HHS Public Access

Author manuscript

Tuberculosis (Edinb). Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:

Tuberculosis (Edinb). 2017 September ; 106: 9–15. doi:10.1016/j.tube.2017.05.003.

Evaluation of 24-locus MIRU-VNTR genotyping in *Mycobacterium tuberculosis* cluster investigations in four jurisdictions in the United States, 2006–2010

Larry D. Teeter[a,1], **J. Steven Kammerer**[b], **Smita Ghosh**[b], **Duc T.M. Nguyen**[a], **Padmaja Vempaty**[b], **Jane Tapia**[c], **Roque Miramontes**[b], **Wendy A. Cronin**[d], and **Edward A. Graviss**[a,*] **on behalf of the Tuberculosis Epidemiologic Studies Consortium**[^]

[a]Houston Methodist Hospital Institute, 6670 Bertner Ave, Houston, TX 77030, U.S.A

[b]Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, GA 30333, Georgia, U.S.A

[c]Emory University School of Medicine, 1648 Pierce Dr NE, Atlanta, GA 30307, U.S.A

[d]Maryland Department of Health and Mental Hygiene, 201 W Preston St, Baltimore, MD 21201, U.S.A

Abstract

The U.S. Centers for Disease Control and Prevention (CDC) uses a combination of spacer oligonucleotide typing (spoligotyping) and mycobacterial interspersed repetitive units-variable number of tandem repeats (MIRU-VNTR) analyses as part of the National TB Genotyping Service (NTGS). The NTGS expansion from 12-locus MIRU-VNTR (MIRU12) to 24-locus MIRU-VNTR (MIRU24) in 2009 enhanced the ability to discriminate *Mycobacterium tuberculosis* strains. In the current study, we investigated the MIRU24 concordance among epidemiologic-linked tuberculosis (TB) patients in four U.S. health jurisdictions. We also evaluated the programmatic benefits of combining MIRU24 and spoligotyping with epidemiologic evidence in identifying potential recent TB transmission. We examined 342 TB patients in 42 spoligotype/MIRU12 (PCRType) clusters (equivalent to 46 spoligotype/MIRU24 [GENType] clusters) to identify epidemiologic links among cases. GENType clusters, when compared to PCRType clusters, had 12 times higher odds of epidemiologic links being identified if patients were younger than 25 years and 3 times higher odds if patients resided in the same zip code, or had HIV infection. Sixty (18%) fewer PCRType-clustered patients would need investigations if clusters are defined using GENType instead of

[*]Corresponding author: Edward A. Graviss, PhD, MPH, FIDSA, Department of Pathology and Genomic Medicine, Houston Methodist Research Institute, Mail Station: R6-414, 6670 Bertner, Houston, TX 77030, 713-441-4323/713-441-3599 (fax); eagraviss@houstonmethodist.org.

[1]Present affiliation: Forensic Research & Analysis, 425 NW 10th Avenue, Ste 306, Portland, OR 97209, U.S.A.

[^]Members of the Tuberculosis Epidemiologic Studies Consortium involved in this research are provided in the Acknowledgments.

Competing interests

None declared.

Ethical approval

CDC's Institutional Review Board (IRB) conducted an expedited review and approved this project based on minimal risk to study subjects. Additionally, IRBs at all four study sites approved the protocol.

Disclaimers

The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the U.S. Department of Health and Human Services.

PCRType. An important advantage of defining clusters by MIRU24 is resource savings related to the reduced number of clustered cases needing investigation.

Keywords

tuberculosis; genotype; cluster investigation; MIRU-VNTR; spoligotype; surveillance

1. Introduction

Tuberculosis (TB) control is a fundamental aspect of the U.S. public health system, whereby cases with TB disease and their contacts at risk are systematically identified, investigated, and treated. TB patients who are proximal to one another are considered clustered when their *Mycobacterium tuberculosis* isolates have indistinguishable genotypes. TB cluster investigations play an important role in tracking transmission of specific *Mycobacterium tuberculosis* (*Mtb*) strains in a population [1]. Among the tools used by investigators, polymerase chain reaction (PCR)-based methods such as spoligotyping (spacer oligonucleotide typing) and MIRU-VNTR (mycobacterial interspersed repetitive units-variable number of tandem repeats) are useful for identification of clusters of *Mtb* strains (through molecular characterization “fingerprints” or genotypes) due to their relatively fast laboratory turnaround time, suitability for discriminatory analyses, and their reproducibility [2–5]. These *Mtb* characterization methodologies have become the routine genotyping methods used by the U.S. Centers for Disease Control and Prevention (CDC) since 2004 [6]. However, the limited discriminatory power of the original 12-locus MIRU-VNTR (MIRU12) sometimes identified clusters that included cases among patients who may not have been connected by recent transmission. Therefore, in 2009, CDC’s National TB Genotyping Service (NTGS) began using 24-locus MIRU-VNTR (MIRU24) to improve the discriminatory power of TB molecular surveillance in the United States (U.S.) [7–9].

To better understand the practical utility, programmatic application, and added value of increased discriminatory power associated with the additional 12 loci in genotyping, TB transmission must be verified by epidemiologic investigations [8, 10–14]. With this background, we investigated the concordance of MIRU24 among epidemiologic-linked patients in four U.S. public health jurisdictions. We also attempted to determine the programmatic benefits of combining MIRU24 and spoligotyping.

2. Materials and Methods

2.1. Study population

The study population included *Mtb* culture-positive patients from four sites: Georgia (GA), Maryland (MD), Massachusetts (MA) and Texas (TX) who were reported to the CDC between January 2006 and October 2010 and whose *Mtb* isolates were genotyped by NTGS. All participating sites except Texas evaluated TB patients for *Mtb* clustering in each county throughout the state. In Texas, only TB patients counted in the jurisdiction of the City of Houston (HOU) were evaluated. The study jurisdictions were part of the Tuberculosis Epidemiologic Studies Consortium (TBESC), a consortium of U.S. sites funded by the CDC

to conduct TB epidemiologic research [15]. TB genotypes from all patients' isolates have been defined by a unique combination of spoligotype and 12-locus MIRU-VNTR results (MIRU12) [16], with each combination assigned a PCRType cluster designation. For patients in selected PCRType-based study clusters with *Mtb* isolates characterized prior to the 2009 transition to MIRU24, retrospective MIRU24 typing was performed. The study sample selection has been described in detail elsewhere [17]. Briefly, PCRType clusters were defined as two or more TB patients with the same PCRType in a given public health jurisdiction during the study period. As mentioned in the parent study, we focused on PCRTypes having three or more patients. Therefore, only PCRType clusters consisting of at least three TB patients residing in the same given public health jurisdiction, whose TB status were reported between January 1, 2006 and the time of cluster evaluation, were eligible for the sample's random selection [17]. Clustering by GENType was defined as two or more patients with indistinguishable spoligotype and MIRU24 pattern in a given county during the study timeframe. A PCRType cluster was defined as having a high proportion of patients clustered by GENType when ≥75% patients in the PCRType cluster were also in GENType clusters. Patients having single locus variant (SLV) were part of the present analysis. Patients having either mixed or missing results for any of the loci (MML) were excluded from the present analysis.

2.3. Epidemiologic link

An epidemiologic link was defined for two TB patients in the same cluster when they likely shared air space (i.e., same place at the same time) while one or both had active TB disease [18–19]. The epidemiologic link was classified as a definite link if the two patients were in the same place at the exact same time, a probable link if the patients were in the same place within the same week, or a possible link if the patients were in the same place either possibly at the same time or if exact time was unknown. A linked patient-pair was defined as a pair of patients who had an identified epidemiologic link with each other. Epidemiologic links were systematically investigated by site study staff for all clustered cases.

2.4. Ethical considerations

The current study was approved by the Institutional Review Boards of CDC, Emory University, Georgia's State Department Human Resources, Maryland's Department of Health and Mental Hygiene, Johns Hopkins University School of Medicine, Massachusetts's Department Public Health, Texas's Department of State Health Services, and the Houston Methodist Research Institute.

2.5. Statistical analysis

Demographic, behavioral, and clinical patient characteristics and genotype clusters were described as frequencies and proportions. Logistic regression analyses with cluster robust option (taking into account the non-independence by cluster) and exact logistic regression modeling were used to identify the characteristics of PCRType and GENType clusters associated with epidemiologic links (i.e., evidence for recent transmission). Variables having p-value < 0.2 in univariate analysis were investigated further in multivariate models. The variable selection process for the multivariate models was performed using the Bayesian model averaging (BMA) method [20–21]. We calculated the unadjusted and adjusted odds

ratios (OR) with associated 95% confidence intervals (95% CI) of finding epidemiologic links for PCRType and GENType clusters separately. The positive predictive value (PPV) of PCRType and GENType was calculated as the proportion of patients that were PCRType and GENType-clustered who also had an epidemiologic link. Differences between the PPVs of PCRType and GENType were compared using the Pearson's chi-square test. We compared the frequency and proportion of epidemiologic-linked patients in GENType clusters with the frequency and proportion of epidemiologic-linked patients in PCRType clusters. All statistical analyses were performed using Stata/MP 14.2 (StataCorp LP, College Station, TX, U.S.) and R version 3.3.1 (R Core Team, Vienna, Austria). A p-value of <0.05 was considered statistically significant.

3. Results

From 44 PCRType clusters randomly selected in our previous study [17], fifty-nine patients with incomplete MIRU24 results (including 13 patients for the entire clusters 3 and 31 from MD and GA respectively) were excluded from the analysis. In total, 42 clusters consisting of 342 patients having complete MIRU24 results were included in the current study (Appendix A). The number of patients in clusters from the GA, MD, MA, and HOU study sites were 73 (21%), 53 (16%), 71 (21%), and 145 (42%), respectively (Figure 1). Patients' median age at TB diagnosis was 47 years (interquartile range: 21–57); 73% were males, 70% were U.S.-born and 55% were black (Table 1). Eighteen of the 42 clusters (43%) had only U.S.-born patients (data not shown). In an effort to determine the extent of cluster homogeneity by birth country, Appendix B shows the birth country details for patients in the 24 clusters having at least one patient with a birth country other than the U.S. Eight out of 42 clusters (18%) contained only foreign-born patients. Among the 342 study patients, 150 (43.9%) were found to have at least one epidemiologic link to another patient in the cluster to form 156 linked patient-pairs. There were 33 patients that had more than one epidemiologic link. The strength of the linkages between each of 156 linked patient-pairs was most often considered definite (97/156, 62%), and less often probable (22/156, 14%) or possible (37/156, 24%) (data not shown).

3.1. 24-locus MIRU-VNTR genotyping

Although all 401 patients' isolates were MIRU24-genotyped, 342 (85%) had complete, unambiguous MIRU24 information at all 24 loci. Among 342 patients, 282 were in 46 GENType clusters (>2 cases) and 60 had unique GENType. From the four study sites, a total of 130 different GENTypes were identified (Table 2). Of note, 10/342 patients were identified as having a MIRU24 difference at one locus (SLV) (data not shown).

3.2. Characteristics associated with epidemiologic links

Among 42 PCRType clusters, 28 (66.7%) had patients with epidemiologic links. By comparison, 30/46 (65.2%) GENType clusters had patients with epidemiologic links. Notably, PCRType clusters with a high proportion of patients clustered by GENType had higher odds of epidemiologic links identified than PCRType clusters with low proportion of patients clustered by GENType (25/28, 89.3% versus 5/14, 35.7%); however, this difference was not statistical significant [adjusted OR (aOR) =6.15, 95% CI: 0.66, 242.55] (Table 3).

Patients in the 30 GENType clusters with epidemiologic links had greater odds of living in the same zip code as another patient in the cluster (aOR=22.71, 95% CI: 1.88, 1383.05) and lower odds of having extrapulmonary TB (aOR=0.04, 95% CI:0.0004, 0.66) or East Asian lineage (aOR=0.05, 95% CI: 0.001, 0.95) (Table 3).

3.3. MIRU24 concordance between epidemiologic-linked patients

The highest average proportion of epidemiologic-linked patients (83/143, 58%) was in PCRType clusters in which all isolates were also clustered by GENType (cluster numbers 01, 02, 04, 08, 09, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 33, 34, 37, 40, 42, 43). Conversely, the lowest average proportion of epidemiologic-linked patients (0/14, 0%) was in PCRType clusters in which none of the case isolates were in GENType clusters (cluster numbers 6, 21, 22, 32). The average proportion of epidemiologic-linked patients in PCRType clusters in which some of the cases' isolates were in GENType clusters was 35% (65/185). In this group, seven clusters (cluster numbers 05, 25, 26, 27, 35, 38, and 41) had no epidemiologic-linked patients identified (Table 3). The proportion of GENType-clustered patients that had epidemiologic links (GENType PPV) was 47.2% (133/282), higher than the proportion of PCRType-clustered patients that had epidemiologic links [PCRType PPV, 43.9% (150/342)] (Table 2) with a non-significant Pearson's p-value = 0.615 (data not shown). The proportion of patients that were epidemiologic-linked among clusters with East Asian strains was significantly lower than for non-East Asian strains [30% vs. 51% for PCRType clusters (p = 0.022), and 32% vs. 54% for GENType clusters (p = 0.035) (data not shown)].

Of 156 linked patient-pairs, 145 (93%) pairs had the same PCRTypes and 135 (87%) pairs also had the same GENTypes. There were 10 patient-pairs which had a MIRU24 difference at one locus (i.e. single locus variant or SLV) (data not shown). Two epidemiologic-linked pairs with "definite" linkage strength had concordant PCRTypes (PCR00002), but were discordant at three or more MIRU24 loci. We identified spoligotype or MIRU24 copy number discordance for 20 (12.8%) of the 156 epidemiologic-linked pairs, of which four were associated with household transmission settings (data not shown).

4. Discussion

This study assessed the impact of using MIRU24 versus MIRU12 by systematically investigating clusters for epidemiologic relationships among patients (i.e., evidence of transmission among patients within a genotype cluster). MIRU24 analysis further differentiated 342 patients in 42 PCRType clusters into 46 GENType clusters comprised of 282 patients, plus 60 patients with unique GENTypes. By routinely using spoligotyping and MIRU24 (GENType) as the primary cluster identification method, follow-up cluster investigations through chart review, extensive interview and contact investigation of 60/342 (18%) patients in PCRType clusters could have been avoided and as a result, TB program's resources for these interventions could have been saved. Additionally, using spoligotyping and MIRU24 (GENType) could provide a more accurate determination of the transmission risk among TB patients clustered by PCRType. This overall finding supports the practice of defining genotype clusters by GENType. While current practice has sometimes been to

consider including patients with a SLV genotype that already has identified links (e.g., through a contact investigation) in a cluster, re-reading or repeating MIRU24 when the results have missing or ambiguous values should be considered in order to definitively rule out a patient from a genotypic cluster.

In this study, 15% of *Mtb* isolates (n=59) had mixed or missing information for one or more of the 12 additional MIRU24 loci. Fortunately, the proportion of patients genotyped by NTGS with missing MIRU24 loci has decreased substantially due to improvement in laboratory methods (< 5% in 2013) since the time of this study (CDC, unpublished data).

More than half of the PCRType clusters (22/42, 52.4%) had 100% of patients who were also GENType clustered. In other clusters, however, each patient had a unique MIRU24 pattern – demonstrating that patients in some PCRType-based clusters should not be considered clustered. In particular, the five East Asian (L2) PCR00002 clusters had significantly decreased odds of epidemiologic linkages using GENType (OR=0.05, 95% CI 0.001, 0.95) and were heterogeneous both by birth country and MIRU24; only 26% (10/38) of patients clustered by GENType, which was significantly less than patients in other clusters (272/304, 90%, p<0.001). These findings appear to be consistent with the insufficient discrimination power of MIRU24 in isolates from the Beijing family that were reported by other authors [22–23].

Our results suggest that defining TB clusters based upon PCRType overestimated clustering, especially for genotypes that are common in the United States. Clusters associated with certain PCRTypes (e.g., PCR00002, PCR00015, PCR00022 and PCR00041, which are each reported in >40 U.S. states and associated with more than 400 TB patients nationally during 2006–2010) were clearly discriminated by MIRU24 (Appendix A). On the other hand, our results also show that SLVs for certain MIRU-VNTR loci can result in underestimation of clustering by MIRU24. Our comparison of the MIRU24 patterns of epidemiologic-linked patients showed ~6% of linked pairs were SLVs (Appendix A). If an epidemiologic linkage between patients exists where the only MIRU24 difference is missing data at a particular locus for one patient and an identified value at the same locus for another patient, it remains equally likely that they are in the same chain of transmission even though their isolates will be assigned different GENTypes. Additionally, the decreased proportion of epidemiologic links in PCRType and GENType clusters containing all foreign-born patients was (62.5% and 40.0%, respectively) and the GENType discordance found in 20 of the 156 epidemiologic linked pairs, which also suggests that using PCRType might be overestimating epidemiologic links in comparison with using GENType.

A notable limitation to this study was exclusion of clinically defined, and therefore non-genotyped, TB patients. In addition, we cannot rule out the possibility of misclassification due to false epidemiologic linkage in patient-pairs. Bennett *et al.* reported a higher proportion of discordant genotypes among epidemiologic linkages with 29% of epidemiologic-linked patient-pairs having discordant IS*6110*-based restriction fragment length polymorphisms (RFLP) patterns and 31% of epidemiologic-linked patient-pairs consisting of household members who were discordant [11]. Although the PPV to identify epidemiologic-linked cases was higher among clustered patients with GENType (47.2%)

than PCRType (43.9%), this relatively small difference between these two methods suggested the discrimination power of MIRU24 is still needed to be further investigated. With increasing applications of whole-genome sequencing (WGS) and phylogenetic analyses for use by TB control programs, additional molecular resolution can help identify patients who are part of a recent chain of transmission [9]. In the United States, it is likely that WGS will be integrated into molecular surveillance and become the gold standard for strain identification and characterization of genetic relatedness among isolates [24].

Because MIRU24 further differentiates most MIRU12 clusters into smaller clusters, one might logically expect that GENType clusters would be associated with higher proportions of epidemiologic links identified. In general, as the proportion of cases clustered by GENType increased, the proportion of patients that were epidemiologic-linked within a cluster also increased. GENType-defined clusters with a high proportion of epidemiologic-linked patients provide strong evidence for recent transmission; prioritization of these clusters would improve public health action. We found that certain characteristics were associated with having epidemiologic links among patients clustered by GENType. Specifically, GENType clusters with epidemiologic links had greater odds of having patients residing in the same zip code as a risk factor and less likely to have had >20% of patients with only extrapulmonary TB or be infected by East-Asian lineage strain. These factors may be useful to guide local programs in determining which GENType clusters warrant field investigations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This project was funded by the Division of TB Elimination, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, through a task order announced and managed by the Tuberculosis Epidemiologic Studies Consortium (TBESC). The task order is designated as Task Order 26.

The authors would like to thank the following Tuberculosis Epidemiologic Studies Consortium collaborators and consultants:

Georgia Site: Henry Blumberg, MD; Michael Leonard, MD; Rose-Marie Sales, MD, MPH; Arathi Chakragiri MPH; Lily Singha

Maryland Site: Susan Dorman, MD; Heather Rutz, MCRP, MHS; Susanna Collins, MPH; Elizabeth Munk, RN, BSN

Massachusetts Site: Sue Etkind, RN, MS; John Bernardo, MD, Andrew Tibbs, MPH

Texas Site: Charles Wallace, PhD, MPH; Joan Taylor RN; Pandora Williams; Kenneth Jost, Jr, MT (ASCP)

CDC Consultants: Patrick Moonan, DrPH, MPH; Thomas Navin, MD;

References

1. Mathema B, Kurepina NE, Bifani PJ, Kreiswirth BN. Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev. 2006; 19:658–85. https://dx.doi.org/10.1128/CMR.00061-05. [PubMed: 17041139]

2. Crawford JT. Genotyping in contact investigations: a CDC perspective. Int J Tuberc Lung Dis. 2003 Dec; 7(12 Suppl 3):S453–7. https://dx.doi.org/10.1186%2Fs12879-016-1937-9. [PubMed: 14677837]

3. Groenen PM, Bunschoten AE, van Soolingen D, van Embden JD. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. Mol Microbiol. 1993 Dec; 10(5):1057–65. http://dx.doi.org/10.1111/j.1365-2958.1993.tb00976.x. [PubMed: 7934856]

4. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol. 1997 Apr; 35(4):907–14. [PubMed: 9157152]

5. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Locht C, van Soolingen D. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol. 2006 Dec; 44(12):4498–510. Epub 2006 Sep 27. https://dx.doi.org/10.1128/JCM.01392-06. [PubMed: 17005759]

6. National TB Controllers Association/CDC Advisory Group on Tuberculosis Genotyping. Guide to the Application of Genotyping to Tuberculosis Prevention and Control. Atlanta, GA: US Department of Health and Human Services, CDC; Jun. 2004 Available: http://www.cdc.gov/tb/programs/genotyping/images/TBGenotypingGuide_June2004.pdf [Accessed 01/05/2014]

7. Oelemann MC, Diel R, Vatin V, Haas W, Rüsch-Gerdes S, Locht C, Niemann S, Supply P. 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. J Clin Microbiol. 2007 Mar; 45(3):691–7. Epub 2006 Dec 27. https://dx.doi.org/10.1128/JCM.01393-06. [PubMed: 17192416]

8. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of Mycobacterium tuberculosis complex isolates. J Clin Microbiol. 2008 Aug; 46(8):2692–9. Epub 2008 Jun 11. https://dx.doi.org/10.1128/JCM.00540-08. [PubMed: 18550737]

9. Kato-Maeda M, Metcalfe JZ, Flores L. Genotyping of Mycobacterium tuberculosis: Application in Epidemiologic Studies. Future Microbiol. 2011 Feb; 6(2):203–16. https://dx.doi.org/10.2217/fmb.10.165. DOI: 10.2217/fmb.10.165 [PubMed: 21366420]

10. Moonan PK, Ghosh S, Oeltmann JE, Kammerer JS, Cowan LS, Navin TR. Using genotyping and geospatial scanning to estimate recent mycobacterium tuberculosis transmission, United States. Emerg Infect Dis. 2012 Mar; 18(3):458–65. https://dx.doi.org/10.3201/eid1803.111107. [PubMed: 22377473]

11. Bennett DE, Onorato IM, Ellis BA, Crawford JT, Schable B, Byers R, et al. DNA fingerprinting of Mycobacterium tuberculosis isolates from epidemiologic-linked case pairs. Emerg Infect Dis. 2002 Nov; 8(11):1224–9. https://dx.doi.org/10.3201/eid0811.020420. [PubMed: 12453346]

12. Teeter LD, Ha NP, Ma X, Wenger J, Cronin WA, Musser JM, Graviss EA. Evaluation of large genotypic Mycobacterium tuberculosis clusters: contributions from remote and recent transmission. Tuberculosis (Edinb). 2013 Dec; 93(Suppl):S38–46. https://dx.doi.org/10.1016/S1472-9792(13)70009-X. [PubMed: 24388648]

13. Klovdahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, et al. Networks and tuberculosis: an undetected community outbreak involving public places. Soc Sci Med. 2001 Mar; 52(5):681–94. http://dx.doi.org/10.1016/S0277-9536(00)00170-2. [PubMed: 11218173]

14. Centers for Disease Control and Prevention (CDC). [Accessed 01/05/2017] GENType: New Genotyping Terminology to Integrate 24-locus MIRU-VNTR. 2002. Available at: http://www.cdc.gov/tb/publications/factsheets/statistics/genotypingterminology.pdf

15. Katz D, Albalak R, Wing JS, Combs V. Tuberculosis Epidemiologic Studies Consortium. Setting the agenda: a new model for collaborative tuberculosis epidemiologic research. Tuberculosis (Edinb). 2007 Jan; 87(1):1–6. Epub 2006 Aug 8. https://dx.doi.org/10.1016/j.tube.2005.12.003. [PubMed: 16895763]

16. Ghosh S, Moonan PK, Cowan L, Grant J, Kammerer S, Navin TR. Tuberculosis genotyping information management system: enhancing tuberculosis surveillance in the United States. Infect Genet Evol. 2012 Jun; 12(4):782–8. https://dx.doi.org/10.1016/j.meegid.2011.10.013. [PubMed: 22044522]

17. Teeter LD, Vempaty P, Nguyen DT, Tapia J, Sharnprapai S, Ghosh S, Kammerer JS, Miramontes R, Cronin WA, Graviss EA8. Tuberculosis Epidemiologic Studies Consortium. Validation of genotype cluster investigations for Mycobacterium tuberculosis: application results for 44 clusters from four heterogeneous United States jurisdictions. BMC Infect Dis. 2016 Oct 21.16(1):594. https://doi.org/10.1186/s12879-016-1937-9. [PubMed: 27769182]

18. Perri BR, Proops D, Moonan PK, Munsiff SS, Kreiswirth BN, Goranson C, Ahuja SD. Mycobacterium tuberculosis Cluster with Developing Drug Resistance, New York, New York, USA, 2003–2009. Emerging Infectious Diseases. 2011; 17(3):372–378. https://doi.org/10.3201/eid1703.101002. [PubMed: 21392426]

19. Thombley, ML., Stier, DD. Menu of Suggested Provisions for State Tuberculosis Prevention Control Laws. Centers for Disease Control and Prevention, National Tuberculosis Controllers Association, Association of State and Territorial Health Officials and National Association of County Health Officials Workshop; Feb 4–5, 2010; Atlanta, GA, USA: CDC; Available at www.cdc.gov/tb/programs/laws/menu/tblawmenu.pdf

20. Wasserman L. Bayesian Model Selection and Model Averaging. Journal of mathematical psychology. 2000; 44:92–107. https://doi.org/10.1006/jmps.1999.1278. [PubMed: 10733859]

21. Raftery Adrian, HJ., Chris, Volinsky, Ian, Painter, Yee, Yeung Ka. [Accessed on Mar 12, 2017] BMA: Bayesian Model Averaging. 2015. Available at http://CRAN.R-project.org/package=BMA

22. Devi KR, Bhutia R, Bhowmick S, Mukherjee K, Mahanta J, Narain K. Genetic Diversity of Mycobacterium tuberculosis Isolates from Assam, India: Dominance of Beijing Family and Discovery of Two New Clades Related to CAS1_Delhi and EAI Family Based on Spoligotyping and MIRU-VNTR Typing. PLoS One. 2015 Dec 23.10(12):e0145860. https://doi.org/10.1371/journal.pone.0145860. [PubMed: 26701129]

23. Zheng C, Zhao Y, Zhu G, Li S, Sun H1, Feng Q, Luo M, Wu F, Li X, Hill V, Rastogi N, Sun Q. Suitability of IS6110-RFLP and MIRU-VNTR for differentiating spoligotyped drug-resistant mycobacterium tuberculosis clinical isolates from Sichuan in China. Biomed Res Int. 2014; 2014:763204. https://doi.org/10.1155/2014/763204. [PubMed: 24724099]

24. MacLean D, Jones JD, Studholme DJ. Application of ‘next-generation’ sequencing technologies to microbial genetics. Nat Rev Microbiol. 2009 Apr; 7(4):287–96. https://dx.doi.org/10.1038/nrmicro2122. [PubMed: 19287448]

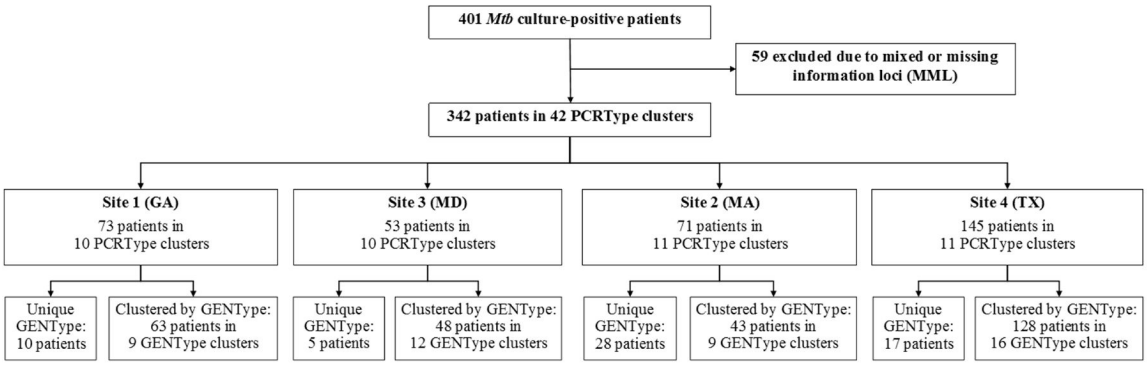

Figure 1.
Patient enrollment

Table 1

Select characteristics of study population

Characteristics	N=342
Age (years), median (IQR)	47 (21, 57)
Gender, n (%)	
Female	93 (27.2)
Male	249 (72.8)
Race/Ethnicity, n (%)	
White	52 (15.2)
Black	187 (54.7)
Hispanic	62 (18.1)
Asian	39 (11.4)
Mixed	2 (0.6)
Foreign-born, n (%)	
No	240 (70.2)
Yes	102 (29.8)
PCRType clusters with all patients being foreign-born	8/42 (19.1%)
GENType clusters with all patients being foreign-born	11/46 (23.9%)
Lineage	
Indo-Oceanic (L1)	13 (3.8%)
East Asian (L2)	115 (33.6%)
East-African-Indian (L3)	3 (0.9%)
Euro-American (L4)	211 (61.7%)
HIV infection	
No	258 (83.0)
Yes	58 (17.0)
TB type, n (%)	
Extra-pulmonary TB	8 (2.3)
Pulmonary TB	334 (97.7)
Vital status at diagnosis, n (%)	
Alive	334 (97.7)
Dead	8 (2.3)

Table 2

Comparison of GENType clustering and Epidemiologic Links for 42 PCRType Clusters

PCRType	State	PCRType Cluster Number	Number of patients in PCRType cluster	PCRType epilinked patients	% epilinked	Number of patients in GENType clusters	GENType epilinked patients	% epilinked	% GENType/ PCRType matched
East Asian									
PCR03412	MA	08	6	5	83.3	6	5	83.3	100.0
PCR04846	TX	09	6	0	0.0	6	0	0.0	100.0
PCR01201	TX	12	26	11	42.3	26	11	42.3	100.0
PCR00224	TX	11	31	17	54.8	30	11	36.7	96.8
PCR00036	TX	41	8	0	0.0	7	0	0.0	87.5
PCR00002	MD	35	9	0	0.0	4	0	0.0	44.4
PCR00002	MA	05	10	0	0.0	4	0	0.0	40.0
PCR00002	MA	38	11	0	0.0	2	0	0.0	18.2
PCR00002	MA	21	4	0	0.0	0	0	0.0	0.0
PCR00002	GA	32	4	2	50.0	0	0	0.0	0.0
Sub-Total			115	35	30%	85	27	32%	73.9
Non-East Asian									
PCR01047	MD	01	14	11	78.6	14	11	78.6	100.0
PCR01674	MD	02	6	3	50.0	6	3	50.0	100.0
PCR02143	MD	04	3	2	66.7	3	2	66.7	100.0
PCR00578	GA	13	4	4	100.0	4	4	100.0	100.0
PCR06732	GA	14	5	4	80.0	5	4	80.0	100.0
PCR01872	GA	15	2	0	0.0	2	0	0.0	100.0
PCR00016	GA	16	18	10	55.6	18	10	55.6	100.0
PCR00044	MD	17	3	3	100.0	3	3	100.0	100.0
PCR00169	MD	18	4	0	0.0	4	0	0.0	100.0
PCR02651	MD	19	3	1	33.3	3	1	33.3	100.0
PCR03994	MD	20	3	3	100.0	3	3	100.0	100.0
PCR00849	MA	23	8	5	62.5	8	5	62.5	100.0

PCRType	State	PCRType Cluster Number	Number of patients in PCRType cluster	PCRType epilinked patients	% epilinked	Number of patients in GENType clusters	GENType epilinked patients	% epilinked	% GENType/ PCRType matched
PCR03405	MA	24	9	7	77.8	9	7	77.8	100.0
PCR04200	MD	33	4	3	75.0	4	3	75.0	100.0
PCR01017	MD	34	4	2	50.0	4	2	50.0	100.0
PCR01034	MA	37	3	0	0.0	3	0	0.0	100.0
PCR01873	TX	40	3	2	66.7	3	2	66.7	100.0
PCR03588	GA	42	6	5	83.3	6	5	83.3	100.0
PCR00017	GA	43	3	2	66.7	3	2	66.7	100.0
PCR00719	TX	10	15	7	46.7	14	6	42.9	93.3
PCR00051	TX	39	29	9	31.0	26	9	34.6	89.7
PCR02397	GA	29	9	7	77.8	8	7	87.5	88.9
PCR00293	GA	30	14	7	50.0	12	5	41.7	85.7
PCR00016	MA	36	10	10	100.0	8	8	100.0	80.0
PCR01046	MA	07	4	1	25.0	3	1	33.3	75.0
PCR00497	TX	26	7	0	0.0	5	0	0.0	71.4
PCR00017	TX	25	3	0	0.0	2	0	0.0	66.7
PCR00015	GA	44	8	3	37.5	5	0	0.0	62.5
PCR00041	TX	28	13	4	30.8	7	3	42.9	53.8
PCR04837	TX	27	4	0	0.0	2	0	0.0	50.0
PCR00724	MA	06	3	0	0.0	0	0	0.0	0.0
PCR00022	MA	22	3	0	0.0	0	0	0.0	0.0
Sub-total			227	115	51%	197	106	54%	86.8
Total			**342**	**150**	**43.9**	**282**	**133**	**47.2**	**82.5**

Based on data from Appendix A;

Epidemiologic-linked (epilinked): two TB patients in the same cluster when they likely shared air space (i.e. same place at the same time) while one or both subjects had active TB disease.

Table 3

Characteristics associated with epidemiologic links for PCRType- and GENType-based clusters in four jurisdictions in the United States, 2006–2010

	PCRType Clusters (N = 42)							GENType Clusters (N=46)						
	Not epilinked (n=14)		Epilinked (n=28)		Epilinked row %	Unadjusted OR[a] (95% CI)	Adjusted OR[b] (95% CI)	Not epilinked (n=16)		Epilinked (n=30)		Epilinked row %	Unadjusted OR[a] (95% CI)	Adjusted OR[b] (95% CI)
	n	%	*n*	%				*n*	%	*n*	%			
Study Site														
Texas (City of Houston)	5	35.7	6	21.4	54.6	0.49 (0.12, 2.06)	-	7	43.8	9	30.0	56.3	0.55 (0.14, 2.13)	-
Massachusetts	6	42.9	5	17.9	45.5	0.29 (0.07, 1.24)	-	4	25.0	5	16.7	55.6	0.60 (0.11, 3.18)	-
Maryland	2	14.3	8	28.6	80.0	2.40 (0.43, 13.50)	-	4	25.0	8	26.7	66.7	1.09 (0.24, 4.95)	-
Georgia	1	7.1	9	32.1	90.0	6.16 (0.68, 56.11)	-	1	6.3	8	26.7	88.9	5.46 (0.59, 50.26)	-
Cluster with ANY patient with given characteristic														
Age younger than 25 yrs	6	42.9	18	64.3	75.0	2.40 (0.64, 9.04)	3.00 (0.17, 202.50)	3	18.8	15	50.0	83.3	4.33 (1.00, 18.72)	13.81 (0.96, 986.23)
Zip code clustered by genotype	7	50.0	23	82.1	76.7	4.60 (1.09, 19.48) 9.17 (0.63, 550.41)	3	18.8	23	76.7	88.5		14.24 (3.03, 66.92)	22.71 (1.88, 1383.1)
Foreign-born	8	57.1	16	57.1	66.7	1.00 (0.27, 3.71)	1.70 (0.13, 91.87)	10	62.5	14	46.7	58.3	0.53 (0.14, 2.01)	0.36 (0.02, 6.47)
Asian	5	35.7	3	10.7	37.5	0.22 (0.04, 1.12)	-	5	31.3	2	6.7	28.6	0.16 (0.02, 15.18)	-
Black	9	64.3	24	85.7	72.7	3.33 (0.72, 15.55)	-	9	56.3	25	83.3	73.5	3.89 (1.00, 15.41)	-
Hispanic	4	28.6	11	39.3	73.3	1.62 (0.40, 6.58)	-	6	37.5	11	36.7	64.7	0.97 (0.26, 3.53)	-
White	6	42.9	11	39.3	64.7	0.86 (0.23, 3.22)	-	3	18.8	11	36.7	78.6	2.51 (0.63, 10.06)	-
HIV infection	6	42.9	18	64.3	75.0	2.40 (0.64, 9.04)	4.44 (0.05, 426.47)	5	31.3	19	63.3	79.2	3.80 (1.03, 13.99)	9.00 (0.88, 447.46)
Homeless[c]	8	57.1	18	64.3	69.2	1.35 (0.36, 5.09)	-	6	37.5	18	60.0	75.0	2.50 (0.72, 8.74)	-
Excess alcohol use[c]	7	50.0	21	75.0	75.0	3.00 (0.76, 11.79)	-	7	43.8	23	76.7	76.7	4.23 (1.04, 17.16)	-
Non-injection drug use[c]	5	35.7	17	60.7	77.3	2.78 (0.72, 10.70)	-	7	43.8	19	63.3	73.1	2.22 (0.55, 9.00)	-
Injection drug use[c]	1	7.1	7	25.0	87.5	4.33 (0.46, 40.43)	-	0	0.0	6	20.0	100.0	na	-
Cluster with ALL patients with given characteristic														
US-born	6	42.9	12	42.9	66.7	1.00 (0.27, 3.71)	-	6	37.5	16	53.3	72.7	1.91 (0.50, 7.30)	-
Foreign-born	3	21.4	5	17.9	62.5	0.80 (0.16, 4.03)	-	6	37.5	4	13.3	40.0	0.26 (0.07, 1.02)	-
Male	2	14.3	6	21.4	75.0	1.64 (0.28, 9.60)	-	8	50.0	8	26.7	50.0	0.36 (0.11, 1.26)	-

	PCRType Clusters (N = 42)							GENType Clusters (N=46)						
	Not epilinked (n=14)		Epilinked (n=28)		Epilinked row %	Unadjusted OR[a] (95% CI)	Adjusted OR[b] (95% CI)	Not epilinked (n=16)		Epilinked (n=30)		Epilinked row %	Unadjusted OR[a] (95% CI)	Adjusted OR[b] (95% CI)
	n	%	*n*	%				*n*	%	*n*	%			
Sputum AFB smear+ (>=50% patients)	8	57.1	21	75.0	72.4	2.25 (0.57, 8.92)	-	11	68.8	24	80.0	68.6	1.82 (0.54, 6.19)	-
Cavitary CXR (>=50% patients)	3	21.4	8	28.6	72.7	1.47 (0.32, 6.81)	-	8	50.0	9	30.0	52.9	0.43 (0.12, 1.54)	-
Extrapulmonary site only (>20% patients)	6	42.9	4	14.3	40.0	0.22 (0.05, 1.01)	0.09 (0.002, 1.26)	5	31.3	3	10.0	37.5	0.25 (0.05, 1.22)	0.04 (0.0004, 0.66)
Cluster size<4	5	35.7	6	21.4	54.6	0.49 (0.12, 2.06)	-	11	68.8	10	33.3	47.6	0.23 (0.06, 0.81)	-
Common genotype in U.S.[d]	5	35.7	3	10.7	37.5	0.22 (0.04, 1.11)	-	5	31.3	2	6.7	28.6	0.16 (0.02, 1.12)	-
East Asian (L2) lineage	6	42.9	4	14.3	40.0	0.22 (0.05, 1.01)	0.08 (0.001, 1.32)	7	43.8	5	16.7	41.7	0.26 (0.04, 1.522)	0.05 (0.001, 0.95)
Euro-American (L4) lineage	8	57.1	22	78.6	73.3	2.75 (0.67, 11.24)	-	9	56.3	23	76.7	71.9	2.56 (0.53, 12.28)	
Clustered by GENType (>=75% patients)	5	35.7	25	89.3	83.3	15.00 (2.91, 77.42)	6.15 (0.66, 242.55)	na	na	na	na	na	na	na

Notes:

Cluster of *M. tuberculosis* isolates defined by common results from spoligotyping and 12-locus MIRU-VNTR (PCRType) or spoligotyping and 24-locus MIRU-VNTR (GENType).

[a]Unadjusted odds ratio. Odds ratio > 1.0 indicates clusters with epidemiologic links identified had greater odds of having the given characteristic than clusters without epidemiologic links; the referent group for comparisons were all cases that did not have the particular characteristic.

[b]Adjusted in the multiple logistic regression model.

[c]Within 12 months prior to TB diagnosis

[d]PCRTypes associated with > 400 TB patients in the United States during 2006–2010 and reported in >40 states (PCR00002, PCR00015, PCR00022, PCR00041) AFB: Acid-fast bacilli

CXR: chest radiograph

Epilinked: epidemiologic-linked