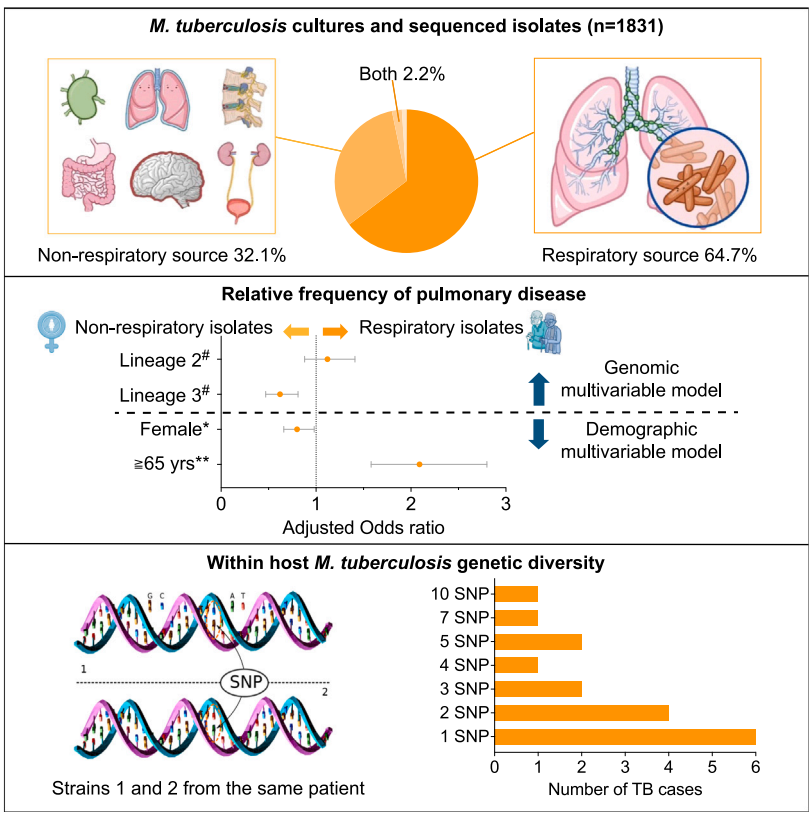
iScience

CellPress
OPEN ACCESS

Article

Genomic characteristics of prospectively sequenced *Mycobacterium tuberculosis* from respiratory and non-respiratory sources



Xiaomei Zhang, Connie Lam, Eby Sim, Elena Martinez, Taryn Crighton, Ben J. Marais, Vitali Sintchenko

xiaomei.zhang@sydney.edu.au

Highlights

More frequent *M. tuberculosis* isolation from respiratory specimens in older adults

Overrepresentation of lineage 2 *M. tuberculosis* strains in respiratory specimens

Within-host *M. tuberculosis* strain-specific genetic variability of up to 10 SNPs

Zhang et al., iScience *27*, 110327
July 19, 2024
https://doi.org/10.1016/j.isci.2024.110327

iScience

CellPress
OPEN ACCESS

Article

Genomic characteristics of prospectively sequenced *Mycobacterium tuberculosis* from respiratory and non-respiratory sources

Xiaomei Zhang,[1,2,3,7,*] Connie Lam,[2,3] Eby Sim,[2,3] Elena Martinez,[2,3,4] Taryn Crighton,[3,4] Ben J. Marais,[1,2,5,6] and Vitali Sintchenko[1,2,3,4,5,6]

SUMMARY

Understanding the differences between *Mycobacterium tuberculosis* strains isolated from respiratory and non-respiratory sources may inform clinical care and control strategies. We examined demographic and genomic characteristics of all culture-confirmed *M. tuberculosis* cultures isolated from respiratory and non-respiratory sources in New South Wales, Australia, from January 2017 to December 2021, using logistic regression models. *M. tuberculosis* strains from 1,831 patients were sequenced; 64.7% were from respiratory, 32.1% from non-respiratory, and 2.2% from both sources. Female patients had more frequent isolation from a non-respiratory source ($p = 0.03$), and older adults (≧65 years) from a respiratory source ($p < 0.0001$). Lineage 2 strains were relatively over-represented among respiratory isolates ($p = 0.01$). Among 39 cases with sequenced isolates from both sources, 43.6% had 1–10 single nucleotide polymorphism differences. The finding that older adults were more likely to have *M. tuberculosis* isolated from respiratory sources has relevance for TB control given the expected rise of TB among older adults.

INTRODUCTION

Tuberculosis (TB) is the leading infectious cause of death globally, with an estimated 1.3 million TB-related deaths in 2022 and major setbacks in global TB control efforts resulting from health system disruption caused by the COVID-19 pandemic.[1] *Mycobacterium tuberculosis* spreads via the aerosol route with pulmonary tuberculosis (PTB) responsible for most transmission events. However, disease may also affect other anatomical sites, referred to as extrapulmonary TB (EPTB).[1,2] The World Health Organization (WHO) reported an estimated 10.6 million new TB cases globally in 2022, with approximately 20% of all cases being EPTB.[1]

M. tuberculosis isolation from a respiratory source, including sputum, induced sputum, nasopharyngeal aspirates (also gastric aspirates or stool in children), and bronchoalveolar lavage or bronchial washings is indicative of PTB. *M. tuberculosis* isolation from a non-respiratory source, such as lymph node biopsies, pleural fluid, cerebrospinal fluid, and various other tissues reflect EPTB,[2] although PTB and EPTB may be present at the same time, in which case, it is programmatically classified as PTB—by convention. The demographic characteristics of patients with PTB and EPTB have been explored in multiple studies,[3–5] but few studies were able to reflect on genomic differences between strains causing PTB and EPTB disease,[6–8] and none have been able to do it in a comprehensive prospective fashion. The implementation of routine whole genome sequencing allows comprehensive genomic characterization of *M. tuberculosis* strains, including lineage/sub-lineage assignment, mixed strain population (simultaneous co-infection with more than one strain), drug resistance, and transmission cluster identification.[9,10] The incorporation of sequencing data into real-time TB case management and control efforts assists clinical decision-making and guides better targeted public health control efforts.[11]

Genomic differences between respiratory and non-respiratory isolates have not been comprehensively assessed in a programmatic setting. The implementation of routine sequencing (since 2016) of all culture-confirmed TB cases in New South Wales (NSW), Australia, presented a unique opportunity to compare genomic characteristics of *M. tuberculosis* strains isolated from different anatomical disease sites.

[1]Centre for Research Excellence in Tuberculosis (TB-CRE), Centenary Institute, Sydney, NSW, Australia
[2]Sydney Infectious Diseases Institute (Sydney ID), The University of Sydney, Sydney, NSW, Australia
[3]Centre for Infectious Diseases and Microbiology-Public Health, Westmead Hospital, Western Sydney Local Health District, Sydney, NSW, Australia
[4]NSW Mycobacterium Reference Laboratory, Centre for Infectious Diseases and Microbiology-Laboratory Services, Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Sydney, NSW, Australia
[5]Senior author
[6]These authors contributed equally
[7]Lead contact
*Correspondence: xiaomei.zhang@sydney.edu.au
https://doi.org/10.1016/j.isci.2024.110327

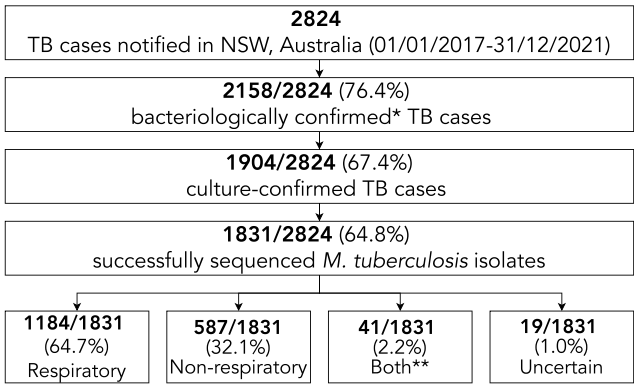


Figure 1. Flowchart of *M. tuberculosis* cultures and sequenced isolates included in the study

NSW: New South Wales; PCR: *M. tuberculosis*-specific Polymerase Chain Reaction; TB: tuberculosis. *Culture and/or PCR (*M. tuberculosis*-specific Polymerase Chain Reaction). **Respiratory and non-respiratory. All culture-confirmed cases were routinely sequenced in a prospective fashion. See also Table S1.

RESULTS

M. tuberculosis cultures and sequenced isolates

Nearly two-thirds (1,831/2,824; 64.8%) of notified TB cases in NSW during the study period were sequenced (Figure 1), including 96.2% (1,831/1,904) of culture-confirmed cases. Of the sequenced *M. tuberculosis* strains, 1,184 (64.7%) were from respiratory and 587 (32.1%) from non-respiratory sources. In 41 instances *M. tuberculosis* was cultured from both respiratory and non-respiratory sources; 1% (19/1,831) had no anatomical collection site specified (Figure 1 and Table S1).

TB patients with *M. tuberculosis* isolated from a respiratory or non-respiratory source

Table 1 presents the demographic characteristics, microbiological findings, and genomic information of TB patients based on their *M. tuberculosis* culture source. Among the 1,812 *M. tuberculosis* cultures, 55.6% (1,007/1,812) were from males, with adults aged 25–44 years accounting for 42.3% (767/1,812) of cases. The incidence of HIV within 1,812 *M. tuberculosis* cultures was found to be less than 1%.[12] All four major *M. tuberculosis* lineages were represented and 10.0% (181/1,812) of strains genomically clustered using a ≤5 SNP difference cut-off. Mixed *M. tuberculosis* strain populations were identified in 10.1% (183/1,812) of sequenced cultures. Among 587 non-respiratory specimens (Figure 2A) the majority were collected from lymph nodes (50.9%), followed by pleura (15.7%), musculoskeletal (10.2%), abdomen (9.0%), genitourinary (6.3%), CNS (2.2%), and other anatomical sites (5.6%); including unspecified abscess, blood, breast, chest wall, mediastinal, pericardial, and other sites (Table S2).

Multivariate analysis of demographic and genomic characteristics associated with *M. tuberculosis* isolation from a respiratory or non-respiratory source

Table 2 compares demographic characteristics between specimens obtained from respiratory and non-respiratory sources. A non-respiratory source was more common among female (36.6%; 289/789) than male (30.4%; 298/980) patients (adjusted odds ratios [aOR] 0.80, 95% confidence interval [CI] 0.66–0.98) (Figure 2B). A respiratory source was more common among older (≧65 years) adults (aOR 2.09, 95% CI 1.58–2.80), compared to younger adults (reference age 25–44 years) (Figure 2B). Respiratory specimens had a higher likelihood of being acid-fast bacilli (AFB) positive (odds ratio [OR] 4.74, 95% CI 3.49–6.57) and being part of a genomic cluster (aOR 1.91, 95% CI 1.31–2.83), with a trend to being drug resistant (DR) to first-line drugs (aOR 1.29, 95% CI 0.94–1.78) (Figure 2C; Table 3). Interestingly, a higher proportion of non-respiratory specimens demonstrated mixed strain infection (11.9%, 70/587) compared to respiratory specimens (9.5%, 112/1,182), although this difference was not statistically significant (aOR 0.77, 95% CI 0.56–1.07) (Table 3). Mixed strain infections were most commonly detected in lymph node specimens (52.9%, 37/70) (Table S2). Compared to all other lineages combined, lineage 3 strains were less likely to be isolated from respiratory specimens (aOR 0.62, 95% CI 0.47–0.81) (Figure 2C; Table 3). A detailed assessment of the relative frequency of *M. tuberculosis* sub-lineages identified in specimens from respiratory and non-respiratory sources did not suggest any sub-lineage specific tissue tropism (Figure 3).

Potential genomic transmission routes and genetic differences observed among TB cases

Figure 4 provides an overview of lineage specific genomic clusters (using a ≤5 SNP cut-off) identified. Interestingly, 36 clustered strains were from non-respiratory isolates, with pleural isolates most likely to be included in a genomic cluster (aOR 2.41; 95% CI 1.08–5.08) (Table S3) compared to all other non-respiratory isolates. Based on genomic analysis and the temporality of specimen receipt, seven patients with *M. tuberculosis* sequenced from a non-respiratory source were identified as potential transmitters; two each with pleural, musculoskeletal, or genitourinary disease and one with lymph node disease (Figure S1). Among the 41 TB cases with isolates from both respiratory

Table 1. Demographic and genomic characteristics of TB patients with *M. tuberculosis* isolated from a respiratory or non-respiratory source

	Specimen source no. (%)			
Characteristic	Respiratory	Non-respiratory	Respiratory and non-respiratory	Total
Gender				
Female	500 (62.3)	289 (36.0)	14 (1.7)	803
Male	682 (67.7)	298 (29.6)	27 (2.7)	1,007
Unknown	2 (100)	0	0	2
Total	1,184 (65.3)	587 (32.4)	41 (2.3)	1,812
Age group (years)				
<25	217 (69.3)	91 (29.1)	5 (1.6)	313
25–44	453 (59.1)	296 (38.6)	18 (2.3)	767
45–64	240 (65.2)	117 (31.8)	11 (3.0)	368
≧65	274 (75.3)	83 (22.8)	7 (1.9)	364
Total	1,184 (65.3)	587 (32.4)	41 (2.3)	1,812
Auramine AFB smear				
Pos	362 (86.4)	50 (11.9)	7 (1.7)	419
Neg	822 (59.0)	537 (38.5)	34 (2.4)	1,393
Total	1,184 (65.3)	587 (32.4)	41 (2.3)	1,812
Major strain lineage				
Lineage 1	371 (65.5)	184 (32.5)	11 (1.9)	566
Lineage 2	374 (69.6)	152 (28.3)	11 (2.0)	537
Lineage 3	152 (53.9)	122 (43.3)	8 (2.8)	282
Lineage 4	287 (67.2)	129 (30.2)	11 (2.6)	427
Total	1,184 (65.3)	587 (32.4)	41 (2.3)	1,812
Strain populations				
Mixed	112 (61.2)	70 (38.3)	1 (0.5)	183
Single	1,072 (65.8)	517 (31.7)	40 (2.5)	1,629
Total	1,184 (65.3)	587 (32.4)	41 (2.3)	1,812
p/gDST				
RR + MDR	33 (73.3)	11 (24.4)	1 (2.2)	45
DR (not RR/MDR)	130 (71.8)	49 (27.1)	2 (1.1)	181
DS	1,021 (64.4)	527 (33.2)	38 (2.4)	1,586
Total	1,184 (65.3)	587 (32.4)	41 (2.3)	1,812
Genomic clusters				
0-SNP	79 (79.8)	17 (17.2)	3 (3.0)	99
2-SNP	124 (77.5)	33 (20.6)	3 (1.9)	160
5-SNP	141 (77.9)	36 (19.9)	4 (2.2)	181

AFB, acid-fast bacilli; DR, drug resistant; DS, drug susceptible; DST, drug susceptibility testing; gDST, genotypic DST; pDST, phenotypic DST; MDR, multidrug-resistant (resistant to both rifampicin and isoniazid); Mixed, mixed strain populations; Neg, negative; Pos, positive; RR, rifampicin resistance; Single, single strain population; SNP, single nucleotide polymorphism.

and non-respiratory sources during the same disease episode, 39 had both isolates successfully sequenced, with a maximum 70 days apart between sample collection. Of these, 22 (56.4%) had 0 SNP differences, 7 (18.0%) had ≥2 SNP differences, and 4 (10.3%) had ≥5 SNP differences—ranging from 5 to 10 SNPs (Figure 2D and Table S4). No identical mutations were found among any of the extra-pulmonary strains.

DISCUSSION

This study presents the first comprehensive description of the demographic and genomic characteristics associated with *M. tuberculosis* strains isolated from respiratory and non-respiratory sources in a low incidence setting. Although most TB patients had pulmonary disease,

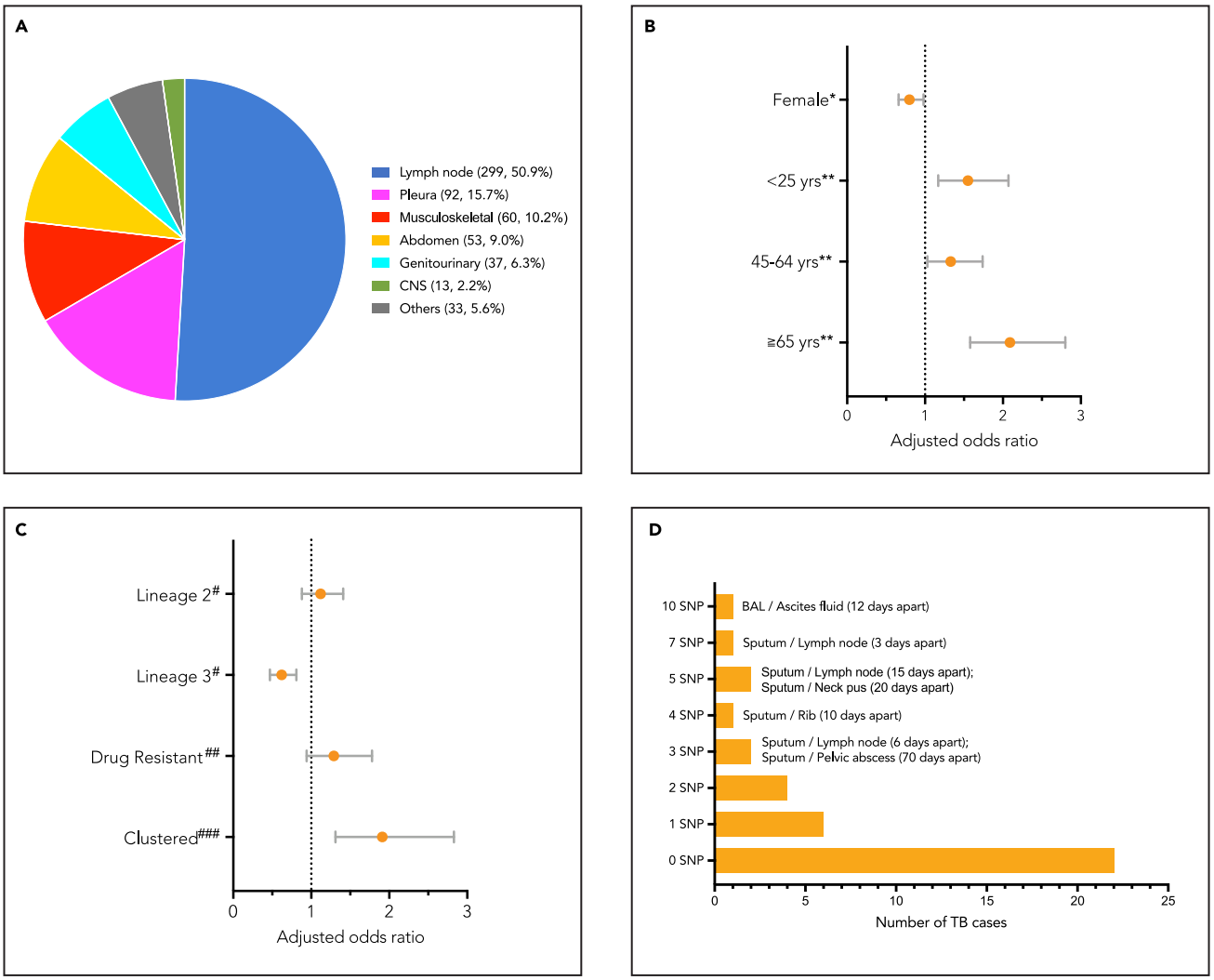


Figure 2. Overview of all sequenced *M. tuberculosis* isolates (*N* = 1,812)

(A) Proportion of sequenced *M. tuberculosis* isolates obtained from different non-respiratory sources (*n* = 587).

(B) Multivariate analysis of demographic features and (C) genomic features of respiratory and non-respiratory *M. tuberculosis* isolates.

(D) Genomic (SNP) distance of *M. tuberculosis* sequenced from both respiratory and non-respiratory sources[+] in the same patient during the same disease episode. BAL, bronco-alveolar lavage; CNS, central nervous system; SNP, single nucleotide polymorphism; TB, tuberculosis; yrs, years. Odds ration below 1 favors non-respiratory isolates, and above 1 favors respiratory isolates.

(A) The "Others" category includes unspecified abscess, blood, breast, chest wall, mediastinal, pericardial, and other sites. See also Table S2.

(B) *Male used as reference; **25–44 year olds used as reference. Odds ratios were adjusted for gender and age group. See also Table 2.

(C) [#]All others combined used as reference; [##]drug susceptible strains used as reference; [###] "Unclustered" strains (>5 SNP threshold) used as reference. Odds ratios were adjusted for lineage 2, lineage 3, presence of drug resistance, and genomic clustering. See also Table 3.

(D) [+]*M. tuberculosis* was isolated from both respiratory and non-respiratory isolates in 41 patients; 39/41 (95.1%) were successfully sequenced from both sources (difference in collection timing in brackets). See also Table S4.

nearly a third of cultures were recovered from non-respiratory specimens. This is broadly similar to the PTB/EPTB case ratio observed in NSW and in global TB notification data.[1,12] Although the early detection and effective treatment of PTB cases is important for disease control, accurate detection of diverse EPTB presentations is important for optimal patient outcomes and patient-centerd care. The diversity of sources from which *M. tuberculosis* were grown, reflects the broad range of clinical presentations and affected organs.[2]

The male predominance observed among TB cases is consistent with findings in other settings,[3,5,12,13] although a greater proportion of non-respiratory specimens in our study were collected from female patients. It has been postulated that EPTB is likely to be more common in patients with HIV infection or other immunocompromising conditions, which preferentially affects women in some settings.[14,15] However, the HIV-infection rate in our cohort was very low and did not support this viable explanation. It may be that women are inherently more vulnerable

Table 2. Multivariate analysis of demographic characteristics associated with *M. tuberculosis* isolation from a respiratory or non-respiratory source

	Specimen source[a]			Crude OR		Adjusted OR[b]	
Characteristic	Respiratory	Non-respiratory	Total	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Gender							
Female	500	289	789	0.76 (0.62–0.92)	0.006	0.80 (0.66–0.98)	0.03
Male	682	298	980	Ref.		Ref.	
Age group (years)							
<25	217	91	308	1.57 (1.18–2.09)	0.0021	1.55 (1.17–2.07)	0.003
25–44	451	296	747	Ref.		Ref.	
45–64	240	117	357	1.35 (1.03–1.76)	0.03	1.33 (1.03–1.74)	0.03
≧65	274	83	357	2.17 (1.63–2.90)	<0.0001	2.09 (1.58–2.80)	<0.0001

CI, confidence interval; OR, odds ratio.
[a]Excluding those with respiratory and non-respiratory isolates.
[b]Gender and age group were included in the multivariable logistic regression. Odds ratios were adjusted for gender and age group used in this model. Odds ration below 1 favors non-respiratory isolates, and above 1 favors respiratory isolates. See also Figure 2B.

to develop extra-pulmonary TB,[16,17] or alternatively, men could be predisposed to pulmonary disease due to intrinsic factors or behaviors such as cigarette smoking that is more common among men than women.[18]

A finding of particular interest is the fact that the PTB/EPTB ratio was highest among older adults (≥65 years), suggesting a potentially increased transmission risk within this age group. The over-representation of pulmonary cases among older adults has relevance for TB control efforts, particularly in regions with an aging population linked to global demographic shifts.[19] This poses a particular challenge in areas with high TB prevalence and a rapidly aging population. These emerging patterns highlight the need for better tailored approaches to address the distinct challenges and risks associated with TB in older adults.[20]

Table 3. Genomic characteristics of *M. tuberculosis* isolates from a respiratory or non-respiratory *M. tuberculosis* source

	Specimen source[a]			Crude OR		Adjusted OR[b]	
Characteristic	Respiratory	Non-respiratory	Total	OR (95% CI)	*p* value	OR (95% CI)	*p* value
Strain lineage							
Lineage 1	370	184	554	1.00 (0.81–1.24)	0.98	–	–
Lineage 2	373	152	525	1.32 (1.06–1.65)	0.014	1.12 (0.88–1.41)	0.36
Lineage 3	152	122	274	0.56 (0.43–0.73)	<0.0001	0.62 (0.47–0.81)	0.0005
Lineage 4	287	129	416	1.14 (0.90–1.45)	0.28	–	–
All others combined				Ref.		Ref.	
Strain populations							
Mixed	112	70	182	0.77 (0.56–1.07)	0.11	–	–
Single	1,070	517	1,587	Ref.			
p/gDST							
Any DR	163	60	223	1.41 (1.03–1.94)	0.03	1.29 (0.94–1.78)	0.12
RR/MDR	33	11	44	1.50 (00.78–3.14)	0.25	–	–
DR (not RR/MDR)	130	49	179	1.36 (0.97–1.93)	0.08	–	–
DS	1,019	527	1,546	Ref.		Ref.	
Genomic clusters (≤5 SNP difference)							
Clustered	141	36	177	2.07 (1.43–3.07)	0.0002	1.91 (1.31–2.83)	0.001
Unclustered	1,041	551	1,592	Ref.		Ref.	

CI, confidence interval; DR, drug resistant; DS, drug susceptible; MDR, multi-drug resistant (resistant to both rifampicin and isoniazid); OR, odds ratio; RR, rifampicin-resistant; SNP, single nucleotide polymorphism.
[a]Excluding those with respiratory and non-respiratory isolates.
[b]Lineage 2, lineage 3, DR, and genomic clusters at 5-SNP level were included in the multivariable logistic regression. Odds ratios were adjusted for lineage 2, lineage 3, DR, and genomic clusters at 5-SNP level used in this model. Odds ration below 1 favors non-respiratory isolates, and above 1 favors respiratory isolates. See also Figure 2C.

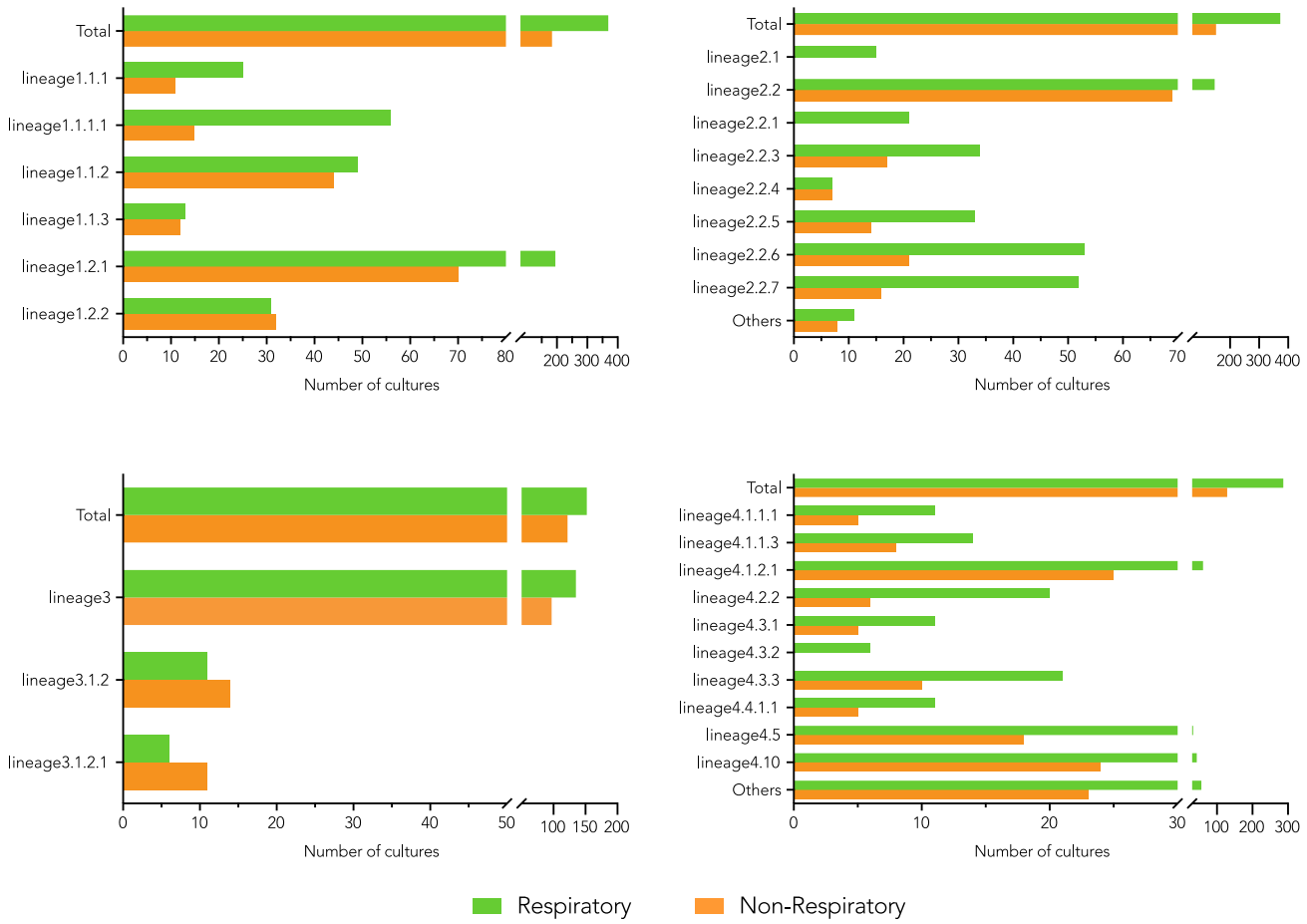


Figure 3. Relative frequency of *M. tuberculosis* lineages and sub-lineages identified in specimens from respiratory and non-respiratory sources
Others encompass sub-lineages with less than 5 representatives each (3 from lineage 2 and 17 from lineage 4).

Variations in the prevalence of *M. tuberculosis* strain lineages and sub-lineages across different anatomical sites may suggest strain-specific tropism or preferences for specific anatomical environments.[6,21,22] For instance, there is a relative overrepresentation of lineage 2 strains and underrepresentation of lineage 3 strains in respiratory specimens, which is consistent with previous reports.[22–25] Recognition of these lineage-specific trends may have clinical relevance if it provides insight into different pathogenesis or transmission patterns. The increased detection of mixed strain infections in non-respiratory isolates aligns with previous findings,[26,27] which demonstrated that lymph nodes remain infected for prolonged periods of time, and that reinfecting strains often co-locate in the same lymph nodes or other extrapulmonary tissues that were previously infected.[28]

Respiratory specimens were more frequently associated with genomic transmission clusters, which is not unexpected given the respiratory route of *M. tuberculosis* transmission. Patients with non-respiratory isolates were mostly identified as "dead end" hosts, with no indication of onwards transmission, but in some instances patients with EPTB may have contributed to transmission. While EPTB cases are not regarded as major sources of infection,[29,30] their occasional contribution to transmission warrants careful consideration. Pleural isolates were more commonly associated with transmission clusters, which suggest that pleural disease may be a proxy of lung involvement and potential transmission risk.[31] More detailed assessment of the clinical phenotype and detailed epidemiological analysis is required to assess potential transmission risk from other non-respiratory specimens.

Our study documented within-host genetic variability of *M. tuberculosis*, with a maximum 10-SNPs difference between strains observed in TB cases where cultures were collected from both respiratory and non-respiratory sources during the same disease episode. These variations likely represent within-host microevolution, which is supported by previous findings that $\leq$10 SNPs differences between isolates from the same patient are indicative of within host clonal diversification.[26,27] These findings highlight the within-host genetic diversity, which complicates absolute SNP cut-off definitions for TB cluster identification, and emphasize the importance epidemiological data to help elucidate TB transmission dynamics, especially in clusters with $\geq$2–5 SNP differences.

In conclusion, the comparative analysis of *M. tuberculosis* isolates from respiratory and non-respiratory specimens in a low incidence setting revealed anatomical site-specific differences in demographic, microbiological, and genomic characteristics. These differences may

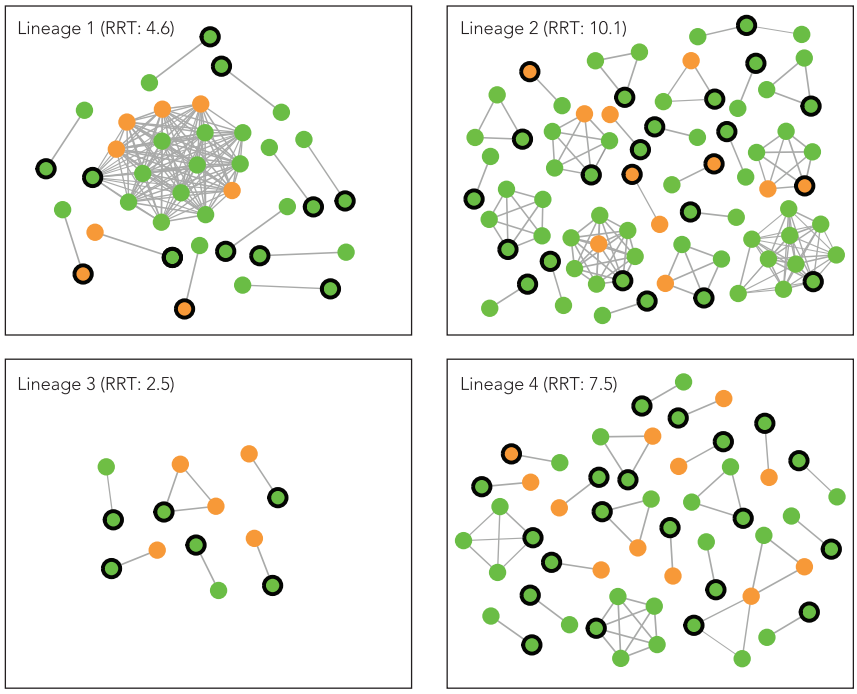


Figure 4. Overview of genomically clustered isolates with indication of respiratory or non-respiratory source

RRT: rate of recent transmission;[9] SNP, single nucleotide polymorphism. Genomically clustered isolates were identified sing a ≤5-SNP cut-off and including 41 patients in whom *M. tuberculosis* was cultured from both respiratory and non-respiratory sources, categorized as a respiratory source or pulmonary disease. Coloured dots indicate the following specimen sources: green respiratory, orange non-respiratory. A black halo identifies the specimen with the earliest collection date within a cluster, indicating likely temporality. The estimated RRT was calculated using the formula (N-C)/T*100, where N is the number of clustered isolates (using a 5-SNP cut off), C the number of clusters and T the total number of isolates analyzed.[9] See also Table S1 and Figure S1.

influence disease presentation, timeliness of diagnosis and treatment initiation, the risk of drug resistance, and transmission dynamics. This emphasizes the importance of individualizing diagnostic and treatment approaches, with careful consideration of the most appropriate public health responses.

Limitations of the study

Important limitations of our study need to be acknowledged. We relied on data captured by our laboratory information management system that lacked individual-level clinical and patient outcome data. Importantly, we were unable to cross-correlate genomic findings with detailed contact mapping and relevant epidemiological information from the field. We acknowledge this as a major limitation and hope to incorporate such data in future investigations to explore the association between genomic characteristics and clinical outcomes, as well as the impact of TB interventions on transmission dynamics. However, the relatively large longitudinal and representative dataset of culture-confirmed TB cases with the vast majority of cases being sequenced, strengthen the representativeness of our findings.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Characteristics assessed
  - Laboratory testing and genome sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110327.

ACKNOWLEDGMENTS

The authors thank the Sydney Informatics Hub and the University of Sydney's high-performance computing cluster, Artemis. The first author (X.Z.) is funded by NHMRC Centre for Research Excellence in Tuberculosis.

Funding: NHMRC Centre for Research Excellence in Tuberculosis (www.tbcre.org.au) and NSW Health Prevention Research Support Program.

AUTHOR CONTRIBUTIONS

B.J.M. and V.S. designed the study and guided the data analysis. X.Z. performed analysis and drafted the manuscript. All authors participated in manuscript revision and approved the final version.

DECLARATION OF INTERESTS

The authors declare no conflicts of interests.

Received: February 5, 2024
Revised: May 23, 2024
Accepted: June 18, 2024
Published: June 20, 2024

REFERENCES

1. World Health Organization Global Tuberculosis Report 2023 (2023 (World Health Organization).
2. Golden, M.P., and Vikram, H.R. (2005). Extrapulmonary tuberculosis: an overview. Am. Fam. Physician *72*, 1761–1768.
3. Rolo, M., González-Blanco, B., Reyes, C.A., Rosillo, N., and López-Roa, P. (2023). Epidemiology and factors associated with Extra-pulmonary tuberculosis in a Low-prevalence area. J. Clin. Tuberc. Other Mycobact. Dis. *32*, 100377. https://doi.org/10.1016/j.jctube.2023.100377.
4. Kang, W., Liu, S., Du, J., Tang, P., Chen, H., Liu, J., Ma, J., Li, M., Qin, J., Shu, W., et al. (2022). Epidemiology of concurrent extrapulmonary tuberculosis in inpatients with extrapulmonary tuberculosis lesions in China: a large-scale observational multi-centre investigation. Int. J. Infect. Dis. *115*, 79–85. https://doi.org/10.1016/j.ijid.2021.11.019.
5. Pang, Y., An, J., Shu, W., Huo, F., Chu, N., Gao, M., Qin, S., Huang, H., Chen, X., and Xu, S. (2019). Epidemiology of Extrapulmonary Tuberculosis among Inpatients, China, 2008-2017. Emerg. Infect. Dis. *25*, 457–464. https://doi.org/10.3201/eid2503.180572.
6. Hesseling, A.C., Marais, B.J., Kirchner, H.L., Mandalakas, A.M., Brittle, W., Victor, T.C., Warren, R.M., and Schaaf, H.S. (2010). Mycobacterial genotype is associated with disease phenotype in children. Int. J. Tuberc. Lung Dis. *14*, 1252–1258.
7. Séraphin, M.N., Doggett, R., Johnston, L., Zabala, J., Gerace, A.M., and Lauzardo, M. (2017). Association between *Mycobacterium tuberculosis* lineage and site of disease in Florida, 2009-2015. Infect. Genet. Evol. *55*, 366–371. https://doi.org/10.1016/j.meegid.2017.10.004.
8. van Leeuwen, L.M., Versteegen, P., Zaharie, S.D., van Elsland, S.L., Jordaan, A., Streicher, E.M., Warren, R.M., van der Kuip, M., and van Furth, A.M. (2019). Bacterial Genotyping of Central Nervous System Tuberculosis in South Africa: Heterogenic *Mycobacterium tuberculosis* Infection and Predominance of Lineage 4. J. Clin. Microbiol. *57*, e00415-19. https://doi.org/10.1128/jcm.00415-19.
9. Zhang, X., Martinez, E., Lam, C., Crighton, T., Sim, E., Gall, M., Donnan, E.J., Marais, B.J., and Sintchenko, V. (2023). Exploring programmatic indicators of tuberculosis control that incorporate routine *Mycobacterium tuberculosis* sequencing in low incidence settings: a comprehensive (2017–2021) patient cohort analysis. Lancet Reg. Health West. Pac. *41*, 100910. https://doi.org/10.1016/j.lanwpc.2023.100910.
10. Zhang, X., Lam, C., Martinez, E., Sim, E., Crighton, T., Marais, B.J., and Sintchenko, V. (2023). Genomic markers of drug resistance in *Mycobacterium tuberculosis* populations with minority variants. J. Clin. Microbiol. *61*, e0048523. https://doi.org/10.1128/jcm.00485-23.
11. Donnan, E.J., Marais, B.J., Coulter, C., Waring, J., Bastian, I., Williamson, D.A., Sherry, N.L., Bond, K., Sintchenko, V., Meumann, E.M., et al. (2023). The use of whole genome sequencing for tuberculosis public health activities in Australia: a joint statement of the National Tuberculosis Advisory Committee and Communicable Diseases Genomics Network. Commun. Dis. Intell. *47*, 47. https://doi.org/10.33321/cdi.2023.47.8.
12. NSW Tuberculosis Program, C.D.B. (2023). Tuberculosis in NSW – Surveillance Report 2021 (Health Protection NSW).
13. Solovic, I., Jonsson, J., Korzeniewska-Koseła, M., Chiotan, D.I., Pace-Asciak, A., Slump, E., Rumetshofer, R., Abubakar, I., Kos, S., Svetina-Sorli, P., et al. (2013). Challenges in diagnosing extrapulmonary tuberculosis in the European Union, 2011. Euro Surveill. *18*, 20432.
14. Mohammed, H., Assefa, N., and Mengistie, B. (2018). Prevalence of extrapulmonary tuberculosis among people living with HIV/AIDS in sub-Saharan Africa: a systemic review and meta-analysis. HIV AIDS (Auckl) *10*, 225–237. https://doi.org/10.2147/hiv.S176587.
15. da Rocha Dias, A.P.G., von Amann, B., Costeira, J., Gomes, C., and Bárbara, C. (2016). Extrapulmonary tuberculosis in HIV infected patients admitted to the hospital. Eur Respiratory Soc *48*, PA2761.
16. Min, J., Park, J.S., Kim, H.W., Ko, Y., Oh, J.Y., Jeong, Y.J., Na, J.O., Kwon, S.J., Choe, K.H., Lee, W.Y., et al. (2023). Differential effects of sex on tuberculosis location and severity across the lifespan. Sci. Rep. *13*, 6023. https://doi.org/10.1038/s41598-023-33245-5.
17. Yang, Z., Kong, Y., Wilson, F., Foxman, B., Fowler, A.H., Marrs, C.F., Cave, M.D., and Bates, J.H. (2004). Identification of risk factors for extrapulmonary tuberculosis. Clin. Infect. Dis. *38*, 199–205. https://doi.org/10.1086/380644.
18. Gaifer, Z. (2017). Epidemiology of extrapulmonary and disseminated tuberculosis in a tertiary care center in Oman. Int. J. Mycobacteriol. *6*, 162–166. https://doi.org/10.4103/ijmy.ijmy_31_17.
19. Teo, A.K.J., Morishita, F., Islam, T., Viney, K., Ong, C.W.M., Kato, S., Kim, H., Liu, Y., Oh, K.H., Yoshiyama, T., et al. (2023). Tuberculosis in older adults: challenges and best practices in the Western Pacific Region. Lancet Reg. Health West. Pac. *36*, 100770. https://doi.org/10.1016/j.lanwpc.2023.100770.
20. Teo, A.K.J., Rahevar, K., Morishita, F., Ang, A., Yoshiyama, T., Ohkado, A., Kawatsu, L., Yamada, N., Uchimura, K., Choi, Y., et al. (2023). Tuberculosis in older adults: case studies from four countries with rapidly ageing populations in the western pacific region. BMC Publ. Health *23*, 370. https://doi.org/10.1186/s12889-023-15197-7.
21. Marais, B.J., Hesseling, A.C., Schaaf, H.S., Gie, R.P., van Helden, P.D., and Warren, R.M. (2009). *Mycobacterium tuberculosis*

transmission is not related to household genotype in a setting of high endemicity. J. Clin. Microbiol. *47*, 1338–1343. https://doi.org/10.1128/jcm.02490-08.

22. Sankar, M.M., Singh, J., Diana, S.C.A., and Singh, S. (2013). Molecular characterization of Mycobacterium tuberculosis isolates from North Indian patients with extrapulmonary tuberculosis. Tuberculosis *93*, 75–83. https://doi.org/10.1016/j.tube.2012.10.005.
23. Holt, K.E., McAdam, P., Thai, P.V.K., Thuong, N.T.T., Ha, D.T.M., Lan, N.N., Lan, N.H., Nhu, N.T.Q., Hai, H.T., Ha, V.T.N., et al. (2018). Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat. Genet. *50*, 849–856. https://doi.org/10.1038/s41588-018-0117-9.
24. Kato-Maeda, M., Gagneux, S., Flores, L.L., Kim, E.Y., Small, P.M., Desmond, E.P., and Hopewell, P.C. (2011). Strain classification of Mycobacterium tuberculosis: congruence between large sequence polymorphisms and spoligotypes. Int. J. Tuberc. Lung Dis. *15*, 131–133.
25. Click, E.S., Moonan, P.K., Winston, C.A., Cowan, L.S., and Oeltmann, J.E. (2012). Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis. Clin. Infect. Dis. *54*, 211–219. https://doi.org/10.1093/cid/cir788.
26. Moreno-Molina, M., Shubladze, N., Khurtsilava, I., Avaliani, Z., Bablishvili, N., Torres-Puente, M., Villamayor, L., Gabrielian, A., Rosenthal, A., Vilaplana, C., et al. (2021). Genomic analyses of *Mycobacterium tuberculosis* from human lung resections reveal a high frequency of polyclonal infections. Nat. Commun. *12*, 2716. https://doi.org/10.1038/s41467-021-22705-z.
27. Guerra-Assunção, J.A., Crampin, A.C., Houben, R.M., Mzembe, T., Mallard, K., Coll, F., Khan, P., Banda, L., Chiwaya, A., Pereira, R.P., et al. (2015). Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife *4*, e05166. https://doi.org/10.7554/eLife.05166.
28. Ganchua, S.K.C., White, A.G., Klein, E.C., and Flynn, J.L. (2020). Lymph nodes-The neglected battlefield in tuberculosis. PLoS Pathog. *16*, e1008632. https://doi.org/10.1371/journal.ppat.1008632.
29. Sandgren, A., Hollo, V., and van der Werf, M.J. (2013). Extrapulmonary tuberculosis in the European Union and European Economic Area, 2002 to 2011. Euro Surveill. *18*, 20431.
30. Sharma, S.K., Mohan, A., and Kohli, M. (2021). Extrapulmonary tuberculosis. Expert Rev. Respir. Med. *15*, 931–948. https://doi.org/10.1080/17476348.2021.1927718.
31. Shaw, J.A., and Koegelenberg, C.F.N. (2021). Pleural Tuberculosis. Clin. Chest Med. *42*, 649–666. https://doi.org/10.1016/j.ccm.2021.08.002.
32. Votintseva, A.A., Pankhurst, L.J., Anson, L.W., Morgan, M.R., Gascoyne-Binzi, D., Walker, T.M., Quan, T.P., Wyllie, D.H., Del Ojo Elias, C., Wilcox, M., et al. (2015). Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. J. Clin. Microbiol. *53*, 1137–1143.
33. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.
34. Hunt, M., Bradley, P., Lapierre, S.G., Heys, S., Thomsit, M., Hall, M.B., Malone, K.M., Wintringer, P., Walker, T.M., Cirillo, D.M., et al. (2019). Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. Wellcome Open Res. *4*, 191.
35. Anyansi, C., Keo, A., Walker, B.J., Straub, T.J., Manson, A.L., Earl, A.M., and Abeel, T. (2020). QuantTB - a method to classify mixed Mycobacterium tuberculosis infections within whole genome sequencing data. BMC Genom. *21*, 80.
36. Stimson, J., Gardy, J., Mathema, B., Crudu, V., Cohen, T., and Colijn, C. (2019). Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. Mol. Biol. Evol. *36*, 587–603.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Mycobacterium tuberculosis	Institute of Clinical Pathology and Medical Research, Westmead hospital, NSW, Australia	
Critical commercial assays		
DNA extraction protocol	Votintseva et al. (2015)[32]	N/A
RNase A	QIAGEN	Cat#19101
DNeasy UltraClean Microbial Kits	QIAGEN	Cat#10196-4
Nextera XT library Prep Kit	Illumina	Cat#FC-131-1024
Deposited data		
The whole genome sequencing data used in this study are available on the NCBI Sequence Read Archive	This paper	NCBI SRA PRJNA899911
Software and algorithms		
Burrows-Wheeler Aligner	Li et al. (2013)[33]	https://github.com/lh3/bwa
Mykrobe predict/master	Hunt et al. (2019)[34]	https://github.com/Mykrobe-tools/mykrobe
Snippy v3.1	Torsten Seemann	https://github.com/tseemann/snippy
QuantTB v1.0	Anyansi et al. (2020)[35]	https://github.com/AbeelLab/quanttb
Transcluster	Stimson et al. (2019)[36]	https://github.com/JamesStimson/transcluster
RedDog v1beta.8	D. J. Edwards, B. J. Pope and K. E. Holt	https://github.com/katholt/RedDog
Prism v9.4.1	GraphPad	N/A

RESOURCE AVAILABILITY

Lead contact

- Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Xiaomei Zhang (Xiaomei.zhang@sydney.edu.au).

Materials availability

- This study did not generate new unique reagents.

Data and code availability

- Raw de-identified pathogen WGS data was deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA). They are publicly available as of the date of publication. BioProject number is listed in the key resources table and accession numbers are listed the Table S1.
- This paper does not report original code.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact Dr. Xiaomei Zhang (Xiaomei.zhang@sydney.edu.au).

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We analysed demographic, specimen location and genomic data of all culture-confirmed and routinely sequenced TB cases in NSW, Australia, with a specimen collection date between 1st January 2017 and 31st December 2021. Routine sequencing was performed at the

New South Wales (NSW) Mycobacterium Reference Laboratory (MRL) at the Institute of Clinical Pathology and Medical Research (ICPMR) NSW Health Pathology. In general, only one culture-positive isolate per patient was sequenced unless there were positive cultures from respiratory and non-respiratory sources. Isolates were classified as respiratory or non-respiratory depending on the specimen type recorded on the laboratory request form. Any specimen collected from the respiratory tract were classified as respiratory. Non-respiratory specimens were classified as lymph node, pleura, musculoskeletal, abdomen, genitourinary, central nervous system (CNS) and other.[2] Anatomical sites that were not specified were designated as 'uncertain' and excluded from comparative analyses.

METHOD DETAILS

Characteristics assessed

We reviewed all characteristics recorded in the NSW MRL database, including collection date, patient gender, patient age, auramine Acid-Fact Bacillus (AFB) smear and phenotypic drug susceptibility testing (DST) results. Genomic characteristics evaluated included strain lineage and sub-lineage, the presence of mixed strain infection or drug resistance conferring mutations, and genomic clusters.

Laboratory testing and genome sequencing

Phenotypic DST was performed using the modified microdilution method in the BACTEC MGIT 960 system with WHO recommended critical concentrations. All cultures identified as *M. tuberculosis* were routinely sequenced using Illumina NextSeq500 (Illumina, San Diego, California) instrument using 2 x 150bp paired-end chemistry and genomic characteristics determined as previously described.[9,10] In brief, *M. tuberculosis* species, major strain lineage and sub-lineage were predicted using Mykrobe predict/master. Instances of mixed *M. tuberculosis* strain infections, defined as simultaneous co-infection with more than one strain during the same disease episode, were detected by QuantTB v1.0, employing a genetic distinctness threshold of ≥ 100 single nucleotide polymorphisms (SNP) differences between strains. Mutations associated with first-line TB drug resistance were identified from the 2021 WHO Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. Genomic clusters were identified utilizing the hierarchical single-linkage agglomerative clustering algorithm in python package with genomic distance of ≤ 5 SNPs. Visualization of the identified genomic clusters were conducted using Transcluster package (https://github.com/JamesStimson/transcluster) by incorporating a 5-SNP cut-off together with the case collection date and an assumed molecular clock of 0.5 SNP per year per genome. The genomic distance between isolates in whom *M. tuberculosis* was cultured from both respiratory and non-respiratory sources were determined using RedDog v1beta.8 (https://github.com/katholt/RedDog) with default settings.

QUANTIFICATION AND STATISTICAL ANALYSIS

We performed descriptive statistical analyses using Prism GraphPad v9.4.1. Comparative analyses employed uni- and multivariable logistic regression models to assess differences between respiratory specimens and non-respiratory specimens. Univariable logistic regression models were used to determine crude odds ratios (OR) with 95% confidence intervals (CIs). Multivariable logistic regression models provided adjusted odds ratios (aOR) with 95% CIs with inclusion if $p < 0.05$ from the univariable logistic regression models. A two-sided p-value of < 0.05 was considered as statistically significant.