

RESEARCH ARTICLE

Whole Genome Sequencing Demonstrates Limited Transmission within Identified *Mycobacterium tuberculosis* Clusters in New South Wales, Australia

Ulzijargal Gurjav^{1,2*}, Alexander C. Outhred^{1,3}, Peter Jelfs^{2,4}, Nadine McCallum^{1,2}, Qinning Wang², Grant A. Hill-Cawthorne^{1,5}, Ben J. Marais^{1,3}, Vitali Sintchenko^{1,2}

1 Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Sydney, Australia, **2** Centre for Infectious Diseases and Microbiology—Public Health, Westmead Hospital, Sydney, Australia, **3** Children’s Hospital at Westmead, Sydney, Australia, **4** NSW Mycobacterium Reference Laboratory, Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research—Pathology West, Sydney, Australia, **5** School of Public Health and Westmead Institute for Medical Research, The University of Sydney, Sydney, Australia

* ugur6096@uni.sydney.edu.au



OPEN ACCESS

Citation: Gurjav U, Outhred AC, Jelfs P, McCallum N, Wang Q, Hill-Cawthorne GA, et al. (2016) Whole Genome Sequencing Demonstrates Limited Transmission within Identified *Mycobacterium tuberculosis* Clusters in New South Wales, Australia. PLoS ONE 11(10): e0163612. doi:10.1371/journal.pone.0163612

Editor: Srinand Sreevatsan, University of Minnesota, UNITED STATES

Received: June 11, 2016

Accepted: September 12, 2016

Published: October 13, 2016

Copyright: © 2016 Gurjav et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The research study was sponsored by the Centre for Infectious Diseases and Microbiology – Public Health, Westmead Hospital, Sydney, Australia and the NSW Mycobacterium Reference Laboratory, Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research – Pathology West, Sydney, Australia. Mrs. Gurjav

Abstract

Australia has a low tuberculosis incidence rate with most cases occurring among recent immigrants. Given suboptimal cluster resolution achieved with 24-locus mycobacterium interspersed repetitive unit (MIRU-24) genotyping, the added value of whole genome sequencing was explored. MIRU-24 profiles of all *Mycobacterium tuberculosis* culture-confirmed tuberculosis cases diagnosed between 2009 and 2013 in New South Wales (NSW), Australia, were examined and clusters identified. The relatedness of cases within the largest MIRU-24 clusters was assessed using whole genome sequencing and phylogenetic analyses. Of 1841 culture-confirmed TB cases, 91.9% (1692/1841) had complete demographic and genotyping data. East-African Indian (474; 28.0%) and Beijing (470; 27.8%) lineage strains predominated. The overall rate of MIRU-24 clustering was 20.1% (340/1692) and was highest among Beijing lineage strains (35.7%; 168/470). One Beijing and three East-African Indian (EAI) clonal complexes were responsible for the majority of observed clusters. Whole genome sequencing of the 4 largest clusters (30 isolates) demonstrated diverse single nucleotide polymorphisms (SNPs) within identified clusters. All sequenced EAI strains and 70% of Beijing lineage strains clustered by MIRU-24 typing demonstrated distinct SNP profiles. The superior resolution provided by whole genome sequencing demonstrated limited *M. tuberculosis* transmission within NSW, even within identified MIRU-24 clusters. Routine whole genome sequencing could provide valuable public health guidance in low burden settings.

was funded by a Mongolian Government Postgraduate scholarship supplemented by Westmead Foundation for Medical Research (Sydney, NSW) and the National Health and Medical Research Council Centre for Research Excellence in Tuberculosis Control (Canberra, ACT, Australia).

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Mycobacterium tuberculosis is a highly successful human pathogen. Estimates suggest that up to a third of global people are infected and that 9 million developed active disease in 2014 [1]. Recent data support the ancient origins of *M. tuberculosis* with evidence of ongoing adaptation and genetic diversification throughout human history [2]. Seven *M. tuberculosis* strain lineages have been identified with distinct geographic distribution patterns, shaped by ancient human migration pathways [2,3]. More recent migration patterns and increased population mobility are reshaping these geographic distributions.

Traditional genotyping methods such as mycobacterial interspersed repetitive unit (MIRU) analysis provided new insight into pathogen diversity and strain-specific transmission dynamics [4]. Different *M. tuberculosis* strain lineages have been associated with variable virulence, transmissibility, disease phenotypes and drug resistance profiles [5–7]. Molecular epidemiology studies also confirmed the transmissibility of drug-resistant strains and highlighted the importance of re-infection in tuberculosis (TB) endemic settings with uncontrolled transmission [8,9]. More recently the availability of whole-genome sequencing (WGS) has provided unprecedented strain resolution to enhance our understanding of *M. tuberculosis* evolution and transmission [10,11].

Australia is a low TB incidence setting with a stable incidence of 5–6 per 100,000 [12]. The most populous state of Australia, New South Wales (NSW), with 7.4 million population reports the highest TB case numbers with significant geographic clustering in and around Sydney [13]. The vast majority of TB cases occur among immigrants from high TB incidence countries [14]. Analysis of routine MIRU typing data showed that East African Indian (EAI) and Beijing lineage strains were most common, with a recent increase in the relative abundance of EAI lineage strains related to changes in migrant flows [4]. It also demonstrated high strain diversity within identified geographic hotspots, suggesting foci of imported disease rather than clusters of local transmission [15].

Despite these insights, the existence of clusters with identical MIRU-24 profiles provided a public health dilemma, since genotypic clustering is suggestive of local transmission that may require targeted public health intervention [4]. It is well recognized that accurate cluster identification is problematic with strains that are highly monomorphic and poorly differentiated using MIRU, such as Beijing lineage strains [16]. In order to assess the contribution of local TB transmission and guide public health intervention efforts we performed WGS of identified MIRU clusters to establish the frequency of true transmission chains within these clusters.

Materials and Methods

Genotyping of bacterial strains

Routine surveillance data collected between 2009 and 2013 at the Mycobacterium Reference Laboratory, NSW containing all culture confirmed *M. tuberculosis* isolates 24-loci MIRU (MIRU-24) genotypes and demographic data were reviewed. MIRU-24 genotyping was performed as described earlier [4] and used to (a) assign strain lineage using the miru-vntrplus.org online database, (b) to build a minimum spanning tree using Bionumerics v.5 (Applied-Maths, Kortrijk, Belgium), (c) to calculate allelic richness (AR) using HP-RARE software v.1 [17] and finally (d) to calculate Hunter Gaston Index of Diversity (HGDI, with the assumption that isolates were unrelated) [18]. In addition, 12-locus MIRU was employed to identify EAI sublineages in the SITVIT v.2 database [19]. Two or more isolates sharing an identical MIRU-24 profile were considered a genotype cluster and aggregated on the minimum spanning tree as a node. Isolates with a difference of one or more MIRU-24 loci were represented by separate

nodes; the distance between respective nodes (indicated by the number of line segments) reflected the number of MIRU-24 loci differences. The “rate of recent transmission” was calculated as follows [(number of clustered isolates—number of clusters) x 100/total number of cultured isolates] [4].

Whole-genome sequencing

The largest four MIRU-24 clusters identified during the study period were selected for WGS. These 4 clusters belonged to EAI Lineage 1 (clusters A and B with 5 and 4 isolates each) and Beijing Lineage 2 (clusters C and D with 11 and 10 isolates each) strains. Selected cultures of *M. tuberculosis* were initially stored at -80°C and then recovered on Middlebrook agar. Genomic DNA was extracted using Wizard Genomic DNA purification kit (Promega, Madison, WI, USA) and libraries were prepared using Ion Xpress Plus Fragment Library kit (Life Technologies, Gaithersburg, MD, USA). Sequencing was performed on an Ion Torrent Personal Genome Machine (Agilent Technologies, Palo Alto, CA, USA) with the Ion 318 chip kit (Life Technologies) as per manufacturer’s instructions. A reference genome for mapping was prepared from NC_000962.3 [20] by substituting gaps in place of repetitive elements (all regions annotated as PE/PPE/PGRS and *cysA* genes, insertion sequences, transposases and prophage components, with gaps representing 6.3% of NC_000962.3.RRE). Single nucleotide polymorphisms (SNPs) were called using the *mem* algorithm of the mapper *bwa* [21], followed by the variant-caller *freebayes* [22] operating on merged lineage-specific BAM files, with “ODDS > 99” used as a quality filter (“ODDS” is a composite value that represents marginal likelihood [23]; the resulting VCF files can be found in the [S1 File](#)). Lineage-specific whole-genome Bayesian inference substitution trees were generated with *mrbayes* [24], using NC_000962.3.RRE patched with SNPs (all non-SNP variants were excluded by filtering for CIGAR1X). A synthetic lineage-specific most recent common ancestor (MRCA) was generated by patching NC_000962.3.RRE with all SNPs shared within that lineage, and included in each tree to serve as a comparator. A difference between libraries of less than 10 SNPs was considered a SNP cluster, suggestive of recent transmission [24].

Statistical analysis and ethics approval

Demographic and clinical characteristics of different *M. tuberculosis* lineages were explored using descriptive statistics, using χ^2 and One-way ANOVA tests where applicable. All statistical analyses were performed using SPSS 23.0 (IBM, USA) and p-values less than 0.05 were considered significant. The study was approved by the Human Research Ethics Committee of the University of Sydney (project number 2013/126).

Results

A total of 1841 culture-confirmed TB cases were identified; 72% of all TB cases in NSW during the study period. Of these, 1692 (91.9%) had complete demographic and MIRU-24 genotyping data. The *M. tuberculosis* population structure included 17 different strain families, with Lineage 1 (EAI, 28%, n = 474) and Lineage 2 (Beijing, 27.8%, n = 470) accounting for more than half of all strains identified. Demographic, clinical and MIRU-24 clustering characteristics of the predominant strain families compared to minority strains are presented in [Table 1](#). The mean age of patients infected with Beijing and EAI lineage strains was similar (43 and 45 years respectively, p = 0.02), but Beijing strains caused disease with a bi-phasic age distribution being most common in young adults (15-29yrs) and in older people (>60yrs). Compared to all other cases, Beijing lineage strains were strongly associated with multi-drug resistant TB (odds ratio 5.0, CI 95% 1.7–14.9, p = 0.04).

Table 1. Demographic, clinical and MIRU-24 clustering characteristics of predominant *M. tuberculosis* strain lineages.

Characteristics	EAI [#]	Beijing	Other ^{##}	Total	p value
	n (%)	n (%)	n (%)	n (%)	
	N = 474	N = 470	N = 748	N = 1692	
Age group					
<15	6 (1.3)	2 (0.4)	14 (1.9)	22 (1.3)	0.001
15–29	101 (21.3)	197 (41.9)	278 (37.2)	576 (34.0)	
30–44	150 (31.6)	79 (16.8)	185 (24.7)	414 (24.5)	
45–59	113 (23.8)	75 (16.0)	110 (14.7)	298 (17.6)	
≥60	104 (21.9)	117 (24.9)	161 (21.5)	382 (22.6)	
Gender					
Male	272 (57.4)	264 (56.2)	437 (58.4)	973 (57.5)	0.74
Site of infection					
Respiratory	291 (61.4)	347 (73.8)	518 (69.3)	115 (68.3)	0.001
Drug resistance					
Isoniazid (low and high level)	40 (8.4)	51 (10.9)	46 (6.1)	137 (8.1)	0.001
MDR/XDR*	4 (0.8)	19 (4.0)	11 (1.5)	34 (2.0)	
MIRU-24 clustering					
No of clusters	32	54	37	123	
No of clustered isolates (%)	76 (16.0)	168 (35.7)	96 (12.7)	340 (20.1)	
Average cluster size (range)	2 (2–5)	3 (2–11)	2 (2–8)	2 (2–12)	
Estimated transmission rate ^{&}	9.3	24.3	7.9	12.8	

[#]EAI—East African Indian lineage strains

^{##} Other—minority strain lineages pooled together

*MDR/XDR—multi-drug resistant tuberculosis (including a single case of extensively drug resistant tuberculosis diagnosed in 2011); MIRU-24—24-locus mycobacterium interspersed repetitive unit strain typing method

[&]Estimated transmission rate = [(number of clustered isolates—number of clusters)/ total number of genotyped isolates]

doi:10.1371/journal.pone.0163612.t001

The TB incidence in NSW showed a decrease from 6.5/100,000 population in 2009 to 5.5/100,000 population in 2013 (Fig 1). Apart from 2011, when nearly a quarter of strains were clustered by MIRU-24, less than 20% of strains were clustered during each of the study years. The average clustering rate of 20.1% (340/1692) suggested that up to 12.8% of cases may have resulted from recent transmission, although the calculated mean cluster sizes were small (2–3 cases). Beijing family strains demonstrated the highest degree of clustering (168/470; 35.7%); comprising 49.4% (168/340) of all clustered stains (Table 1 and S1 Fig). The estimated recent transmission rate was highest among Beijing strains (24.3%), with cluster sizes varying between 2 and 11 cases (Table 1).

MIRU-24 based minimum spanning tree analysis of Beijing lineage strains identified a single clonal complex with four large nodes, while three independent complexes, each with 1–2 large nodes, were observed among EAI lineage strains (Fig 2). The HGDI index for EAI strains (0.43) was slightly higher than for Beijing lineage strains (0.25), as was the mean allelic richness (3.6 vs. 2.7; p = 0.065); MIRU loci 1955 and 2163 displayed the highest variability (S2 Fig). The nodes identified within the three EAI clonal complexes belonged to the following MIRU international types (MIT): complex 1 MIT56; complex 2 MIT59 and MIT272; and complex 3 MIT69 and MIT409 (S3 Fig).

All generated raw reads from whole-genome sequencing were submitted to the European Nucleotide Archive of the European Bioinformatics Institute under study accession number PRJEB11778. The ENA sample identification numbers are listed in S1 Table. After mapping,

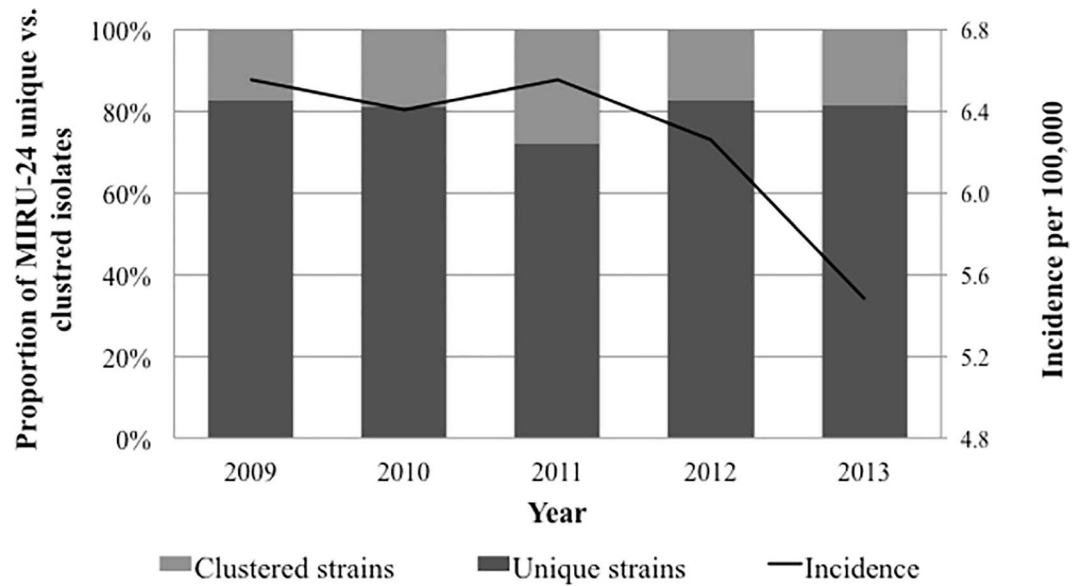


Fig 1. Tuberculosis incidence and genotypic clustering rate in NSW, Australia. MIRU-24–24-locus mycobacterium interspersed repetitive unit strain typing method.

doi:10.1371/journal.pone.0163612.g001

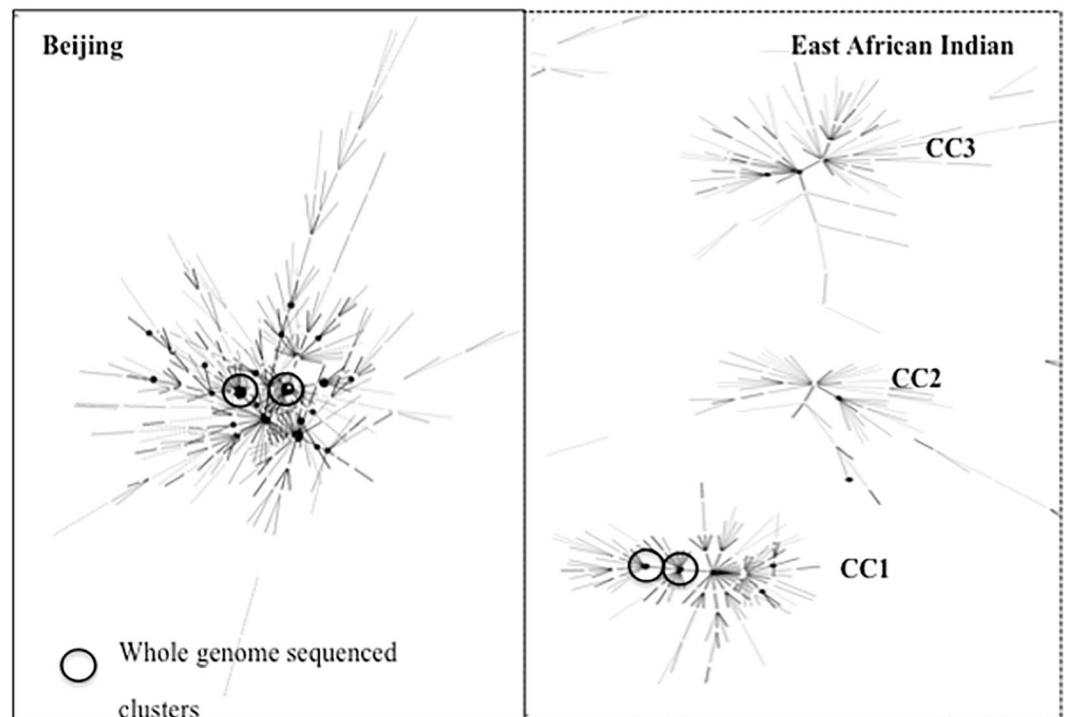


Fig 2. MIRU-24 minimum spanning tree of the predominant *M. tuberculosis* lineages identified in NSW, Australia. MIRU-24–24-locus mycobacterium interspersed repetitive unit strain typing method; Solid box shows single large clonal complex for Beijing; Dotted box shows 3 independent clonal complexes of East African Indian strain lineage; CC1 –clonal complex 1; CC2 –clonal complex 2; CC3 –clonal complex 3; Circles indicate clusters that were subjected to whole genome sequencing.

doi:10.1371/journal.pone.0163612.g002

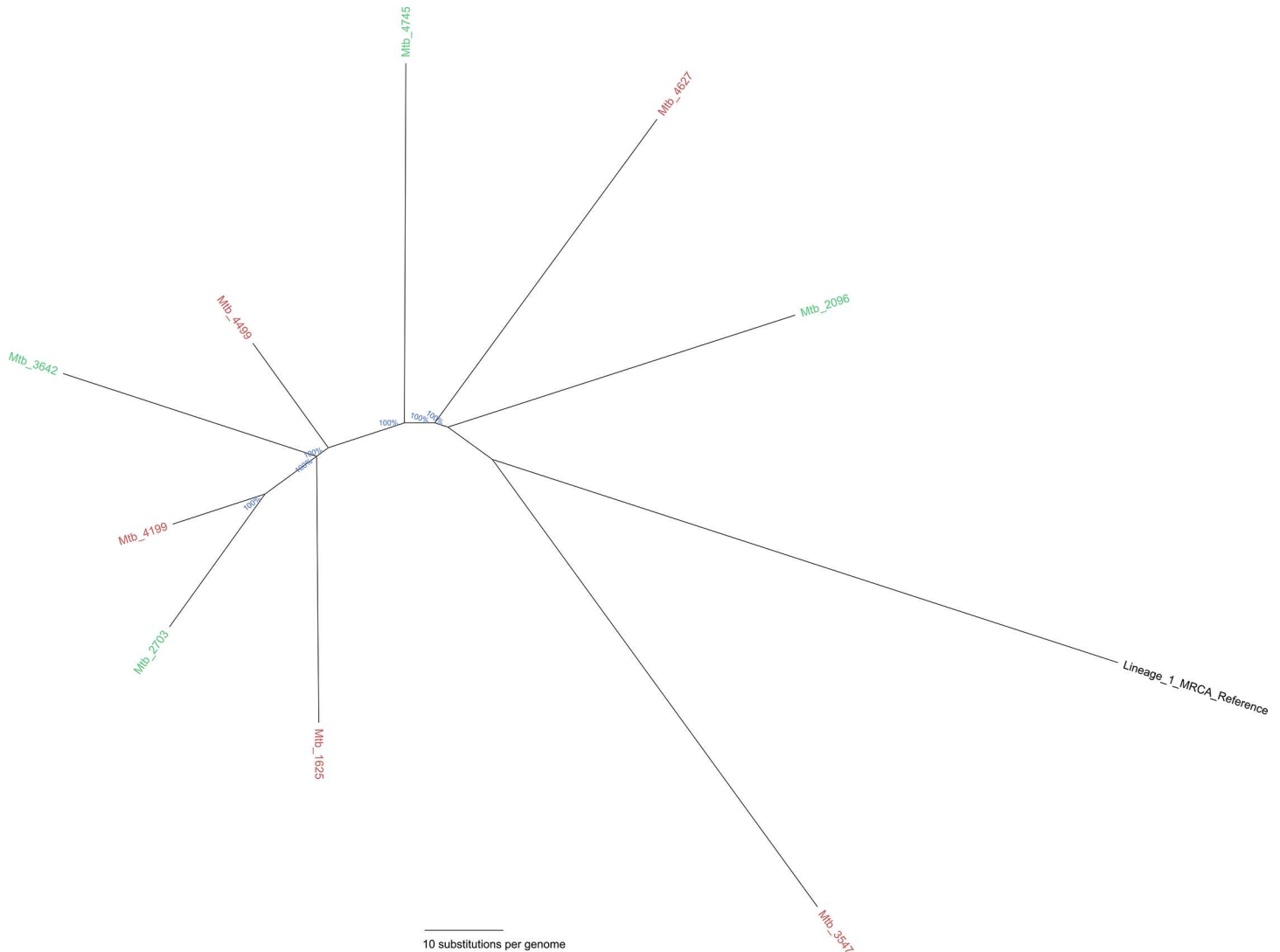


Fig 3. Whole-genome Bayesian inference distance tree of Lineage 1 East African Indian (EAI) clusters with identical MIRU-24 profiles (MIRU-24 clusters). MIRU-24 cluster A is labeled in red, and MIRU-24 cluster B in green and were only distantly related (>100 SNP differences) on whole genome sequencing without any SNP clusters identified.

doi:10.1371/journal.pone.0163612.g003

median read depth ranged from 40- to 100-fold, with a single library at 22-fold depth; for each library, reads covered >91% of the reference genome, and >91% of reads were mapped to the reference genome. VCF files listing the variants found (including filtered, informative SNPs annotated with snpEff [25]) are including in the [S1 File](#). Resistance-associated SNPs that could have contributed to homoplasy (affecting subsequent phylogenetic inference) were sought but not found.

[Fig 3](#) reflects the whole-genome Bayesian inference substitution tree for Lineage 1 MIRU-24 clusters A and B (identical MIRU-24 profiles); both from clonal complex1. WGS demonstrated no SNP clusters within MIRU-24 clusters A and B, thereby reducing the MIRU-based calculated transmission rate by 100%. Demographic and epidemiological data demonstrated no link between these cases, suggesting that strains with identical MIRU-24 profiles were acquired overseas; 7 in the Philippines, one in Ethiopia and the country of origin remained unknown for one patient. SNP-based phylogenetic analysis of clusters C and D from within the Lineage 2

clonal complex identified by MIRU-24 (Fig 4) demonstrated 3 small SNP clusters embedded within. MIRU-24 cluster C included a single 3-member SNP cluster (0–2 SNP differences). Importantly, two of three isolates from this SNP cluster were classified as probable laboratory cross-contamination following careful assessment by laboratory and public health investigators. MIRU-24 cluster D contained two SNP clusters with two members each (0 SNP differences); one of these clusters contained another isolate that was classified as probable laboratory contamination by laboratory and public health investigators. In hindsight, of the 21 Beijing/Lineage 2 isolates clustered by MIRU-24, only a single 2-member SNP cluster represented likely transmission.

Discussion

WGS demonstrated that large *M. tuberculosis* genotype clusters identified with routine MIRU-24 typing were not indicative of local transmission in this low TB incidence setting dominated by imported disease. The use of WGS-defined SNP clusters reduced the MIRU-24 potential secondary case rate for Beijing lineage strains by 79% (from 19/19 to 4/19) with a further 16% reduction (from 4/19 to 1/19) once likely laboratory contamination events were excluded. Overall SNP-based analysis of the four large MIRU-24 clusters (30 specimens in total) suggested only a single case of local transmission (reducing the number of potential secondary cases from 26 to 1). The fact that laboratory contaminated isolates were clustered by WGS allowed for a critical review of specimen collection and processing methods and offered practical guidance to public health authorities. We reconfirmed previous observations [26] that Beijing lineage strains are most prevalent among young adults, with increased rates of respiratory disease and drug resistance compared to other strains. Our findings highlighted the challenge of interpreting MIRU-24 clusters for highly homoplastic Beijing lineage strains [27], particularly in low incidence settings dominated by imported disease.

Interestingly, WGS also demonstrated added value in the evaluation of EAI lineage clusters, despite higher MIRU-24 allelic diversity observed in these strains. Not a single transmission chain was identified within the largest MIRU-24 clusters selected (9 isolates in total). The observation that EAI lineage strains were more prevalent than Beijing among patients with non-respiratory TB and demonstrated equal spread across the age spectrum, support previous observations that suggested reduced transmissibility of EAI strains [28,29]. The EAI MIRU-24 clonal complexes identified are well recognized within the Asia-Pacific region and may provide a clue as to the likely geographic origin of these strains. For example MIT56 identified within clonal complex 1 has been found in the Philippines and Japan [30,31], while MIT59 and MIT272 identified within clonal complex 2 is known to circulate in Vietnam [32,33]. Together these findings shed important light on the likely routes of importation of various *M. tuberculosis* strains into Australia, which may influence pre- and post-immigration screening practices [15] and direct Australian support for TB control initiatives in the region [34].

The sub-optimal resolution associated with MIRU-24 genotyping could be attributable to a slow molecular clock (depending on the experiment, it ranges from 1×10^{-5} to 1×10^{-2} per locus per year) and the fact that it considers less than 1% of the *M. tuberculosis* genome, compared to an estimated SNP mutation rate of 0.3 SNP per genome per year [20,35] with an assessment of the majority of genome; excluding only PE/PPE and PGRS family genes from the genomic comparison. The lower MIRU-24 allelic diversity and single large MIRU-24 cluster of Beijing lineage strains identified are not unexpected, given that Beijing lineage strains have less genetic diversity than EAI lineage strains [36]. Interestingly, despite its reduced MIRU-24 allelic diversity (Table 1), Beijing lineage strains have been associated with increased rates of drug resistance. However, the mechanisms underlying the generation and spread of drug resistant mutations are completely different to those generating MIRU-24 allelic diversity.

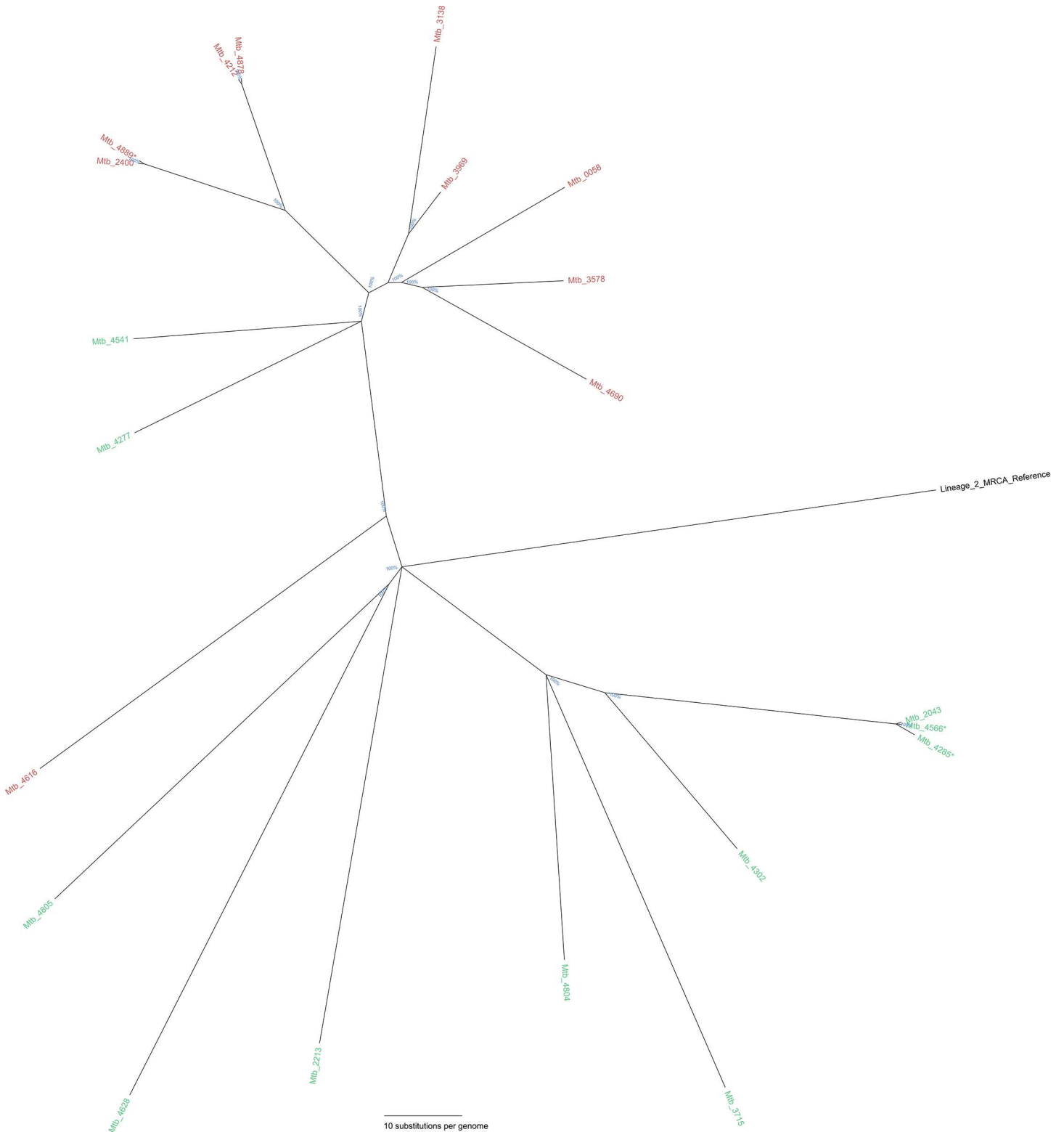


Fig 4. Whole-genome Bayesian inference distance tree of Lineage 2 Beijing clusters with identical MIRU-24 profiles (MIRU-24 clusters). MIRU-24 cluster C is labeled in green, and MIRU cluster D in red and three SNP clusters were identified. Three libraries of the two SNP clusters were determined to represent cross-contamination during diagnostic culture are marked with an asterisk. Branch support probabilities are displayed as percentages in blue. The scale bars represents 10 substitutions per genome for the corresponding distance from a node. Parameters, output and version information for *mrbayes* can be found in [S2 File](#).

doi:10.1371/journal.pone.0163612.g004

Study limitations include the fact that we sequenced only 4 large genotype clusters, two each from Beijing and EAI lineage strains. These two lineages accounted for 55.8% of all culture-confirmed TB cases in NSW and makes up 71.8% of all clustered cases. Although this limited WGS analysis does not represent all transmission events in NSW, it does indicate that false MIRU-24 clustering is a major problem in low incidence settings and that local transmission is less common than MIRU-24 clustering suggests. A recent study from the United Kingdom demonstrated that routine WGS is poised to replace conventional methods in mycobacterial reference laboratories, given the rapidly reducing cost as well as clinical and public health relevance of the findings [37]. Apart from improving the accuracy of assessing recent transmission events, WGS may reduce costs related to unnecessary contact investigations in low incidence settings [38].

In conclusion, MIRU-24 typing could lead to false cluster identification, especially with highly monomorphic *M. tuberculosis* strains lineages, such as Beijing. This is a particular problem in low burden settings with minimal local TB transmission, where most TB cases represent imported disease. Application of WGS to assess SNP clusters improves the accuracy of recent transmission estimates and provides valuable public health guidance. Thus we suggest that routine WGS should replace traditional genotyping methods in low burden settings with adequate resources.

Supporting Information

S1 Table. The European Nucleotide Archive sample identification numbers.

(DOCX)

S1 Fig. *M. tuberculosis* strain lineage distribution among genotypically clustered and unique strains.

(PPTX)

S2 Fig. Comparison of MIRU-24 loci allelic richness for EAI and Beijing strain lineages.

(PPTX)

S3 Fig. Minimum spanning tree of all culture-confirmed and MIRU-24 typed *M. tuberculosis* isolates identified between 2009–2013 in NSW, Australia.

(PPTX)

S1 File. VCF files describing the variants found using whole-genome sequencing.

(ZIP)

S2 File. Settings, output and version information for mrbayes.

(ZIP)

Acknowledgments

The authors thank the staff of the Centre for Infectious Diseases and Microbiology and NSW Mycobacterium Reference Laboratory for technical assistance in MIRU-24 typing and sequencing. UG was funded by a Mongolian Government Postgraduate scholarship and the Centre for Infectious Diseases and Microbiology–Public Health supplemented by a grant from the NHMRC Centre for Research Excellence in Tuberculosis Control and The Westmead Foundation for Medical Research.

Author Contributions

Conceptualization: UG BM VS.

Methodology: AO NM QW.

Resources: PJ.

Supervision: BM VS.

Validation: AO.

Writing – original draft: UG AO.

Writing – review & editing: UG AO GHC BM VS.

References

1. World Health Organization. Global Tuberculosis Report. 2015. <www.who.int>
2. Comas I, Coscollà M, Luo T, Borrell S, Holt KE, Kato-Maeda M et al. 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 45:1176–82. doi: [10.1038/ng.2744](https://doi.org/10.1038/ng.2744) PMID: [23995134](https://pubmed.ncbi.nlm.nih.gov/23995134/)
3. Gagneux S, Small PM. 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 7:328–337. doi: [10.1016/S1473-3099\(07\)70108-1](https://doi.org/10.1016/S1473-3099(07)70108-1) PMID: [17448936](https://pubmed.ncbi.nlm.nih.gov/17448936/)
4. Gurjav U, Jelfs P, McCallum N, Marais BJ, Sintchenko V. 2014. Temporal dynamics of *Mycobacterium tuberculosis* genotypes in New South Wales, Australia. *BMC Infect Dis* 14:455. doi: [10.1186/1471-2334-14-455](https://doi.org/10.1186/1471-2334-14-455) PMID: [25149181](https://pubmed.ncbi.nlm.nih.gov/25149181/)
5. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. 2013. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* 45:784–790. doi: [10.1038/ng.2656](https://doi.org/10.1038/ng.2656) PMID: [23749189](https://pubmed.ncbi.nlm.nih.gov/23749189/)
6. Thwaites G, Caws M, Chau TT, D'Sa A, Lan NT, Huyen MN, et al. 2008. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J Clin Microbiol* 46:1363–1368. doi: [10.1128/JCM.02180-07](https://doi.org/10.1128/JCM.02180-07) PMID: [18287322](https://pubmed.ncbi.nlm.nih.gov/18287322/)
7. Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. 2011. *Mycobacterium tuberculosis* Beijing genotype: a template for success. *Tuberculosis* 91:510–523. doi: [10.1016/j.tube.2011.07.005](https://doi.org/10.1016/j.tube.2011.07.005) PMID: [21835699](https://pubmed.ncbi.nlm.nih.gov/21835699/)
8. Marais BJ, Mlambo CK, Rastogi N, Zozzio T, Duse A, Victor T, et al. 2013. Epidemic spread of multi-drug-resistant (MDR) tuberculosis in Johannesburg, South Africa. *J Clin Microbiol* 51: 1818–1825. doi: [10.1128/JCM.00200-13](https://doi.org/10.1128/JCM.00200-13) PMID: [23554196](https://pubmed.ncbi.nlm.nih.gov/23554196/)
9. Verver S, Warren RM, Beyers N, Richardson M, van der Spuy GD, Borgdorff MW, et al. 2005. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. *Am J Respir Crit Care Med* 171: 1430–1435. doi: [10.1164/rccm.200409-1200OC](https://doi.org/10.1164/rccm.200409-1200OC) PMID: [15831840](https://pubmed.ncbi.nlm.nih.gov/15831840/)
10. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. 2015. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 47:242–249. doi: [10.1038/ng.3195](https://doi.org/10.1038/ng.3195) PMID: [25599400](https://pubmed.ncbi.nlm.nih.gov/25599400/)
11. Outhred AC, Holmes N, Sadsad R, Martinez E, Jelfs P, Hill-Cawthorne GA, et al. 2016. Identifying likely transmission pathways within a 10-year community outbreak of tuberculosis by deep whole genome sequencing. *PLoS ONE* 11:e0150550. doi: [10.1371/journal.pone.0150550](https://doi.org/10.1371/journal.pone.0150550) PMID: [26938641](https://pubmed.ncbi.nlm.nih.gov/26938641/)
12. Barry C, Waring J, Stapledon R, Konstantinos A, and the National Tuberculosis Advisory Committee, for the Communicable Diseases Network Australia. 2012. Tuberculosis notifications in Australia, 2008 and 2009. *Commun Dis Intell* 36:82–94. PMID: [23153084](https://pubmed.ncbi.nlm.nih.gov/23153084/)
13. Massey PD, Durrheim DN, Stephens N, Christensen A. 2013. Local level epidemiological analysis of TB in people from a high incidence country of birth. *BMC Public Health* 13:62. doi: [10.1186/1471-2458-13-62](https://doi.org/10.1186/1471-2458-13-62) PMID: [23339706](https://pubmed.ncbi.nlm.nih.gov/23339706/)
14. Lowbridge C, Christensen A, McAnulty JM. 2013. EpiReview: Tuberculosis in NSW, 2009–2011. *NSW Public Health Bull* 24:3. doi: [10.1071/NB12115](https://doi.org/10.1071/NB12115) PMID: [23849020](https://pubmed.ncbi.nlm.nih.gov/23849020/)
15. Gurjav U, Jelfs P, Hill-Cawthorne GA, Marais BJ, Sintchenko V. 2015. Genotype heterogeneity of *Mycobacterium tuberculosis* within geospatial hotspots suggests foci of imported infection in Sydney, Australia. *Infection Genet Evol Epub ahead of print* doi: [10.1016/j.meegid.2015.07.014](https://doi.org/10.1016/j.meegid.2015.07.014)
16. Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CR, Hoek KG, et al. 2008. Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of *Mycobacterium tuberculosis*

- Beijing strains in a setting of high incidence of tuberculosis. *J Clin Microbiol* 46:3338–3344. doi: [10.1128/JCM.00770-08](https://doi.org/10.1128/JCM.00770-08) PMID: [18716230](https://pubmed.ncbi.nlm.nih.gov/18716230/)
17. Kalinowski S. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 5:187–189. doi: [10.1111/j.1471-8286.2004.00845.x](https://doi.org/10.1111/j.1471-8286.2004.00845.x)
 18. Hunter PR, Gaston MA. 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* 26:2465–2466. PMID: [3069867](https://pubmed.ncbi.nlm.nih.gov/3069867/)
 19. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. 2012. SITVITWEB—a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 12:755–66. doi: [10.1016/j.meegid.2012.02.004](https://doi.org/10.1016/j.meegid.2012.02.004) PMID: [22365971](https://pubmed.ncbi.nlm.nih.gov/22365971/)
 20. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998 Jun 11; 393(6685):537–44. doi: [10.1038/31159](https://doi.org/10.1038/31159) PMID: [9634230](https://pubmed.ncbi.nlm.nih.gov/9634230/)
 21. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio] [Internet]. 2013 Mar 16; Available from: <http://arxiv.org/abs/1303.3997>
 22. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv:12073907 [q-bio] [Internet]. 2012 Jul 17; Available from: <http://arxiv.org/abs/1207.3907>
 23. Ronquist F, Teslenko M, Mark P van der, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol*. 2012 May 1; 61(3):539–42. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/)
 24. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 13:137–146. doi: [10.1016/S1473-3099\(12\)70277-3](https://doi.org/10.1016/S1473-3099(12)70277-3) PMID: [23158499](https://pubmed.ncbi.nlm.nih.gov/23158499/)
 25. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012 Jun; 6(2):80–92. doi: [10.4161/fly.19695](https://doi.org/10.4161/fly.19695) PMID: [22728672](https://pubmed.ncbi.nlm.nih.gov/22728672/)
 26. Couvin D, Rastogi N. 2015. Tuberculosis—A global emergency: Tools and methods to monitor, understand, and control the epidemic with specific example of the Beijing lineage. *Tuberculosis* 95:S177–189. doi: [10.1016/j.tube.2015.02.023](https://doi.org/10.1016/j.tube.2015.02.023) PMID: [25797613](https://pubmed.ncbi.nlm.nih.gov/25797613/)
 27. Luo T, Yang C, Peng Y, Lu L, Sun G, Wu J, et al. 2014. Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis* 94:434–440. doi: [10.1016/j.tube.2014.04.005](https://doi.org/10.1016/j.tube.2014.04.005) PMID: [24888866](https://pubmed.ncbi.nlm.nih.gov/24888866/)
 28. Albanna AS, Reed MB, Kotar KV, Fallow A, McIntosh FA, Behr MA, et al. 2011. Reduced transmissibility of East African Indian strains of *Mycobacterium tuberculosis*. *PLoS ONE* 6:e25075. doi: [10.1371/journal.pone.0025075](https://doi.org/10.1371/journal.pone.0025075) PMID: [21949856](https://pubmed.ncbi.nlm.nih.gov/21949856/)
 29. Gallego B, Sintchenko V, Jelfs P, Coiera E, Gilbert GL. 2010. Three-year longitudinal study of genotypes of *Mycobacterium tuberculosis* in a low prevalence population. *Pathology* 42:267–72. doi: [10.3109/00313021003631346](https://doi.org/10.3109/00313021003631346) PMID: [20350221](https://pubmed.ncbi.nlm.nih.gov/20350221/)
 30. Sia IG, Buckwalter SP, Doerr KA, Lugos S, Kramer R, Orillaza-Chi R, et al. 2013. Genotypic characteristics of *Mycobacterium tuberculosis* isolated from household contacts of tuberculosis patients in the Philippines. *BMC Infect Dis* 13:571. doi: [10.1186/1471-2334-13-571](https://doi.org/10.1186/1471-2334-13-571) PMID: [24308751](https://pubmed.ncbi.nlm.nih.gov/24308751/)
 31. Millet J, Miyagi-Shiohira C, Yamane N, Sola C, Rastogi N. 2007. Assessment of mycobacterial interspersed repetitive unit-QUB markers to further discriminate the Beijing genotype in a population-based study of the genetic diversity of *Mycobacterium tuberculosis* clinical isolates from Okinawa, Ryukyu Islands, Japan. *J Clin Microbiol* 45:3606–3615. doi: [10.1128/JCM.00348-07](https://doi.org/10.1128/JCM.00348-07) PMID: [17898160](https://pubmed.ncbi.nlm.nih.gov/17898160/)
 32. Nguyen VA, Choisy M, Nguyen DH, Tran TH, Pham KL, Thi Dinh PT, et al. 2012. High prevalence of Beijing and EAI4-VNM genotypes among *M. tuberculosis* isolates in Northern Vietnam: Sampling effect, rural and urban disparities. *PLoS ONE* 7:e45553. doi: [10.1371/journal.pone.0045553](https://doi.org/10.1371/journal.pone.0045553) PMID: [23029091](https://pubmed.ncbi.nlm.nih.gov/23029091/)
 33. Caws M, Thwaites G, Dunstan S, Hawn TR, Lan NT, Thuong NT, et al. 2008. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathogens* 4:e1000034. doi: [10.1371/journal.ppat.1000034](https://doi.org/10.1371/journal.ppat.1000034) PMID: [18369480](https://pubmed.ncbi.nlm.nih.gov/18369480/)
 34. Schwartzman K1, Oxlade O, Barr RG, Grimard F, Acosta I, Baez J, et al. 2005. Domestic returns from investment in the control of tuberculosis in other countries. *New Engl J Med* 353:1008–1020. doi: [10.1056/NEJMsa043194](https://doi.org/10.1056/NEJMsa043194) PMID: [16148286](https://pubmed.ncbi.nlm.nih.gov/16148286/)
 35. Ragheb MN, Ford CB, Chase MR, Lin PL, Flynn JL, Fortune SM. 2013. The mutation rate of mycobacterial repetitive unit loci in strains of *M. tuberculosis* from cynomolgus macaque infection. *BMC Genomics* 14:145. doi: [10.1186/1471-2164-14-145](https://doi.org/10.1186/1471-2164-14-145) PMID: [23496945](https://pubmed.ncbi.nlm.nih.gov/23496945/)

36. Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, et al. 2008. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. PLoS Pathogens 4:e1000160. doi: [10.1371/journal.ppat.1000160](https://doi.org/10.1371/journal.ppat.1000160) PMID: [18802459](https://pubmed.ncbi.nlm.nih.gov/18802459/)
37. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et al. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 4:49–58 doi: [10.1016/S2213-2600\(15\)00466-X](https://doi.org/10.1016/S2213-2600(15)00466-X) PMID: [26669893](https://pubmed.ncbi.nlm.nih.gov/26669893/)
38. Kwong JC, McCallum N, Sintchenko V, Howden BP. 2015. Whole genome sequencing in clinical and public health microbiology. Pathology 47:199–210. doi: [10.1097/PAT.000000000000235](https://doi.org/10.1097/PAT.000000000000235) PMID: [25730631](https://pubmed.ncbi.nlm.nih.gov/25730631/)