



Published in final edited form as:

Int J Tuberc Lung Dis. 2025 March 31; 29(4): 193–195. doi:10.5588/ijtld.24.0464.

Use of different genotyping methods to estimate TB transmission in the United States, 2020–2021

K.R. Schildknecht^{1,2}, L.S. Cowen¹, J.E. Posey¹, S. Talarico¹, M.B. Haddad¹, J.M. Wortham¹, J.S. Kammerer¹

¹Division of Tuberculosis Elimination, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA;

²Epidemic Intelligence Service, Centers for Disease Control and Prevention, Atlanta, GA, USA.

Keywords

epidemiology; whole-genome sequencing; disease transmission; infectious

Dear Editor,

Most TB in the United States arises from reactivation of past *Mycobacterium tuberculosis* infection.^{1–4} Identifying and treating recently infected persons are priority prevention and control activities because recently acquired *M. tuberculosis* infection can rapidly progress to TB disease.⁵ Therefore, health departments in the United States routinely use local TB genotyping results to gauge their success in suppressing recent transmission (RT). According to PCR-based genotyping methods used for over a decade, approximately 12% of all TB appears attributable to RT.^{4,6,7} In 2018, whole-genome sequencing (WGS) became routine for each culture-confirmed TB case, and in 2022, PCR-based genotyping (i.e., GENType designations) was discontinued. These new WGS methods produce a whole-genome multilocus sequence typing (wgMLSType) designation. wgMLSType is assigned by comparing the sequences of 2,672 genetic loci, chosen based on presence in ≥95% of genomes in a test dataset of 5,000 genomes, and assigning the same wgMLSType to cases that have matching patterns at ≥99.7% of the loci, a cutoff that ensures clustering of isolates that are within seven single nucleotide polymorphisms (SNPs). However, wgMLSType designation does not provide the number of pairwise SNP differences between isolates, and cases with a matching wgMLSType might not necessarily be closely related enough genetically to be consistent with RT. As recommended in a recent publication regarding best practices for WGS,⁸ whole-genome SNP comparison and phylogenetic trees, which are more discriminatory than wgMLSType, are produced by CDC for a select subset of wgMLSType clusters. These are used along with timing, geography, and other epidemiologic characteristics to further assess the likelihood that cases are related by RT for cluster investigation.

Correspondence to: Kimberly R. Schildknecht, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Mailstop H24-4, Atlanta, GA 30329, USA. trg8@cdc.gov.

Conflicts of interest: none declared.

Using TB cases reported to the National Tuberculosis Surveillance System⁴ from 2018–2021 (i.e., the 4 years when both GENType and wgMLSType designations were assigned), we produced new sequencing-based RT estimates for the 2020–2021 cases. We applied the previously validated plausible source case method; that is, if ≥ 1 plausible source case with a matching wgMLSType designation was identified during the previous 2 years within 10 miles, the TB case was attributed to RT.^{6,7,9} We compared the proportion of TB cases attributed to RT using the older GENType and newer wgMLSType designations, both in the overall population and stratified by origin of birth. We identified counties with ≥ 10 TB cases during 2020–2021 that had $\geq 10\%$ absolute difference in RT using the two genotyping methods. We described those counties by their rural-urban classification (U.S. Department of Agriculture),¹⁰ poverty level (U.S. Census American Community Survey),¹¹ and annual TB count per 100,000 persons (National Tuberculosis Surveillance System).⁴

We found that with the PCR-based GENType designation, 12.3% of TB cases nationally were attributed to RT; the corresponding proportion was 11.8% using the sequencing-based wgMLSType designation. Among U.S.-born persons with TB, the proportion of TB cases attributed to RT increased from 24.8% to 28.2%. Conversely, RT estimates decreased from 7.7% to 5.9% among non-U.S.-born persons. In total, 14 U.S. counties reported ≥ 10 TB cases during 2020–2021 and had $\geq 10\%$ absolute difference in the proportion of cases attributed to RT. The Table shows that estimates of RT increased $\geq 10\%$ in 10 counties and decreased $\geq 10\%$ in 4 counties. The 10 counties with new higher RT estimates varied widely in population size, composition, and annual TB counts, but nearly all had substantial poverty. In these 10 counties, 37 persons whose TB was previously not attributed to RT were reclassified as having TB attributed to RT because wgMLSType grouped closely related isolates with different PCR-based GENTypes. All 37 were either U.S.-born ($n = 22$) or Mexico-born ($n = 15$). The four counties with new lower RT estimates were all urban with larger populations and generally lower poverty levels. Among the 36 total persons reclassified as not having TB attributed to RT, 29 (80.6%) were born in China or the Philippines. However, small absolute numbers drove the $\geq 10\%$ of case reclassifications in these 14 counties.

In conclusion, the overall proportion of TB cases attributed to RT was similar with either genotyping method, but the sequencing-based wgMLSType designation produced a somewhat higher RT estimate among U.S.-born persons and a somewhat lower estimate among non-U.S.-born persons. Poverty in counties with increased RT estimates suggests fewer resources and greater challenges for TB prevention and control at the local level. This poverty context also emphasizes the importance of the direct connections with the communities most affected by RT and including those voices in developing public health strategies to decrease RT.⁸ Transition to WGS-based genotyping will require changes in methods to estimate RT to ensure accuracy; accurate RT estimates will facilitate assessments of efforts to suppress RT. Further work is necessary to validate the findings of this analysis in the context of local epidemiology and to optimize RT estimates for WGS-based genotyping. In the meantime, these findings of similar overall RT estimates in the context of differences by birth origin and specific locations and settings might be helpful context for public health decision-making.

Acknowledgments

We thank the clinicians, laboratory personnel, and public health staff who report cases and contribute data to the U.S. National TB Surveillance System. The views and opinions expressed in this article are those of the authors and do not necessarily represent an official position of the Centers for Disease Control and Prevention. All authors are employees of the U.S. government.

References

1. Cain KP, et al. Tuberculosis among foreign-born persons in the United States. *JAMA* 2008;300(4):405–412. [PubMed: 18647983]
2. Ricks PM, et al. Estimating the burden of tuberculosis among foreign-born persons acquired prior to entering the U.S., 2005–2009. *PLoS ONE* 2011;6(11):e27405. [PubMed: 22140439]
3. Tsang CA, et al. Tuberculosis among foreign-born persons diagnosed ≥10 years after arrival in the United States, 2010–2015. *MMWR Morb Mortal Wkly Rep.* 2017;66(11):295–298. [PubMed: 28333913]
4. CDC. Reported tuberculosis in the United States, 2021. U.S. Department of Health and Human Services, CDC; Atlanta: 2022.
5. Taylor Z, Nolan CM, Blumberg HM, American Thoracic Society, Centers for Disease Control and Prevention, Infectious Diseases Society of America. Controlling tuberculosis in the United States. Recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. *MMWR Recomm Rep* 2005;54(RR-12):1–81.
6. France AM, et al. A field-validated approach using surveillance and genotyping data to estimate tuberculosis attributable to recent transmission in the United States. *Am J Epidemiol* 2015;182(9):799–807. [PubMed: 26464470]
7. Yuen CM, et al. Recent transmission of tuberculosis — United States, 2011–2014. *PloS ONE* 2016;11:e0153728. [PubMed: 27082644]
8. Denholm JT, et al. Considering best practice standards for routine whole-genome sequencing for TB care and control. *IJTL Open* 2024;1(10):431–436. [PubMed: 39398433]
9. Centers for Disease Control and Prevention. Whole-genome sequencing. Atlanta, GA, USA: CDC, 2024.
10. U.S. Department of Agriculture. Rural-Urban Continuum Codes. Washington DC, USA: U.S. Department of Agriculture, 2025.
11. U.S. Census Bureau. American Community Survey. 2017–2021, Table ACSST5Y2021.S1701 Washington DC, USA: U.S. Census Bureau, 2021.

Table.

Counties with $\geq 10\%$ change in estimates of recent transmission after switching from PCR-based genotyping methods to the wgMLSType designation based on whole-genome sequencing, United States, 2020–2021.

County or county equivalent	Rural-urban continuum classification for the county and population size ^{*†}	Proportion of population under 200% of federal poverty level [‡] (%)	Annual TB count per 100,000 persons [‡]	Absolute change in county's recent transmission (RT) estimate (%) [§]	Absolute number reclassified as RT	Absolute number reclassified as not RT
Ten U.S. counties with an increase in estimated RT after change to sequencing-based wgMLSType designation						
County A	Rural with a population of <2,500	70.5	126.9	+28.6% (from 42.9% to 71.4%)	5 U.S.-born persons	1 U.S.-born person
County B	Rural with a population of 2,500–19,999	55.1	107.3	+25.8% (from 70.4% to 96.2%)	7 U.S.-born persons	—
County C	Adjacent to a metropolitan area with a population of $\geq 20,000$	60.6	11.5	+27.3% (from 18.2% to 45.5%)	1 U.S.-born person 2 Non-U.S.-born persons	—
County D	Metropolitan area with a population of 250,000–1,000,000	48.0	8.3	+14.7% (from 35.3% to 50.0%)	2 U.S.-born persons 3 Non-U.S.-born persons	—
County E	Metropolitan area with a population of 250,000–1,000,000	54.4	4.0	+12.2% (from 21.7% to 33.9%)	2 U.S.-born persons 5 Non-U.S.-born persons	—
County F	Metropolitan area with a population of 250,000–1,000,000	34.4	1.6	+11.3% (from 8.7% to 20.0%)	1 U.S.-born person 1 Non-U.S.-born persons	—
County G	Metropolitan area with a population of 250,000–1,000,000	36.0	1.5	+11.1% (from 22.2% to 33.3%)	1 U.S.-born person	—
County H	Metropolitan area with a population of 250,000–1,000,000	31.4	5.5	+10.0% (from 15.0% to 25.0%)	4 Non-U.S.-born persons	—
County I	Metropolitan area with a population of ≥ 1 million	43.2	3.6	+13.3% (from 33.3% to 46.7%)	2 U.S.-born persons	—
County J	Metropolitan area with a population of ≥ 1 million	22.5	0.7	+10.0% (from 20.0% to 30.0%)	1 U.S.-born person	—
Four U.S. counties with a decrease in estimated RT after change to sequencing-based wgMLSType designation						
County K	Metropolitan area with a population of ≥ 1 million	32.5	2.4	–10.3% (from 13.8% to 3.5%)	—	2 U.S.-born persons 7 Non-U.S.-born persons
County L	Metropolitan area with a population of ≥ 1 million	20.5	8.1	–10.5% (from 14.3% to 3.8%)	2 U.S.-born persons	13 Non-U.S.-born persons
County M	Metropolitan area with a population of ≥ 1 million	14.7	6.6	–14.8% (from 16.0% to 1.3%)	—	12 Non-U.S.-born persons
County N	Metropolitan area with a population of ≥ 1 million	20.1	1.5	–16.7% (from 25.0% to 8.3%)	—	2 Non-U.S.-born persons

* Source: U.S. Department of Agriculture, <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/>

[†]Estimated county population size on July 1, 2020, and July 1, 2021, obtained from the American Community Survey, Table CO-EST2022-POP, available at <https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html>. Poverty proportions derived from American Community Survey, 2017–2021, Table ACSST5Y2021.S1701, available at <https://www.census.gov/topics/income-poverty/poverty/data/tables/acs.html>.

[‡]Annual TB count per 100,000 calculated by halving the total number of reported cases in the county during 2020–2021 from the National TB Surveillance System and dividing that by the mean of the two midyear population estimates from the American Community Survey.

[§]No rounding was performed on the minuend or subtrahend, which accounts for apparent rounding errors on the differences in Counties A, I, and M.

PCR = polymerase chain reaction; wgMLSType = whole-genome multilocus sequence typing; RT = recent transmission.