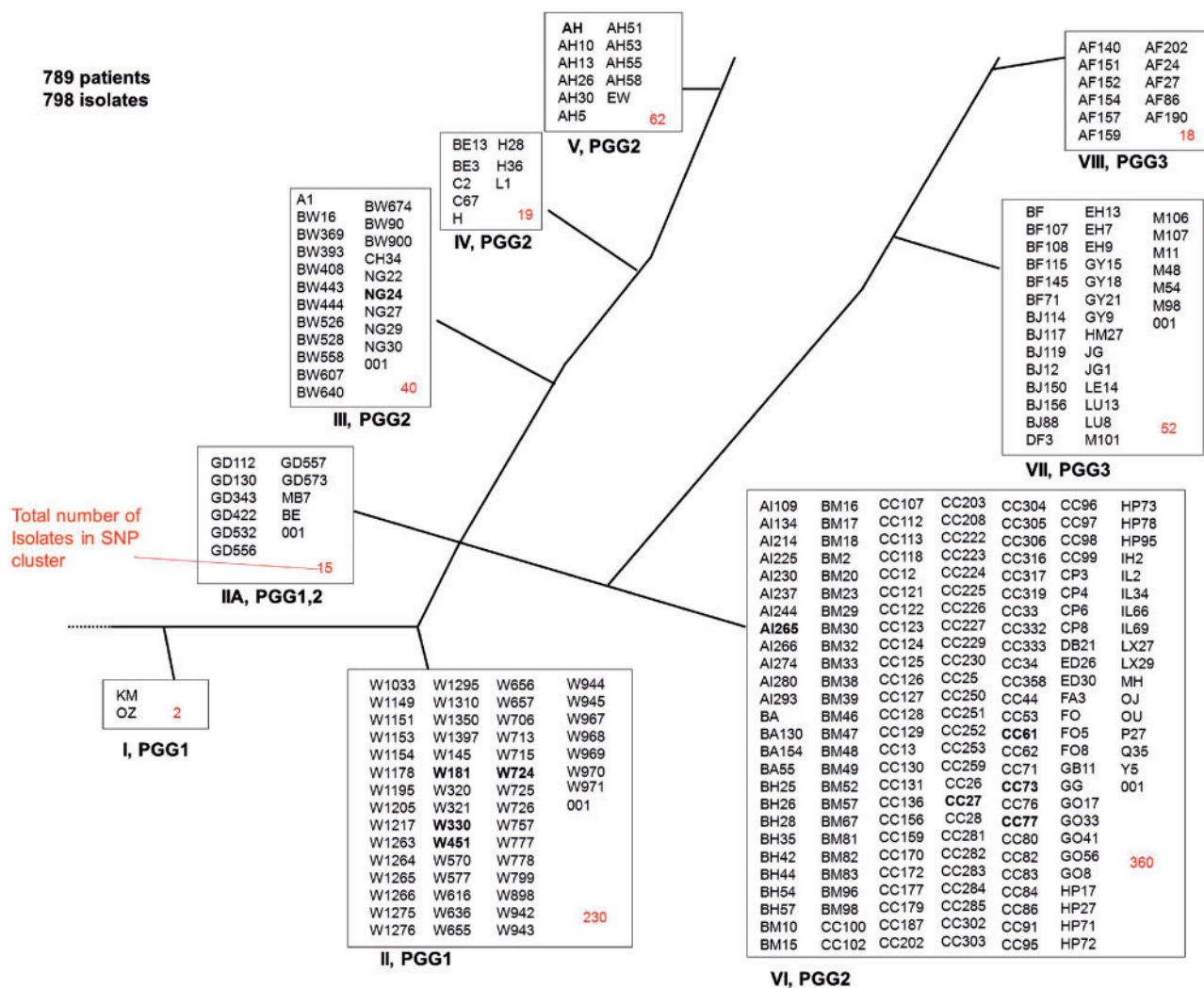


Figure 2. Phylogenetic tree of strains identified in the community from 2001 to 2010. Successful strains are emboldened

positivity ($P=0.32$). Persistently successful strains were positively associated with treatment completion ($P=0.05$) and negatively associated with both mortality ($P=0.03$) and any drug resistance compared with unsuccessful strains ($P=0.02$) (Table 2).

In multivariate logistic regression models adjusted for age and gender, patients with persistently successful strains had a higher odds of HIV-infection (odds ratio [OR]: 1.6; 95% CI: 1.1–2.3), but no association with treatment completion ($P=0.29$), mortality ($P=0.10$) or drug resistant TB ($P=0.10$). The association between HIV status and success persisted in multivariate models adjusting for study year.

Persistently successful vs transiently successful strains

In bivariate analysis, persistently successful strains were positively associated with treatment completion ($P=0.003$) and negatively associated with mortality ($P=0.01$), MDR-TB ($P=0.001$) and any drug resistance ($P<0.001$) compared with transiently successful strains (Table 2).

In multivariate logistic regression models adjusted for age and gender, persistently successful strains remained less likely to be drug resistant (OR: 0.13; 95% CI: 0.04–0.5) or multi-drug resistant (OR: 0.1; 95% CI: 0.01–0.7) compared with transiently successful strains, but there was no association with treatment completion ($P=0.10$), mortality ($P=0.95$) or HIV status ($P=0.09$).

Persistently and transiently successful strains were grouped together to form an overall group of ‘strain success’ and this was compared with unsuccessful strains. No associations were noted.

Strain success was not associated with any of the social factors assessed (Table 2).

Sensitivity analyses

The association between persistently successful strains and HIV infection remained when persistently successful strains were re-defined as those with persistence and a

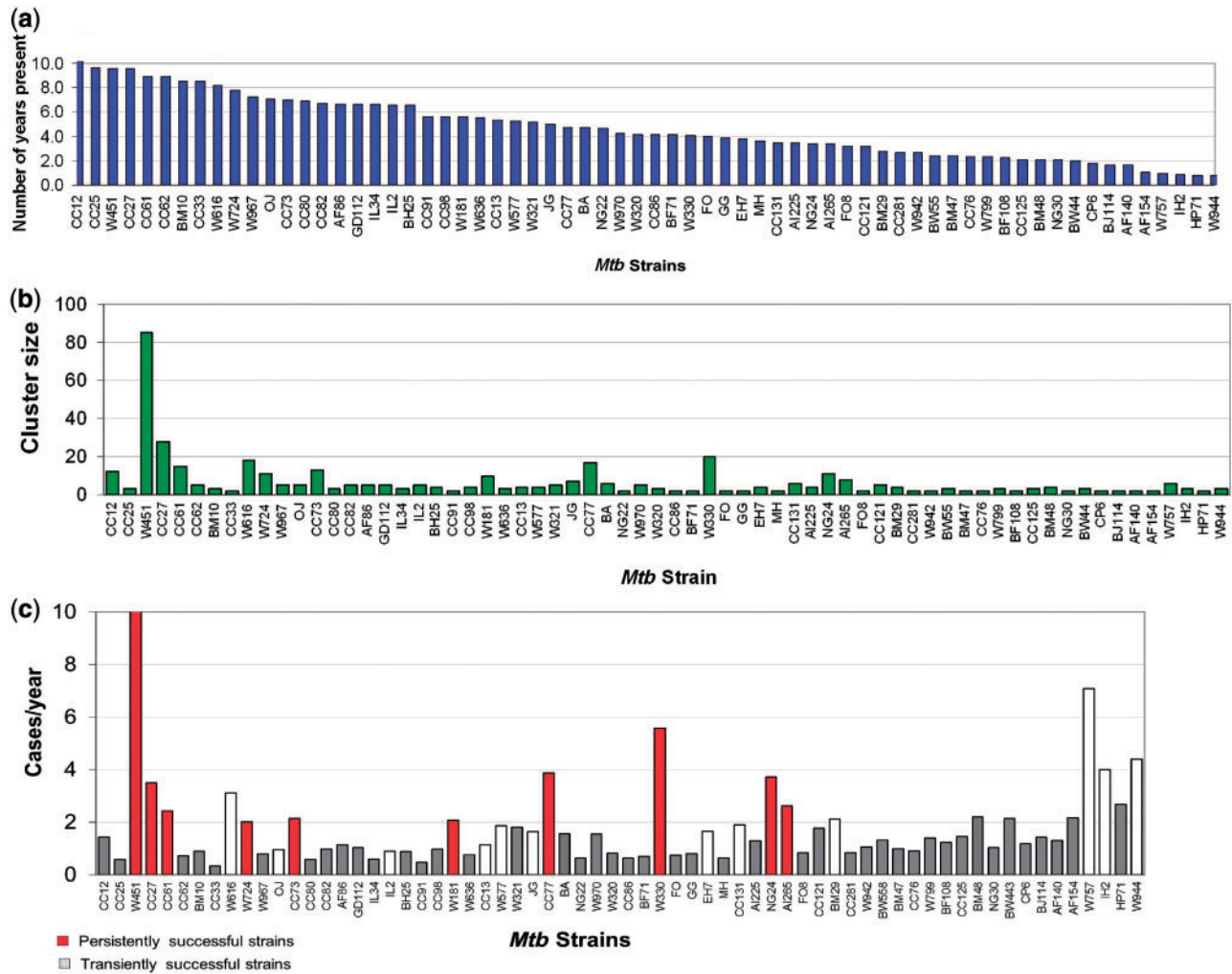


Figure 3. (a) Time (in years) in the community of clustered strains. Only strains present in the community for ≥ 2 years are illustrated. (b) Number of strains in each cluster. (c) Mean number of cases per year by cluster

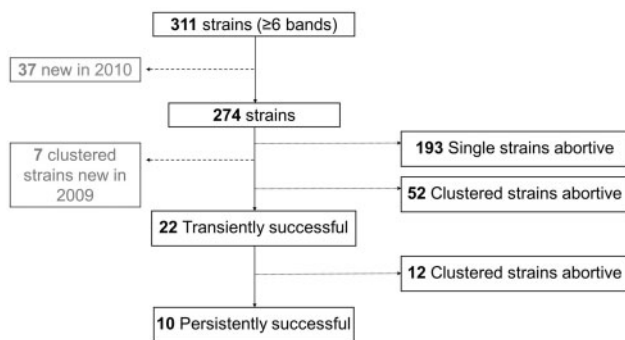


Figure 4. The success of strains in the study community

cluster density of ≥ 1 cases/year, but did not remain at a definition of cluster density of ≥ 3 cases/year ($P=0.16$). The association between transiently successful strains and drug resistance remained at both revised definitions of persistently successful strains. The association with drug

resistance also remained if transiently successful strains were re-defined as strains occurring in 3 consecutive years and at a cluster density of ≥ 2 cases/year.

In sensitivity analyses restricted to a study period of 2006 to 2010, the association between transiently successful strains and drug resistance remained ($P=0.002$ for any drug resistance and $P=0.04$ for MDR-TB), but the association between persistently successful strains and HIV infection did not ($P=0.24$).

Discussion

This study extends traditional cluster analysis by incorporating the concepts of cluster density and persistence over time in an assessment of *Mtb* strain success in a high-prevalence community. We described high *Mtb* strain diversity with 330 strains identified, and all nine sSNP-defined phylogenetic lineages represented in this cohort. This finding is consistent with reports from other Southern African

Table 2. Characteristics of persistently successful strains compared with those of transiently successful and unsuccessful strains

	Persistently successful <i>n</i> = 218 <i>n</i> (%)	Transiently successful <i>n</i> = 69 <i>n</i> (%)	Unsuccessful <i>n</i> = 353 <i>n</i> (%)	OR (95% CI) ^a	OR (95% CI) ^b
Patient characteristics					
Age, years, median (IQR)	32 (26–39)	34 (29–41)	33 (26–39)	0.97 (0.95–1.00)	0.9 (0.9–1.0)
Gender (female)	95 (44%)	26 (38%)	147 (41%)	1.3 (0.7–2.2)	1.1 (0.8–1.5)
HIV-positive	137 (69%)	36 (60%)	190 (60%)	1.5 (0.8–2.7)	1.5 (1.0–2.2)*
Receiving ART	36 (26%)	9 (25%)	42 (22%)	1.1 (0.5–2.5)	1.3 (0.8–2.1)
Median CD4 ^c (IQR)	129 (33–245)	143 (58–327)	132 (68–243)	0.99 (0.99–1.00)	0.99 (0.9–1.0)
TB characteristics					
Re-treatment TB	69 (32%)	23 (33%)	90 (26%)	0.8 (0.5–1.6)	1.4 (0.9–2.0)
Smear-positive TB	180 (83%)	56 (81%)	294 (83%)	1.1 (0.5–2.2)	0.9 (0.6–1.5)
Treatment completion	180 (83%)	45 (65%)	266 (76%)	2.5 (1.4–4.6)*	1.5 (1.0–2.3)*
Mortality	2 (0.9%)	5 (7.3%)	16 (4.6%)	0.1 (0.02–0.6)*	0.2 (0.04–0.9)*
MDR	1 (0.5%)	9 (13%)	11 (3%)	0.03 (0.004–0.2)*	0.1 (0.02–1.1)
Any drug-resistant pattern	4 (1.8%)	13 (19%)	22 (6%)	0.08 (0.03–0.3)*	0.3 (0.1–0.8)*
Social characteristics^d					
	<i>n</i> = 173	<i>n</i> = 47	<i>n</i> = 238		
Employed	55 (32%)	17 (36%)	78 (33%)	1.2 (0.6–2.3)	1.0 (0.7–1.6)
Informal housing	149 (86%)	43 (91%)	217 (91%)	1.7 (0.6–5.2)	1.7 (0.9–3.1)
Number of rooms used for sleeping, median (IQR)	1 (1–2)	1 (1–2)	1 (1–2)	1.2 (0.8–2.0)	0.9 (0.7–1.2)
Number of adults in house, median (IQR)	1 (1–2)	1 (1–2)	1 (1–2)	1.0 (0.8–1.3)	0.9 (0.8–1.1)
Use minibus taxi for transport	165 (95%)	44 (94%)	222 (93%)	1.7 (0.2–2.8)	0.7 (0.3–1.6)
Number of times taxi in past month, median (IQR)	3 (2–12)	2 (2–4)	2 (1–8)	1.0 (0.9–1.1)	1.0 (0.9–1.0)

^aUnadjusted comparison of transiently successful strains to persistently successful strains.

^bUnadjusted comparison of unsuccessful strains to persistently successful strains.

^cCD4 count measured within ± 90 days of TB diagnosis.

^d458 (72%) patients completed the epidemiological questionnaire.

* $P < 0.05$.

settings.^{3,4} The study period enabled the evaluation of *Mtb* clustering and strain success over 10 years.³ Of the 267 *Mtb* strains analysed, only 4% were persistently successful. Strain success was negatively associated with drug-resistant TB.

The substantial diversity reported in this study and in other South African settings^{3,4} may indicate high levels of competition for success between strains in high TB-prevalent areas. This is supported by the low percentage of successful strains, despite high transmission rates.²⁰

Household crowding, transport and employment were not identified as predictors of success. However, it should be noted that our assessment utilized crude or surrogate measures for possible social influences, and more in-depth studies of these factors are required. Host factors of age and gender were not associated with strain success. However, there was a trend towards patients with successful strains having increased odds of HIV infection. This finding may suggest a link between the success of individual strains and the immunological characteristics of the host, but this association was not robust in sensitivity analyses.

In addition to social and host factors, success may be driven by the third component of the traditional epidemiological triad: pathogen characteristics.²¹ Molecular epidemiological studies have shown that some strains may spread and cause disease more efficiently than others,^{22,23} and large clusters may be indicative of high transmissibility or high pathogenicity. This study was unable to demonstrate increased transmission of successful strains as we did not investigate the number of individuals infected per case, and there was no association between sputum positivity and strain success. However, our findings suggest that strain success may be associated with increased bacterial pathogenicity. Laboratory-based studies have reported increased potential for intracellular growth in W-Beijing strains that may be associated with increased pathogenicity^{22,23} and a clinical cohort study reported increased pathogenicity linked to W-Beijing strains.²⁴ In this study, a range of strain families were persistently successful, indicating that particular strain families may not be predictive of success.

Successful strains were negatively associated with drug-resistant TB. The role of DR-TB in success of *Mtb* strains

is complex. Whereas diagnostic and treatment delays may increase DR-TB prevalence in communities, increased mortality associated with MDR-TB may negatively affect transmission and it is possible that drug resistance confers fitness cost on strains.^{4,25} Another Cape Town community has reported an association between drug susceptibility and increasing representation of W-Beijing strains. However, that study was from a low HIV-prevalence setting, and did not report on association of W-Beijing success with host and social factors.⁴

With the exception of strain W616, we found no evidence of high rates of transmission of the DR-TB strains suggestive of outbreaks. This community may therefore provide an opportunity for better understanding of DR-TB occurring in a hyper-endemic setting, and not associated with a DR-TB outbreak or nosocomial transmission. Recent literature indicates that a large proportion of DR-TB is due to transmission, rather than acquired resistance due to poor TB management,^{25,26} and other data from this community support these findings.²⁷

This study has several limitations. Genotyping data was not obtained on all eligible patients in this community and the incomplete sampling may have introduced selection bias. However, all known characteristics and confounders were assessed and most were found to be equally distributed across the two groups. We were less likely to obtain genotyping data in patients who died, possibly due a lower sputum retrieval rate from the local hospital where sicker patients were diagnosed, less time to obtain a specimen from patients before they died or the severity of disease resulting in inability to produce sputum specimens. We were more likely to obtain specimens from patients with MDR-TB, and this may reflect multiple specimens collected for drug susceptibility testing. RFLP analysis poorly differentiates strains with more than six bands and may therefore overestimate clustering. For this reason strains with more than six bands were excluded from the analysis. Sampling rates increased over the study period and this variation would impact on cluster density calculations. Although an incomplete adjustment for missing data, the cluster density calculations were adjusted by a sampling correction factor. In addition, a sensitivity analysis was performed, restricted to the period of more complete sampling (2006 to 2010) and no change was found in study inferences. The definitions of strain success were generated according to descriptive analysis of the data. However, sensitivity analyses were performed and the negative association between strain success and drug resistance persisted in these analyses.

In conclusion, in this community with a high *Mtb* strain diversity, only 4% of strains were persistently successful. Social characteristics did not appear to be associated with strain success and, of the host factors assessed, only

HIV-infection may be associated with success. The role of pathogen characteristics in strain success appears of greatest import and was supported by the negative association of success with DR-TB. This study supports the need for further investigations into the role of pathogen characteristics in strain success. Such studies may utilize whole-genome sequencing and *in vitro* and *in vivo* assays to evaluate the role of strain evolution and pathogenicity in success.

Funding

The National Institutes of Health supported this work [Comprehensive Integrated Programme of Research on AIDS grant 1U19AI053217 to K.M., L.G.B., L.M. and R.W., 1U19AI05321, and RO1 AI058736-02 to RW, RO1 AI66046, RO1 AI 080737 to B.K. and RO1 AI 54361 to G.K.]. K.M. was also supported by the Wellcome Trust [Strategic award: Clinical Infectious Diseases Research Initiative WT084323M] and a Hasso Plattner Foundation award via the University of Cape Town.

Conflict of interest: None declared.

References

1. World Health Organization. *Global Tuberculosis Report*. 2012. http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502_eng.pdf (27 November 2013, date last accessed).
2. Williams BG. TB and HIV: deadly liaison or manageable threat? *Sci Transl Med* 2012;**4**:135 fs15.
3. Warren R, Hauman J, Beyers N *et al*. Unexpectedly high strain diversity of *Mycobacterium tuberculosis* in a high-incidence community. *S Afr Med J* 1996;**86**:45–49.
4. van der Spuy GD, Kremer K, Ndabambi SL *et al*. Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. *Tuberculosis (Edinb)* 2009;**89**: 120–25.
5. Lockman S, Sheppard JD, Braden CR *et al*. Molecular and conventional epidemiology of *Mycobacterium tuberculosis* in Botswana: a population-based prospective study of 301 pulmonary tuberculosis patients. *J Clin Microbiol* 2001;**39**:1042–47.
6. Verver S, Warren RM, Munch Z *et al*. Transmission of tuberculosis in a high incidence urban community in South Africa. *Int J Epidemiol* 2004;**33**:351–57.
7. Glynn JR, Crampin AC, Traore H *et al*. Determinants of cluster size in large, population-based molecular epidemiology study of tuberculosis, northern Malawi. *Emerg Infect Dis* 2008;**14**: 1060–66.
8. Glynn JR, Alghamdi S, Mallard K *et al*. Changes in *Mycobacterium tuberculosis* genotype families over 20 years in a population-based study in Northern Malawi. *PLoS One* 2010;**5**:e12259.
9. Wood R, Middelkoop K, Myer L *et al*. Undiagnosed tuberculosis in a community with high HIV prevalence: implications for tuberculosis control. *Am J Respir Crit Care Med* 2007;**175**:87–93.
10. Middelkoop K, Bekker LG, Myer L *et al*. Antiretroviral therapy and TB notification rates in a high HIV prevalence South African community. *J Acquir Immune Defic Syndr* 2011;**56**:263–69.

11. Middelkoop K, Bekker LG, Myer L *et al.* Antiretroviral program associated with reduction in untreated prevalent tuberculosis in a South African township. *Am J Respir Crit Care Med* 2010;**182**: 1080–85.
12. Department of Health Republic of South Africa. *National Tuberculosis Management Guidelines*. 2009; http://familymedicine.ukzn.ac.za/Libraries/Guidelines_Protocols/TB_Guidelines_2009.sflb.ashx (27 November 2013, date last accessed).
13. Health Systems Trust. *Cape Town TB Control. Progress report 1997-2003*. 2004. www.hst.org.za/publications/618 (27 November 2013, date last accessed).
14. van Embden JD, Cave MD, Crawford JT *et al.* Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;**31**:406–09.
15. Bifani PJ, Mathema B, Liu Z *et al.* Identification of a W variant outbreak of Mycobacterium tuberculosis via population-based molecular epidemiology. *JAMA* 1999;**282**:2321–27.
16. Gutacker MM, Mathema B, Soini H *et al.* Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. *J Infect Dis* 2006;**193**:121–28.
17. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev* 2006;**19**:658–85.
18. Cox DR, Stuart A. Some quick sign tests for trend in location and dispersion. *Biometrika* 1955;**42**:80–95.
19. Gagneux S, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *Lancet Infect Dis* 2007;**7**:328–37.
20. Middelkoop K, Bekker LG, Liang H *et al.* Force of tuberculosis infection among adolescents in a high HIV and TB prevalence community: a cross-sectional observation study. *BMC Infect Dis* 2011;**11**:156.
21. Mathema B, Kurepina N, Yang G *et al.* Epidemiologic consequences of microvariation in Mycobacterium tuberculosis. *J Infect Dis* 2012;**205**:964–74.
22. Hanekom M, van der Spuy GD, Streicher E *et al.* A recently evolved sublineage of the Mycobacterium tuberculosis Beijing strain family is associated with an increased ability to spread and cause disease. *J Clin Microbiol* 2007;**45**:1483–90.
23. Wu S, Barnes PF, Samten B *et al.* Activation of the eis gene in a W-Beijing strain of Mycobacterium tuberculosis correlates with increased SigA levels and enhanced intracellular growth. *Microbiology* 2009;**155**:1272–81.
24. de Jong BC, Hill PC, Aiken A *et al.* Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. *J Infect Dis* 2008;**198**:1037–43.
25. Luciani F, Sisson SA, Jiang H, Francis AR, Tanaka MM. The epidemiological fitness cost of drug resistance in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2009;**106**:14711–15.
26. Kruuner A, Hoffner SE, Sillastu H *et al.* Spread of drug-resistant pulmonary tuberculosis in Estonia. *J Clin Microbiol* 2001;**39**: 3339–45.
27. Middelkoop K, Bekker LG, Shashkina E, Kreiswirth B, Wood R. Retreatment tuberculosis in a South African community: the role of re-infection, HIV and antiretroviral treatment. *Int J Tuberc Lung Dis* 2012;**16**:1510–16