CrossMark

*Epidemiol. Infect.* (2017), **145**, 503–514. © Cambridge University Press 2016
doi:10.1017/S0950268816002399

# Exploring genotype concordance in epidemiologically linked cases of tuberculosis in New York City

R. S. ROBBINS[1], B. R. PERRI[1], S. D. AHUJA[1], H. A. ANGER[2], J. SULLIVAN MEISSNER[1], E. SHASHKINA[3], B. N. KREISWIRTH[3] AND D. C. PROOPS[1]*

[1] *New York City Department of Health and Mental Hygiene, New York, NY, USA*
[2] *Gynuity Health Projects, New York, NY, USA*
[3] *Public Health Research Institute Tuberculosis Center, Rutgers University, Newark, NJ, USA*

*Received 17 May 2016; Final revision 19 September 2016; Accepted 24 September 2016; first published online 21 November 2016*

## SUMMARY

Comparing genotype results of tuberculosis (TB) isolates from individuals diagnosed with TB can support or refute transmission; however, these conclusions are based upon the criteria used to define a genotype match. We used a genotype-match definition which allowed for variation in IS*6110* restriction fragment length polymorphism (RFLP) to support transmission between epidemiologically linked persons. Contacts of individuals with infectious TB (index cases) diagnosed in New York City from 1997 to 2003 who subsequently developed TB (contact cases) from 1997 to 2007 were identified. For each contact case and index case (case-pair), isolate genotypes (spoligotype and RFLP results) were evaluated. Isolates from case-pairs were classified as exact or non-exact genotype match. Genotypes from non-exact match case-pairs were reviewed at the genotyping laboratory to determine if the isolates met the near-genotype-match criteria (exactly matching spoligotype and similar RFLP banding patterns). Of 118 case-pairs identified, isolates from 83 (70%) had exactly matching genotypes and 14 (12%) had nearly matching genotypes (supporting transmission), while the remaining 21 (18%) case-pairs had discordant genotypes (refuting transmission). Using identical genotype-match criteria for isolates from case-pairs epidemiologically linked through contact investigation may lead to underestimation of transmission. TB programmes should consider the value of expanding genotype-match criteria to more accurately assess transmission between such cases.

**Key words**: Genotyping, transmission, tuberculosis.

## INTRODUCTION

Contact investigation around infectious tuberculosis (TB) cases can decrease or eliminate future TB transmission through identification and treatment of TB infection in contacts of TB cases. Historically, the strongest indication of TB transmission has been the diagnosis of active TB in a contact of an infectious case. Genotyping results for *Mycobacterium tuberculosis* isolates can support or refute transmission assumptions between epidemiologically linked cases [1–4].

Transmission assessments based on genotype comparisons depend both on the genotyping methods used and how genotype concordance is defined. IS*6110*-based restriction fragment length polymorphism (RFLP) analysis has been a widely used genotyping method to

* Address for correspondence: D. Proops, MD, MPH, Bureau of Tuberculosis Control, Office of Surveillance and Epidemiology, NYC Department of Health and Mental Hygiene, 42-09 28th Street, 21st Floor, CN-72B, Long Island City, NY 11101-4132, USA.
(Email: dproops2@health.nyc.gov)

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

define TB strains. There has been conflicting data on the validity of clustering TB isolates of *nearly* matching genotype, even when prior evidence of transmission exists. Some studies have shown that the IS*6110* site is relatively stable and the rate of gain or loss of IS*6110* is estimated to be low [5]. Employing match criteria other than requiring an exact match between cases may result in an overestimation of transmission when utilizing RFLP [6, 7]. However, studies examining serial isolates obtained from the same patient or from a known transmission chain have shown IS*6110* pattern changes [8–10]. Accounting for these events can impact transmission assessment between epidemiologically linked TB cases [7, 8, 11–15].

There has been limited description of how expanding genotype concordance definitions impact transmission assessments in epidemiologically linked TB cases [16–18]. Studies that have evaluated *M. tuberculosis* genotypic relationships between linked cases are often from high-incidence countries [19], focus on cluster investigations [20, 21], or determine strain relatedness by only one molecular method [2, 7, 10, 16, 21, 22]. While exact-matching genotype concordance criteria (using IS*6110* patterns) have traditionally been used to characterize transmission between linked cases, this does not account for IS*6110* changes that may occur during or after TB transmission [17]. To better understand TB transmission dynamics in New York City (NYC), we reviewed index-case *M. tuberculosis* isolate genotyping results and those of their contacts who subsequently developed active TB across two molecular methods. To account for possible IS*6110* changes, we explored the use of an expanded genotype concordance definition, and estimated the additional transmission this would reveal. We anticipated that contacts that develop TB a short time after being exposed to the index case are more likely to have isolates with an exactly matching genotype than those that develop TB after a longer period of time.

## MATERIALS AND METHODS

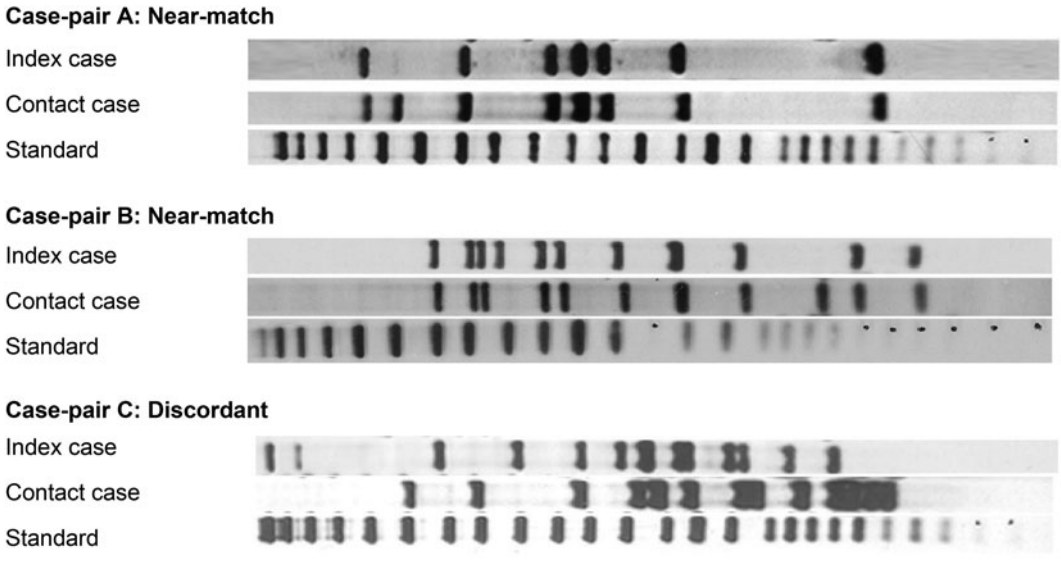
### Study population

The NYC Department of Health and Mental Hygiene TB Registry contains information on TB cases reported in NYC as well as on persons identified as having been exposed to an infectious TB case (contacts) during contact investigation. This retrospective cohort study is based on a TB preventability analysis that includes both index cases and their associated contact cases (contacts of an index case that subsequently developed active TB). Study population selection methods have been described previously [23]. In brief, the NYC TB Registry was used to identify contacts of TB cases (aged $\geqslant 5$ years) that were diagnosed with TB in NYC from 1 January 1997 to 31 December 2003. These contacts were then matched to TB cases diagnosed in NYC from 1 January 1997 to 31 December 2007 by name, sex, date of birth, and country of birth. Included contact cases must have been living in NYC when identified as a contact and could not have been treated for active TB in the year prior to the index case's diagnosis. Our inclusion criteria differ from those of the parent study in that we included contact cases diagnosed throughout the study period as well as multidrug-resistant TB cases and their contact cases [23]. To assess comparability, we compared study population contact-case demographics to those of the overall population of contact cases in NYC 1997–2007.

For individuals identified as a contact multiple times, the most recent contact event was used. We hypothesized that isolates from contact cases were more likely to be discordant with the index cases' isolates when more time elapsed between exposure to the index case and diagnosis of active TB disease. To assess this influence of time, we classified contact cases as either prevalent (active TB diagnosed up to 9 months after being identified as a contact) or incident (active TB diagnosed more than 9 months after being identified as a contact). During contact investigation, tuberculin skin tests (TSTs) were administered to contacts unless there was a documented positive TST result before the contact investigation (prior positive) or a prior TB diagnosis. If negative, the TST was repeated after the window period (8 weeks after last day of known exposure to the index case) to allow time for the immune system to manifest a response to a recent infection. From the contact record, TST results at the time of contact investigation were abstracted and classified as follows: positive ($\geqslant 5$ mm induration, obtained either during or after the window period); negative (<5 mm induration, obtained after the window period); window negative (a negative TST result during the window period with no subsequent test result); prior positive; prior TB diagnosis; or not tested.

We compared demographic (age, region of birth, sex, and race/ethnicity), clinical (time between index case and contact-case diagnoses, TB exposure setting, TST result, and HIV status at TB diagnosis), and

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press




[a] Near-match is defined as two RFLP patterns deemed by a laboratory-based genotyping expert to be closely related and differing by ≤2 bands. All other non-matching genotype pairs were categorized as discordant.
[b] A case-pair consists of an index case and an associated contact that subsequently developed TB (contact case).

**Fig. 1.** Examples of near-matching and discordant[a] restriction fragment-length polymorphism (RFLP) analysis results in tuberculosis (TB) case pairs[b], New York City, 1997–2007.

social characteristics (history of drug use and homelessness at TB diagnosis). To explore the effect of time on contact-case genotype, we stratified data by prevalent (identified as a case during the contact investigation of the index case) and incident (identified as a case following conclusion of contact investigation of index case) time periods.

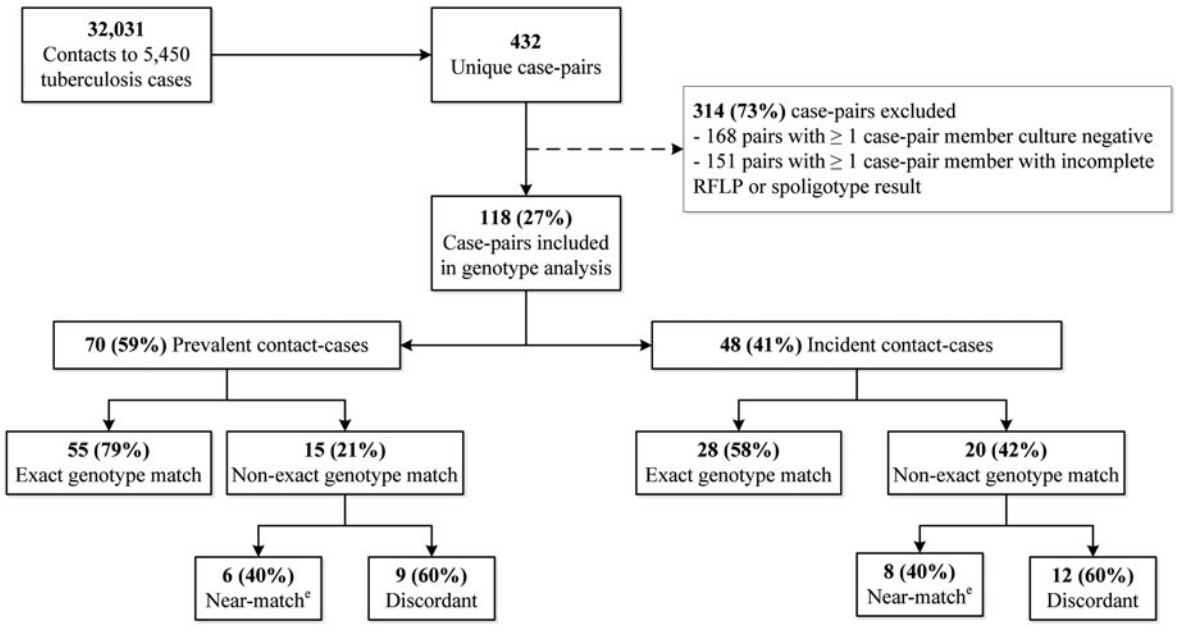
### Genotyping and concordance classification

Since 2001, all initial culture-positive TB isolates have been routinely genotyped [21, 24, 25]. During the study period, the NYC Health Department used two genotyping methods to characterize TB strains: spacer oligonucleotide typing (spoligotyping) and IS*6110* RFLP analysis, which were performed at the New York State Department of Health's Wadsworth Center in Albany, New York and at the Public Health Research Institute at Rutgers University in Newark, New Jersey, respectively. Details on both genotyping methods have been described previously [26–30].

Isolate genotype data were abstracted for index cases and associated contact cases (together considered a case-pair), and only case-pairs with complete genotype results (RFLP and spoligotype) were included in our analysis. Initially, case-pairs were classified as either an exact or a non-exact genotype match by examining both spoligotype and RFLP results. We performed bivariate analyses to compare contact cases' clinical, social, and demographic characteristics in exact genotype-match case-pairs to those of non-exact genotype-match case-pairs within the prevalent/incident classification of the contact case.

To account for genotype changes in TB bacteria that may have occurred during or after TB transmission [7, 17, 19], isolate genotypes of non-exact genotype-match case-pairs were re-evaluated. These case-pairs were further categorized as near-match or genotype-discordant based on a non-blinded review of the RFLP patterns by a TB genotyping expert (Fig. 1). A near-match genotype was defined as a case-pair with the same spoligotype and RFLP patterns deemed to be closely related and differing by $\leqslant 2$ bands [10, 23]. Case-pairs with genotype results that fell outside of the 'near-match' definition were classified as discordant. Based on this case-pair re-categorization, we repeated bivariate analyses of characteristics in exact genotype-match, near-match, and discordant contact cases with further stratification by prevalent or incident status.

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press




[a]TB genotype is defined as spoligotype and IS*6110* restriction fragment length polymorphism (RFLP) result.
[b]A prevalent contact case is defined as a contact case that was diagnosed with active tuberculosis within 9 months of the index case's diagnosis.
[c]An incident contact case is defined as a contact case that was diagnosed with active tuberculosis >9 months after the index case's diagnosis.
[d]A case-pair consists of an index case and an associated contact that subsequently developed TB (contact case).
[e]A near-match was defined as genotypes having exact-matching spoligotype results and RFLP patterns deemed to be closely related and differing by ≤2 bands.

**Fig. 2.** Genotype[a] concordance in prevalent[b] and incident[c] tuberculosis (TB) case-pairs[d] identified in New York City, 1997–2007.

### Statistical analysis

We used Pearson's $\chi^2$ and Fisher's exact tests for categorical data analyses and Wilcoxon rank-sum test for comparing medians; $P$ values <0·05 were considered statistically significant. All statistical analyses were performed using SAS v. 9.2 (SAS Institute Inc., USA).

## RESULTS

### Study population

Of 32 031 contacts of 5450 infectious TB cases reported in NYC during 1997–2003, 432 case-pairs were identified, 118 (27%) of which were included in the final study population (Fig. 2). These 118 contact cases were linked to 104 index cases (median of 1 contact case per index case, range 1–4; data not shown). Compared to all TB contact cases, the contact cases included in the study were less likely to be aged <5 years at TB diagnosis (6% *vs.* 18%, respectively, $P$ = 0·002) and more likely to be aged 18–44 years at TB diagnosis (44% *vs.* 58%, respectively, $P$ = 0·010). Additionally, included contact cases were significantly more likely to be born outside the United States (53% *vs.* 40% of all contact cases, $P$ = 0·012) (Table 1).

### Prevalent and incident contact cases

Of the 118 contact cases included, 70 (59%) were considered prevalent contact cases, and 48 (41%) were incident contact cases. Although there were no significant differences in demographic characteristics observed between incident and prevalent contact cases (Table 2), prevalent contact cases were more likely than incident contact cases to have had a positive TST result at the time of contact investigation (69% *vs.* 50% respectively, $P$ = 0·04), and less likely to have a history of homelessness at the time of TB

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

Table 1. *Demographic characteristics of tuberculosis (TB) cases identified as contacts (contact cases) to infectious TB cases (index cases), New York City, 1997–2007: all contact cases vs. study population contact cases*

| | All TB contact cases | | Study population contact cases | | |
|---|---|---|---|---|---|
| | *N* | % | *N* | % | *P* value[a] |
| Total | 432 | 100 | 118 | 100 | |
| Age group, years[b] | | | | | |
| 0–4 | 76 | 18 | 7 | 6 | 0·002 |
| 5–17 | 61 | 14 | 11 | 9 | 0·171 |
| 18–44 | 191 | 44 | 68 | 58 | 0·010 |
| 45–64 | 81 | 19 | 26 | 22 | 0·425 |
| ⩾65 | 23 | 5 | 6 | 5 | 0·918 |
| Median age[b] (range) | 29 (0–95) | | 33 (0–95) | | 0·146 |
| Female sex | 193 | 45 | 52 | 44 | 0·906 |
| Region of birth | | | | | |
| Unknown | 2 | 0 | 0 | 0 | 1·000 |
| Foreign-born | 171 | 40 | 62 | 53 | 0·012 |
| US-born[c] | 259 | 60 | 56 | 47 | 0·015 |
| Race/ethnicity (among US-born) | | | | | |
| Black non-Hispanic | 189 | 73 | 43 | 77 | 0·557 |
| White non-Hispanic | 50 | 19 | 8 | 14 | 0·380 |
| Hispanic | 8 | 3 | 2 | 4 | 0·694 |
| Asian | 10 | 4 | 3 | 5 | 0·709 |
| Other | 2 | 1 | 0 | 0 | 1·000 |

[a] *P* values generated by Pearson's $\chi^2$ or Fisher's exact tests for proportions, Wilcoxon rank sum test for medians.
[b] Age at TB diagnosis.
[c] Includes birth in US territories.

diagnosis (1% *vs.* 15% respectively, $P = 0{\cdot}01$, see Table 2).

### Case-pair genotype-match analyses

Although 82% ($n = 97$) of all case-pairs were ultimately categorized as near or exact genotype match, this proportion was greater in prevalent case-pairs than incident case-pairs; however, this difference was not significant (87% *vs.* 75% respectively, $P = 0{\cdot}090$. Prevalent case-pairs were more likely than incident case-pairs to be classified as exact genotype match (79% *vs.* 58%, respectively, $P = 0{\cdot}02$, Table 2). When applying our expanded genotype concordance definition to include non-exact genotype-match case-pairs, the proportion of prevalent and incident case-pairs reclassified as near-match genotypes was the same, at 40%.

### Patient comparisons by genotype concordance (expanded definition)

We examined contact-case characteristics by the expanded genotype concordance classifications within the prevalent/incident categorization. Overall, we found no significant demographic differences between exact- and near-genotype-match contact cases in either the prevalent or incident contact-case groupings (Tables 3 and 4).

## DISCUSSION

In this study, genotype concordance with the index case was found in 70% of the 118 contact cases (79% of the prevalent contact cases, and 58% in the incident contact cases) using an exact match criteria between epidemiologically linked case-pairs. However, by accounting for minimal changes in IS*6110* RFLP patterns, we found near-matching genotypes in an additional 12% ($n = 14$). Ultimately, using an expanded definition of genotype concordance, genotyping results supported TB transmission in 82% ($n = 97$) of all contact cases, highlighting the need for programmes to evaluate their criteria for determining what constitutes a genotype match when making transmission inferences.

Our study population of contact cases who had full genotyping results available differed slightly from all

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

Table 2. *Demographic, clinical, and social characteristics of tuberculosis (TB) cases identified as contacts (contact cases) to infectious TB cases (index cases) by prevalence status[a], New York City, 1997–2007*

| | Prevalent contact cases | | Incident contact cases | | |
|---|---|---|---|---|---|
| | *N* | % | *N* | % | *P* value[b] |
| Total | 70 | 100 | 48 | 100 | |
| Age group (years) at TB diagnosis | | | | | |
| 0–4 | 6 | 9 | 1 | 2 | 0·238 |
| 5–17 | 7 | 10 | 4 | 8 | 1·000 |
| 18–44 | 37 | 53 | 31 | 65 | 0·205 |
| 45–64 | 18 | 26 | 8 | 17 | 0·244 |
| ⩾65 | 2 | 3 | 4 | 8 | 0·223 |
| Median age in years at TB diagnosis (range) | 33 (0–95) | | 32·5 (4–81) | | 0·850 |
| Female sex | 33 | 47 | 19 | 40 | 0·417 |
| US-born[c] | 37 | 53 | 19 | 40 | 0·156 |
| Race/ethnicity (among US-born) | | | | | |
| Black non-Hispanic | 29 | 78 | 14 | 74 | 0·694 |
| White non-Hispanic | 6 | 16 | 2 | 11 | 0·565 |
| Hispanic | 1 | 3 | 1 | 5 | 0·625 |
| Asian | 1 | 3 | 2 | 11 | 0·218 |
| Household exposure setting | 52 | 74 | 29 | 60 | 0·111 |
| Tuberculin skin test (TST) result[d] | | | | | |
| Positive | 48 | 69 | 24 | 50 | 0·042 |
| Negative | 6 | 9 | 10 | 21 | 0·056 |
| No TST administered: prior TB diagnosis | 4 | 6 | 1 | 2 | 0·647 |
| No TST administered: prior positive TST | 8 | 11 | 4 | 8 | 0·759 |
| Window negative | 3 | 4 | 3 | 6 | 0·686 |
| Not tested | 1 | 1 | 6 | 13 | 0·018 |
| TST converted from negative to positive | | | | | |
| Eligible[e] | 8 | 11 | 14 | 29 | n.a. |
| Converted (among eligible) | 5 | 63 | 8 | 57 | n.a. |
| Initial genotype concordance status[f] | | | | | |
| Exact match | 55 | 78 | 28 | 58 | 0·018 |
| Non-exact match | 15 | 21 | 20 | 42 | |
| Final genotype concordance status[g] | | | | | |
| Exact or near-match | 61 | 87 | 36 | 75 | 0·090 |
| Discordant | 9 | 13 | 12 | 25 | |
| HIV status at TB diagnosis | | | | | |
| Infected | 15 | 21 | 8 | 17 | 0·638 |
| Not infected | 46 | 66 | 31 | 65 | 0·899 |
| Unknown | 9 | 13 | 9 | 19 | 0·381 |
| History of homelessness ever | | | | | |
| Yes | 1 | 1 | 7 | 15 | 0·008 |
| No | 6 | 9 | 21 | 44 | <0·001 |
| Missing | 63 | 90 | 20 | 42 | <0·001 |
| History of illicit drug use ever | | | | | |
| Yes | 12 | 17 | 7 | 15 | 0·710 |
| No | 56 | 80 | 40 | 83 | 0·648 |
| Missing | 2 | 3 | 1 | 2 | 0·793 |

n.a., Not applicable.

[a] Prevalent contact cases were diagnosed within 9 months of the date of diagnosis of the associated index case. Incident contact cases were diagnosed >9 months after the date of diagnosis of the associated index case.

[b] *P* values generated by Pearson's $\chi^2$ or Fisher's exact tests for proportions, Wilcoxon rank sum test for medians.

[c] Includes birth in US territories.

[d] TST result when contact case was originally evaluated as a contact. Contacts that had a negative TST result in the window period (within 8 weeks of last known date of exposure) but did not have a subsequent test were assigned a window negative TST result.

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

Table 2 notes (*cont.*)

[e] Contacts were eligible for TST conversion if they either had a known TST induration result within 2 years of the first TST after being identified as a contact, or if they had a negative (<5 mm induration) TST during the window period and then a second TST after the window period. An increase of 10 mm induration between the TST qualifies as a conversion.
[f] Contact case and index case genotypes are comprised of both spoligotype and IS*6110* restriction fragment length polymorphism (RFLP) results. Genotype concordance was determined by comparing the genotypes of case-pairs (consisting of the contact case and the associated index case). Initially, genotypes that matched exactly were considered exact matches and all others were categorized as non-exact matches.
[g] In non-exact matches, contact cases were further categorized as either near-match (case-pair with the same spoligotype and RFLP patterns differing by $\leqslant 2$ bands). Non-exact matches that did not meet the near-match definition were classified as discordant.

TB contact cases initially identified, namely in that there are fewer included contact cases who were aged <5 years at the time of diagnosis and included contact cases were more likely to be foreign-born. Due to the nature of our data, these results are expected. Patients aged <5 years diagnosed with TB are not likely to produce a culture-positive sputum sample, which is necessary to perform genotyping. Additionally, universal genotyping was mandated in NYC in 2001 [21, 24, 25]. Since that time, the majority of TB cases diagnosed and reported in NYC have been in foreign-born populations [31]. Therefore, there would be a smaller pool of US-born TB patients that met our eligibility criteria.

TB isolate genotype concordance has historically been used as a potential indicator of TB transmission within a specific population; a supposition that is strengthened when cases are epidemiologically linked by contact investigation. However, even in linked cases, the definition of genotype concordance influences transmission assessments. The inclusion of the near-match genotype in contact cases more accurately captures transmission events, accounting for IS*6110* changes that can occur over time or when a TB strain is transmitted [14, 17]. When changes occur, most studies estimated that these alterations occur at a higher rate when active TB disease has developed and prior to effective anti-TB treatment, when replication of the *M. tuberculosis* bacterium slows [14].

In prevalent case-pairs, we expect genotype concordance, as prevalent contact cases had a documented exposure to TB and were diagnosed shortly thereafter. The finding of discordant genotype in nine prevalent contact cases is unexpected. Of these, six tested TST positive during contact investigation, and three converted their TST. Although we typically categorize conversion of TST as evidence of recent transmission, it is possible that this instead represented a boosted TST response of a remote infection. All nine were born in countries with a TB incidence rate $\geqslant 15$ times that of the United States [32], and undocumented TB infection prior to identification as a contact in NYC is possible.

As expected, compared to prevalent case-pairs, we found increased *M. tuberculosis* genotype discordance in incident case-pairs. These incident contact cases (most of whom were born in a high-TB incident country of birth [32]) had more time to either reactivate a latent infection (acquired prior to identification as a contact) or to have been infected (or re-infected) due to undocumented TB exposures subsequent to their identification as a contact in our study. Surprisingly, we did not find a statistically significant over-representation of traditional TB risk factors (e.g. age <5 years, birth in foreign country, homelessness, HIV infection, etc.) in incident genotype discordant contact cases when compared to exact-match or near-match incident contact cases.

We attempted to account for the relative stability of the TB genome, but at the same time to allow for the possibility of minor genomic changes over time by including an additional genotyping method, spoligotyping, in the overall genotype result of isolates in our study. The finding of the same proportion of incident and prevalent near-match case-pairs (40%) in case-pairs initially classified as a non-exact genotype match was unexpected, as we anticipated that the incident contact cases would have had increased opportunity for genotype change. This finding indicates that further study of *M. tuberculosis* genotypic changes over time is warranted.

Our results differ from a similar study conducted in San Francisco in 1998 [16]. The study authors, who defined patients as having matching genotypes if the IS*6110* patterns were either the same or differed by one band, found a 30% genotype discordance proportion, a twofold increase compared to our study. These conflicting results may be explained by the differences in the definition of genotype concordance as well as differences in the study populations. During the study

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

Table 3. *Demographic, clinical, and social characteristics of prevalent contact-cases[a] by index case isolate genotype concordance[b]*

| | Exact genotype match | | Near genotype match | | *P* value (exact *vs.* near) | Discordant genotype | | *P* value (exact *vs.* discordant) |
|---|---|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | | *N* | % | |
| Total | 55 | 100 | 6 | 100 | | 9 | 100 | |
| Age group (years) at TB diagnosis | | | | | | | | |
| 0–4 | 4 | 7 | 2 | 33 | 0·102 | 0 | 0 | 1·000 |
| 5–17 | 7 | 13 | 0 | 0 | 1·000 | 0 | 0 | 0·580 |
| 18–44 | 29 | 53 | 2 | 33 | 0·425 | 6 | 67 | 0·494 |
| 45–64 | 14 | 26 | 2 | 33 | 0·648 | 2 | 22 | 1·000 |
| ⩾65 | 1 | 2 | 0 | 0 | 1·000 | 1 | 11 | 0·263 |
| Median age (years) at TB diagnosis (range) | 33 (0–95) | | 28·5 (1–52) | | 0·417 | 43 (19–77) | | 0·285 |
| Female sex | 26 | 47 | 3 | 50 | 1·000 | 4 | 44 | 1·000 |
| US-born[c] | 28 | 51 | 5 | 83 | 0·205 | 4 | 44 | 1·000 |
| Race/ethnicity (among US-born) | | | | | | | | |
| Black non-Hispanic | 20 | 71 | 5 | 100 | 0·302 | 4 | 100 | 0·550 |
| White non-Hispanic | 6 | 21 | 0 | 0 | 0·556 | 0 | 0 | 0·566 |
| Hispanic | 1 | 4 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| Asian | 1 | 4 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| Median days to TB diagnosis (range) | 54 (1–272) | | 79 (4–125) | | 0·634 | 77 (22–271) | | 0·721 |
| Household exposure setting | 40 | 73 | 5 | 83 | 1·000 | 7 | 78 | 1·000 |
| Tuberculin skin test (TST) result[d] | | | | | | | | |
| Positive | 38 | 69 | 4 | 67 | 1·000 | 6 | 67 | 1·000 |
| Negative | 4 | 7 | 0 | 0 | 1·000 | 2 | 22 | 0·196 |
| No TST administered: prior TB diagnosis | 4 | 7 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| No TST administered: prior positive TST | 7 | 13 | 1 | 17 | 1·000 | 0 | 0 | 0·580 |
| Window negative | 2 | 4 | 1 | 17 | 0·271 | 0 | 0 | 1·000 |
| Not tested | 0 | 0 | 0 | 0 | n.a. | 1 | 11 | 0·141 |
| TST converted from negative to positive | | | | | | | | |
| Eligible[e] | 4 | 7 | 0 | 0 | n.a. | 4 | 44 | n.a. |
| Converted (among eligible) | 2 | 50 | 0 | 0 | n.a. | 3 | 75 | n.a. |
| HIV status at TB diagnosis | | | | | | | | |
| Infected | 12 | 22 | 1 | 17 | 1·000 | 2 | 22 | 1·000 |
| Not infected | 35 | 64 | 4 | 67 | 1·000 | 7 | 78 | 0·707 |
| Unknown | 8 | 15 | 1 | 17 | 1·000 | 0 | 0 | 0·587 |
| History of homelessness ever | | | | | | | | |
| Yes | 1 | 2 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| No | 5 | 9 | 0 | 0 | 1·000 | 1 | 11 | 1·000 |
| Missing | 49 | 89 | 6 | 100 | 1·000 | 8 | 89 | 1·000 |
| History of illicit drug use ever | | | | | | | | |
| Yes | 10 | 18 | 0 | 0 | 0·577 | 2 | 22 | 0·672 |
| No | 43 | 78 | 6 | 100 | 0·588 | 7 | 78 | 1·000 |
| Missing | 2 | 4 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |

n.a., Not applicable.

[a] Prevalent contact cases were diagnosed within 9 months of the date of diagnosis of the associated index case.

[b] Contact-case and index-case genotypes are comprised of both spoligotype and IS*6110* restriction fragment length polymorphism (RFLP) results. Genotype concordance was determined by comparing the genotypes of case-pairs (consisting of the contact case and the associated index case). Genotypes that matched exactly were considered genotype-concordant. Near-match contact cases were originally categorized as genotype discordant in Table 2. Near match defined as a difference in no more than 2 bands (but in the same family) between index and contact-case isolate genotypes that share identical spoligotypes. All remaining were categorized as genotype discordant.

[c] Includes birth in US territories.

[d] TST result when contact case was originally evaluated as a contact. Contacts that had a negative TST result in the window period (within 8 weeks of last known date of exposure) but did not have a subsequent test were assigned a window negative TST result.

[e] Contacts were eligible for TST conversion if they either had a known TST induration result within 2 years of the first TST after being identified as a contact, or if they had a negative (<5 mm induration) TST during the window period and then a second TST after the window period. An increase of 10 mm induration between the TST qualifies as a conversion.

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

Table 4. *Demographic, clinical, and social characteristics of incident contact cases[a] by index-case isolate genotype concordance status[b]*

| | Exact genotype match | | Near genotype match | | *P* value (exact *vs.* near) | Discordant genotype | | *P* value (exact *vs.* discordant) |
|---|---|---|---|---|---|---|---|---|
| | *N* | % | *N* | % | | *N* | % | |
| Total | 28 | 100 | 8 | 100 | | 12 | 100 | |
| Age group (years) at TB diagnosis | | | | | | | | |
| 0–4 | 1 | 4 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| 5–17 | 2 | 7 | 1 | 13 | 0·541 | 1 | 8 | 1·000 |
| 18–44 | 20 | 71 | 4 | 50 | 0·397 | 7 | 58 | 0·476 |
| 45–64 | 4 | 14 | 2 | 25 | 0·596 | 2 | 17 | 1·000 |
| ⩾65 | 1 | 4 | 1 | 13 | 0·400 | 2 | 17 | 0·210 |
| Female sex | 11 | 39 | 4 | 50 | 0·694 | 4 | 33 | 1·000 |
| US-born[c] | 11 | 39 | 4 | 50 | 0·694 | 4 | 33 | 1·000 |
| Race/ethnicity (among US-born) | | | | | | | | |
| Black non-Hispanic | 7 | 64 | 3 | 75 | 1·000 | 4 | 100 | 0·517 |
| White non-Hispanic | 2 | 18 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| Hispanic | 1 | 9 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| Asian | 1 | 9 | 1 | 25 | 0·476 | 0 | 0 | 1·000 |
| Household exposure when identified as a contact | 16 | 57 | 7 | 88 | 0·213 | 6 | 50 | 0·681 |
| Tuberculin skin test (TST) result[d] | | | | | | | | |
| Positive | 15 | 54 | 4 | 50 | 1·000 | 5 | 42 | 0·731 |
| Negative | 6 | 21 | 1 | 13 | 1·000 | 3 | 25 | 1·000 |
| No TST administered: prior TB diagnosis | 1 | 4 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |
| No TST administered: prior positive TST | 1 | 4 | 3 | 38 | 0·028 | 0 | 0 | 1·000 |
| Window negative | 3 | 11 | 0 | 0 | 1·000 | 0 | 0 | 0·541 |
| Not tested | 2 | 7 | 0 | 0 | 1·000 | 4 | 33 | 0·055 |
| TST converted from negative to positive | | | | | | | | |
| Eligible[e] | 9 | 32 | 3 | 33 | n.a. | 2 | 18 | n.a. |
| Converted (among eligible) | 6 | 67 | 2 | 67 | n.a. | 0 | 0 | n.a. |
| HIV status at TB diagnosis | | | | | | | | |
| Infected | 3 | 11 | 0 | 0 | 1·000 | 5 | 42 | 0·039 |
| Not infected | 19 | 68 | 7 | 88 | 0·397 | 5 | 42 | 0·166 |
| Unknown | 6 | 21 | 1 | 13 | 1·000 | 2 | 17 | 1·000 |
| History of homelessness ever | | | | | | | | |
| Yes | 4 | 14 | 0 | 0 | 0·555 | 3 | 25 | 0·410 |
| No | 13 | 46 | 4 | 50 | 1·000 | 4 | 33 | 0·505 |
| Missing | 11 | 39 | 4 | 50 | 0·694 | 5 | 42 | 1·000 |
| History of illicit drug use ever | | | | | | | | |
| Yes | 3 | 11 | 1 | 13 | 1·000 | 3 | 25 | 0·341 |
| No | 24 | 86 | 7 | 88 | 1·000 | 9 | 75 | 0·410 |
| Missing | 1 | 4 | 0 | 0 | 1·000 | 0 | 0 | 1·000 |

n.a., Not applicable.

[a] Incident contact cases were diagnosed more than 9 months after the date of diagnosis of the associated index case.

[b] Contact-case and index-case genotypes are comprised of both spoligotype and IS*6110* restriction fragment length polymorphism (RFLP) results. Genotype concordance was determined by comparing the genotypes of case-pairs (consisting of the contact case and the associated index case). Genotypes that matched exactly were considered genotype-concordant. Near-match contact cases were originally categorized as genotype discordant in Table 2. Near match defined as a difference in no more than 2 bands (but in the same family) between index and contact-case isolate genotypes that share identical spoligotypes. All remaining were categorized as genotype discordant.

[c] Includes birth in US territories.

[d] TST result when contact case was originally evaluated as a contact. Contacts that had a negative TST result in the window period (within 8 weeks of last known date of exposure) but did not have a subsequent test were assigned a window negative TST result.

[e] Contacts were eligible for TST conversion if they either had a known TST induration result within two years of the first TST after being identified as a contact, or if they had a negative (<5 mm induration) TST during the window period and then a second TST after the window period. An increase of 10 mm induration between the TST qualifies as a conversion.

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

period in San Francisco (1991–1996), the majority of TB cases were diagnosed in foreign-born individuals [33], whereas in NYC, foreign-born predominance in cases did not occur until 1997 [31]. Contacts from high-incidence countries are more likely to have had TB exposures prior to their identification as a contact of a TB case in NYC compared to US-born contacts. Thus, the contacts from high-incidence countries may be less likely to be infected in the United States, which may explain why the San Francisco authors found higher discordance in their study population. Additionally, the study authors did not specify when contact cases were diagnosed with TB, and if a large number of included cases were diagnosed after a long period of time post-contact investigation, a higher rate of discordance would be expected.

This study had some limitations. Although our study reported on a relatively large number of epidemiologically linked case-pairs, we were limited in our ability to detect statistical significance in some variables of interest due to missing data. Misclassification of 'index' and 'contact' cases may have occurred. Some contact cases were named as a contact multiple times, but only the most recent contact event was used for this analysis, and contact cases may have been infected by an earlier index case. In prevalent case-pairs, the designated contact case may have been the true index case but was diagnosed at a later date. Similarly, a contact case may have been infected by an unidentified case.

By confining our 'near-match' definition to genotypes differing by $\leqslant 2$ bands, it is possible that genotypes in case-pairs in which transmission did occur were misclassified as discordant. This may be particularly true in strains with a high number of bands or complex banding patterns, where the number and position of bands may be difficult to accurately interpret in the laboratory [34]. We also did not consider additional alternative definitions of genotype concordance in our study (exact RFLP and near-match spoligotype or near-match RFLP and near-match spoligotype). However, only three case-pairs met either of these additional concordance definitions. Finally, studies have shown that individuals may harbour multiple strains of the *M. tuberculosis* bacterium [35]. Case-pair genotype concordance would therefore be directly dependent upon which of multiple strains was identified that such an individual may have had, and could have led to an increased proportion of case-pairs with discordant genotypes.

Despite limitations, our study covered nearly 10 years of TB case data and additionally included data on persons when they were identified as a contact, which few programmes routinely collect. The study included all case-pairs diagnosed during the study period that had complete isolate genotype data and is the largest of its kind to date.

Our study highlights the need for TB control programmes to further evaluate *M. tuberculosis* genotype data in previously linked cases. Nearly half (47%) of all non-exact genotype-match case-pairs were ultimately re-classified as near-match, emphasizing the importance of developing methods to account for changes in genotype to aid in transmission detection and assessment. Including TB contact events and genotyping data for all TB cases in a comprehensive TB registry facilitate TB control programmes' ability to assess transmission between epidemiologically linked cases. This study's methods provide an additional tool by which TB control programmes should determine when to conserve resources (when genotyping refutes transmission), or identify additional transmission by using alternative definitions of genotype concordance to support transmission assessments in epidemiologically linked cases.

## ACKNOWLEDGEMENTS

We acknowledge our laboratory partners at the Public Health Laboratory, the New York State Wadsworth Center and the Public Health Research Institute. In addition, the authors thank Lisa Trieu for assistance with data analysis and the field and clinic staff who conduct contact investigations at the New York City Department of Health and Mental Hygiene Bureau of Tuberculosis Control.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **Oelemann MC, *et al.*** Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *Journal of Clinical Microbiology* 2007; **45**: 691–697.

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

2. **Grant J, *et al.*** Tuberculosis genotyping – United States, 2004–2010. *Morbidity and Mortality Weekly Report* 2012; **61**: 723–725.
3. **Barnes PF, Cave MD.** Molecular epidemiology of tuberculosis. *New England Journal of Medicine* 2003; **349**: 1149–1156.
4. **Dahle UR, *et al.*** Tuberculosis in contacts need not indicate disease transmission. *Thorax* 2005; **60**: 136–137.
5. **van Soolingen D, *et al.*** Occurrence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *Journal of Clinical Microbiology* 1991; **29**: 2578–2586.
6. **Jonsson J, *et al.*** Comparison between RFLP and MIRU-VNTR genotyping of Mycobacterium tuberculosis strains isolated in Stockholm 2009 to 2011. *PLoS ONE* 2014; **9**: e95159.
7. **Niemann S, *et al.*** Stability of IS6110 restriction fragment length polymorphism patterns of Mycobacterium tuberculosis strains in actual chains of transmission. *Journal of Clinical Microbiology* 2000; **38**: 2563–2567.
8. **Benjamin WH Jr., *et al.*** Identification of a contaminating Mycobacterium tuberculosis strain with a transposition of an IS6110 insertion element resulting in an altered spoligotype. *Journal of Clinical Microbiology* 2001; **39**: 1092–1096.
9. **Allix-Beguec C, *et al.*** Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci for subtyping of Mycobacterium tuberculosis Beijing isolates. *Journal of Clinical Microbiology* 2014; **52**: 164–72.
10. **Cave MD, *et al.*** Epidemiologic import of tuberculosis cases whose isolates have similar but not identical IS6110 restriction fragment length polymorphism patterns. *Journal of Clinical Microbiology* 2005; **43**: 1228–1233.
11. **de Boer AS, *et al.*** Analysis of rate of change of IS6110 RFLP patterns of Mycobacterium tuberculosis based on serial patient isolates. *Journal of Infectious Diseases* 1999; **180**: 1238–44.
12. **de Boer AS, van Soolingen D, Borgdorff MW.** Genetic mutations occur gradually in in vivo populations of Mycobacterium tuberculosis bacteria. *Journal of Clinical Microbiology* 2001; **39**: 3814.
13. **de Boer AS, *et al.*** Genetic heterogeneity in Mycobacterium tuberculosis isolates reflected in IS6110 restriction fragment length polymorphism patterns as low-intensity bands. *Journal of Clinical Microbiology* 2000; **38**: 4478–4784.
14. **Warren RM, *et al.*** Calculation of the stability of the IS6110 banding pattern in patients with persistent Mycobacterium tuberculosis disease. *Journal of Clinical Microbiology* 2002; **40**: 1705–1708.
15. **Tanaka MM, Francis AR.** Methods of quantifying and visualising outbreaks of tuberculosis using genotypic information. *Infection, Genetics and Evolution* 2005; **5**: 35–43.
16. **Behr MA, *et al.*** Predictive value of contact investigation for identifying recent transmission of Mycobacterium tuberculosis. *American Journal of Respiratory and Critical Care Medicine* 1998; **158**: 465–469.
17. **Warren RM, *et al.*** Evolution of the IS6110-based restriction fragment length polymorphism pattern during the transmission of Mycobacterium tuberculosis. *Journal of Clinical Microbiology* 2002; **40**: 1277–1282.
18. **Benedetti A, *et al.*** How close is close enough? Exploring matching criteria in the estimation of recent transmission of tuberculosis. *American Journal of Epidemiology* 2010; **172**: 318–326.
19. **van der Spuy GD, *et al.*** Use of genetic distance as a measure of ongoing transmission of Mycobacterium tuberculosis. *Journal of Clinical Microbiology* 2003; **41**: 5640–5644.
20. **Lindquist S, *et al.*** Prioritizing tuberculosis clusters by genotype for public health action, Washington, USA. *Emerging Infectious Diseases* 2013; **19**: 493–496.
21. **McNabb SJ, *et al.*** Added epidemiologic value to tuberculosis prevention and control of the investigation of clustered genotypes of Mycobacterium tuberculosis isolates. *American Journal of Epidemiology* 2004; **160**: 589–597.
22. **Case C, *et al.*** Examining DNA fingerprinting as an epidemiology tool in the tuberculosis program in the Northwest Territories, Canada. *International Journal of Circumpolar Health* 2013; **72**: 20067.
23. **Anger HA, *et al.*** Active case finding and prevention of tuberculosis among a cohort of contacts exposed to infectious tuberculosis cases in New York City. *Clinical Infectious Diseases* 2012; **54**: 1287–1295.
24. **Clark CM, *et al.*** Universal genotyping in tuberculosis control program, New York City, 2001–2003. *Emerging Infectious Diseases* 2006; **12**: 719–724.
25. **New York City Public Health Code.** 2000. Article 13, Section 13·05.
26. **Groenen PM, *et al.*** Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Molecular Microbiology* 1993; **10**: 1057–1065.
27. **Kamerbeek J, *et al.*** Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *Journal of Clinical Microbiology* 1997; **35**: 907–914.
28. **Kremer K, *et al.*** Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. *Journal of Clinical Microbiology* 1999; **37**: 2607–2618.
29. **van Embden JD, *et al.*** Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *Journal of Clinical Microbiology* 1993; **31**: 406–409.
30. **Thierry D, *et al.*** IS6110, an IS-like element of Mycobacterium tuberculosis complex. *Nucleic Acids Research* 1990; **18**: 188.
31. **New York City Department of Health and Mental Hygiene.** Bureau of Tuberculosis Control Annual


https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press

Tuberculosis Summary, 2012 (http://www.nyc.gov/html/doh/downloads/pdf/tb/tb2012.pdf).
32. **World Health Organization.** Global Tuberculosis Report 2013 (http://apps.who.int/iris/bitstream/10665/91355/1/9789241564656_eng.pdf).
33. **San Francisco Department of Public Health.** San Francisco County Annual Report on Tuberculosis, 2008 (http://sfcdcp.org/document.html?id=968).
34. **Braden CR, Crawford JT, Schable BA.** Quality assessment of *Mycobacterium tuberculosis* genotyping in a large laboratory network. *Emerging Infectious Diseases* 2002; **8**: 1210–1215.
35. **Warren RM, *et al.*** Patients with active tuberculosis often have different strains in the same sputum specimen. *American Journal of Respiratory and Critical Care Medicine* 2004; **169**: 610–614.

https://doi.org/10.1017/S0950268816002399 Published online by Cambridge University Press