

Methods Used to Investigate the Transmission of Tuberculosis Using Whole Genome Sequencing: A Systematic Review

Chacha M Issarow*

Department of Integrative Biomedical Sciences, Faculty of Health Sciences, University of Cape Town, South Africa.

*Corresponding Author: Chacha M Issarow

Email: chacha.issarow@uct.ac.za

Abstract

Background: Whole-Genome Sequencing (WGS) is a powerful tool in the investigation of Tuberculosis (TB) transmission and disease recurrence due to its high discriminatory power compared to other genotyping methods. This study aimed to review methods used to assess and describe the transmission dynamics of Mycobacterium Tuberculosis (MTB) in low and high TB incidence settings using WGS data.

Method: Four electronic databases were used for searching studies that used WGS to assess TB transmission either in low or high TB burden settings. Studies published in English from 2008 onward applying WGS to describe TB transmission dynamics in humans were eligible for review.

Results: The majority of studies reviewed (28 studies) used single nucleotide polymorphism (SNP) threshold methods to assess and describe TB transmission by computing genetic distance between paired isolates, based on the notion that shorter genetic distances indicate recent TB transmission. Of 28 studies that used the SNP threshold approach, eight studies also used the presence of drug resistance conferring mutations to assess and describe recent TB transmission, assuming that isolates with the same mutations originated from the same source. Three studies developed new methods to assess and describe TB transmission, of which two studies formed transmission trees using novel Bayesian and epidemiological model approaches, and one study identified clusters based on probabilistic matching. One study assessed and described TB transmission based on SNP distances by determining the number of isolates per isolate in a range of ≤ 10 SNPs referred to as transmission index.

Conclusion: While all methods have technical pros and cons, the SNP threshold method was primarily used to investigate TB transmission in both low and high TB incidence settings. Contact tracing was applied in very few studies based in low incidence settings as it is challenging in high incidence areas due to the overall TB burden.

Keywords: Mutation; Recent TB Transmission; Drug Resistance; SNP Threshold.

Introduction

Tuberculosis (TB) remains a global public health concern, particularly in low- and middle-income countries [34]. The emergence of Drug-Resistant TB (DR-TB), including Multidrug Resistant (MDR) and extensively Drug-Resistant TB (XDR-TB) remains a challenge in TB treatment and control worldwide [34,46,47]. Due to the complex nature of TB, including DR-TB,

traditional epidemiological methods are insufficient to infer the exact transmission in populations. The complexity of the TB epidemic is due to the fact that the majority of exposed or infected individuals do not develop disease, and that there is often a long-time frame between exposure and disease. Additionally, TB complexity may be influenced by many other factors, including the biology of the bacillus, failure of the public health care and comorbidities, such as HIV. Reducing

Citation: Issarow CM. Methods Used to Investigate the Transmission of Tuberculosis Using Whole Genome Sequencing: A Systematic Review. Med Discoveries. 2026; 5(3): 1289.

ongoing transmission is key to effective control of TB [60]. A better understanding of how, where and when TB is transmitted would help to triage appropriate strategies for prevention of transmission.

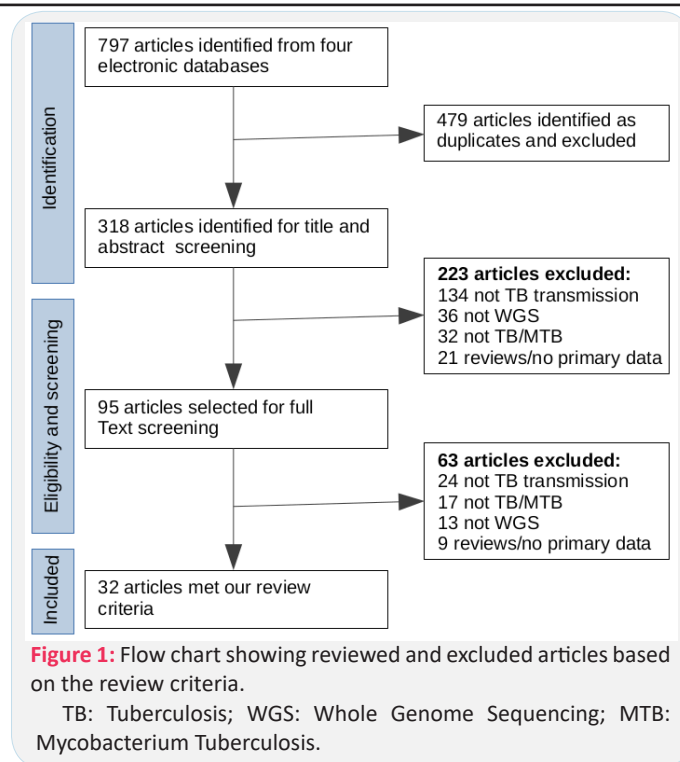
Previously, it was assumed that DR-TB was mainly attributable to the acquisition of drug resistant conferring mutations during inadequate therapy or poor treatment adherence [47]. However, recent modelling and epidemiological data show that drug resistance caused by direct transmission of already drug-resistant TB strains from person-to-person plays an important role even among previously treated cases [4,20,22,38,45]. Additionally, molecular epidemiological studies have demonstrated that direct transmission can account for up to 84% of the notified DR-TB cases, particularly in high burden regions [51]. Mycobacterium Tuberculosis (MTB) comprises different phylogenetic lineages [39], and it was suggested that genetic differences between bacterial clades could account for different epidemiological characteristics such as transmission [36,37,39]. Additionally, different evolutionary rates of MTB strains might be among the key factors influencing the genetic diversity of these bacteria [31,62].

Due to decreasing costs, Whole-Genome Sequencing (WGS) is becoming more widely applied to study TB transmission as well as the evolution of drug resistance, and has begun to inform knowledge around these important aspects of TB pathogenesis [7,19]. WGS identifies sequence variations at the whole genome level, and, therefore, has the potential to help in quantifying and describing transmission with more accuracy compared to traditional genotyping methods [7,19,22,56]. Compared to other existing genotyping methods, WGS is better able to discriminate between relapse and re-infection mechanisms of TB recurrence, especially when patients are re-infected with closely related strains [40,41,58], identify transmission chains [2,21], measure within-host diversity [63-67], and differentiate between direct transmission and acquired drug resistance [19,52]. WGS is therefore increasingly being used to predict epidemiological links between TB cases and to assist in transmission investigation and interruption [43,48,49,54,57-59]. Researchers have used a range of methods to assess TB transmission across low and high TB burden settings and with respect to TB drug resistance. This review aimed to describe all the methods that used WGS to assess TB transmission (drug susceptible or DR-TB) either in low or high TB burden settings, and to describe potential challenges and limitations.

Methods

Search strategy: Four electronic databases (PubMed, Scopus, Escudos and Web of Science) were searched for cross-sectional and cohort studies published from 2008 onward, assessing and describing TB transmission using WGS. In the database, the keywords “Tuberculosis”, or “TB”, and “transmission”, and “whole genome sequencing” were used for searching.

Review criteria: All articles published in English describing TB transmission in humans using WGS data either qualitatively or quantitatively were eligible for review. Therefore, publications aligned with our review criteria of assessing or describing TB transmission dynamics using WGS published in English from 2008 onward were identified and imported into Zotero for references. Review and articles with no primary data were excluded.



Results

Study identification and review: Based on our review search key terms, a total of 798 publications were identified in the four electronic databases. Of the 798 identified articles, 480 were excluded as duplicates, and 318 articles remained for title and abstract screening. Of the 318 articles screened, 95 articles were selected for full text screening, of which 32 met our review criteria (Figure 1). Of the 32 studies included, 17 referred to DR-TB and the remaining 15 studies referred to Drug Susceptible TB (DS-TB). Of the 32 studies that met our review criteria as mentioned in the methods, 14 studies were conducted in low (high-income countries) and 18 studies in high (low-income countries) TB burden settings [16,34].

Of the 32 studies reviewed in this study, genetic distance between paired isolates (SNP threshold) and similarity of drug resistance conferring mutations were among the most common methods used to infer recent or direct transmission of TB strains from person-to-person [20]. A total of 28 studies applied SNP threshold method, of which eight studies also used presence of drug resistance conferring mutations to confirm recent or direct TB transmission, and two studies [27,28], used both SNP threshold and newly developed methods to assess and describe TB transmission. Four [29-32], of 32 studies developed new methods to assess and describe TB transmission.

TB transmission description using the SNP threshold approach: The SNP threshold (defined as a cut-off number of SNPs that differ between paired isolates) places two or more individuals in the same transmission cluster if there is a shorter genetic distance (number of SNPs that differ between two sequences) than a specified SNP threshold between their genomes. One study [28], assumed that paired isolates or cluster of genomically linked cases with a genetic distance of <6 SNPs confirm recent TB transmission. Clustering isolates with shorter genetic distances than a specified SNP threshold together with identical drug resistance conferring mutations was used in some studies to define recent TB transmission. For example, one study [22], found that 32% of MDR strains

were in a cluster that differed by ≤ 12 SNPs, indicating recent transmission of MDR strains. However, in some clustering methods, a pair of isolates with a greater genetic distance than a specified SNP threshold might be in the same transmission cluster if linked by chains of intermediate unsampled cases [31], which highlights that assessment of transmission is highly dependent on the completeness of sampling.

Of 32 studies reviewed in this study, the majority (28 studies) used the SNP threshold method to assess and describe TB transmission [24]. Specifically, five studies [3,9-12,21], used a genetic distance of ≤ 3 SNPs, ten studies [1-11,14,15,19,26], used a genetic distance of ≤ 5 SNPs, four studies [1,6,16,27], used a genetic distance of ≤ 10 SNPs to confirm recent or direct transmission and four studies [13,17,20,22], used a genetic distance of ≤ 12 SNPs to form transmission clusters. Of these four studies that used ≤ 12 SNPs to infer transmission, three studies [13,17,20], combined contact tracing and WGS to confirm recent TB transmission, and were conducted in a low-TB burden setting. Contact tracing complemented with genotyping of MTB isolates is important for understanding disease transmission [17]. The SNP threshold method was therefore used together with contact tracing in three reviewed studies from low incidence settings to assess and describe TB transmission. However, TB contact tracing is mainly used in low incidence settings (high-income countries), as it is challenging in high incidence areas due to the overall TB burden [15]. However, TB contact screening is primarily household based in high incidence settings and often focusses on children.

Bjorn-Mortensen et al [9], suggest that defining a minimum distance of 12 SNPs as a threshold for unlikely recent transmission and a maximum distance of 5 SNPs as the threshold for likely transmission is adequate in low incidence settings, but such a threshold is more difficult to define in high TB burden settings. In high burden settings, certain strains may have been circulating for a long time, and thus the MTB diversity will be limited with many of these “endemic” strains differing by fewer than 12 SNPs. Many of the studies reviewed here used WGS based on the SNP threshold approach to suggest TB transmission without comparison to epidemiological contact data. However, differences in sample processing prior to WGS, variability in data analysis approaches (pipelines) and differences in sampling intensity remain significant challenges [31,42]. Multiple sample processing and WGS data analysis pipelines exist that differ widely in output formats, making SNP threshold standardization difficult even in the same setting [42].

TB transmission assessment using drug resistance conferring mutations: Apart from the SNP threshold approach, comparison of first-line or second-line drug resistance conferring mutations was used in eight studies to infer recent or direct TB transmission [8]. It is assumed that isolates with the same drug resistance mutational profile originated from the same common ancestor (if only DR strains are sampled and the SNP threshold is below a certain value), which indicates recent or direct transmission [8]. Of the 32 studies reviewed, eight studies [19-26], used clustering based on both SNP thresholds and comparison of drug resistance conferring mutations to assess TB transmission. For example, using a genetic distance of ≤ 12 SNPs and assuming that resistance via canonical mutation does not occur in parallel, Yang et al [22], found that 89.5% of the genomic clusters had resistance mutations for isoniazid and rifampicin that were consistent among clustered strains, confirming direct transmission of MDR strains rather than

acquired resistance. Additionally, Caselli et al [21], found that strains containing the same mutations conferring resistance to rifampicin, isoniazid, streptomycin and ethambutol clustered phylogenetically, thus confirming direct transmission of DR-TB strains from an inferred common ancestor. Using a genetic distance of < 50 SNPs to define transmission clusters, Clark et al [23], found that isolates in the same cluster had almost (not all) identical mutations conferring resistance to rifampicin and isoniazid, suggesting direct transmission of MDR-TB. However, this study (Clark et al) did not require that all isolates within a cluster had identical drug resistance conferring mutations to confirm transmission of MDR strain. This implies the fact that there are highly common drug resistance conferring mutations in clinical isolates, so that having the same mutations in clustering isolates does not necessarily mean that transmission took place.

The majority of reviewed studies used the MTB H37Rv, which is a lineage 4 strain [42], as the reference genome to identify site specific sample-specific genomic variations. However, one of the reviewed studies [18], suggested that existing methods for comparative analysis of isolates of using a single MTB strain, such as H37Rv, as the reference genome, may limit resolution (discriminatory power). The study [18], therefore used a pan-genome reference and formed clusters based on SNP threshold approach to assess TB transmission and compared results with other methods that used the MTB H37Rv strain as the reference. The pan-genome is derived from more than 100 MTB reference genomes representing lineages 1-4 for read mapping prior to variant calling. This approach allows the comparison and clustering of a large number of diverse samples using a pan-genome reference sequence inferred computationally. Using a genetic distance of < 13 SNPs for transmission cluster link detection, the authors suggested that the pan-genome approach is superior to previously published methods in several datasets and across different MTB lineages, as its characteristics allow the comparison of a high number of diverse samples in one analysis [18].

TB transmission assessment using newly developed methods: Diderot and colleagues [29], developed a new Bayesian based method for inferring TB transmission called Transpyloric. The method uses a Susceptible, Infected, removed (SIR) epidemiological model assuming that the transmission bottleneck is complete (only a single pathogen variant is transmitted from infector to susceptible per transmission event), and that all cases comprising an outbreak have been sampled. In their approach, they constructed a timed phylogenetic tree using Bayesian evolutionary analysis by sampling trees (using BEAST software), which becomes an input for Transpyloric to infer transmission networks via Markov Chain Monte Carlo simulation. The robustness of the method is based on the fact that it takes into consideration within-host diversity, which has been shown to be significant for MTB in some settings [65]. The main difference from a typical BEAST output is that Transpyloric infers a transmission tree that defines specific transmission events or new infections indicated by red stars and a change in branch colour (Figure 2). The method also produces a transmission tree from the original phylogeny with vertical arrows representing the occurrence of transmission from person-to-person [27,33]. The difference between transmission tree and phylogenetic tree is that a transmission tree can be inferred from a phylogeny while accounting for within-host genetic diversity by colouring the branches of a phylogeny according to which host those branches were in [29].

Some studies used SNP threshold combined with other new methods, such as phylogenetic modelling and a Transpyloric approach to assess and confirm recent TB transmission. For example, Yang et al [27], used both SNP threshold (≤ 10 SNPs) and Transpyloric methods to assess TB transmission dynamics among internal migrants in Shanghai, China. The study quantified the relative importance of latent TB importation with that of local transmission by comparing the approximated time of transmission (using Transpyloric) with reported time of arrival in Shanghai. Based on the combination of these methods (SNP threshold and Transpyloric), the authors found that the primary mechanism driving local incidence of TB in Shanghai was locally transmitted between both migrants and residents [28]. Additionally, Arabian et al [28], used SNP threshold (< 6 SNPs and 6 - 12 SNPs) and Transpyloric methods to assess TB transmission among immigrants and local individuals in Norway. The study estimated that about 25% of the patients contracted TB (via direct transmission) after having lived in Norway for almost 20 years.

Eldholm et al [30], developed a new method using epidemiological modelling approach to explore the transmission of MDR-TB and HIV co-infection. The overall aim of the study was to apply a newly developed method to explore the impact of HIV on TB transmission, MTB evolution and whether HIV co-infection accelerates drug resistance evolution. The authors considered a SEIR (susceptible, exposed, infected, removed) epidemiological model and assumed that within-host diversity happens at a constant rate, α as applied in Transpyloric method [29]. The SEIR epidemiological model assumed in the development of the method implies that there is random mixing between the individuals, with every infectious individual being equally likely to infect any susceptible individual.

The Eldholm et al [30], method described above is an extension of the Transpyloric approach (by Diderot et al [29], with a timed phylogenetic tree from BEAST used as input for SEIR model simulation. The outputs of the tool are transmission trees (transmission chains) with different branch colours (Figure 3), which denote direct transmission events from person-to-person as defined in Transpyloric method [29]. The main difference between the two methods is that Transpyloric uses SIR model (assuming that susceptible individuals move directly to disease after acquiring infection), while the method applied by Eldholm et al [30], applies a SEIR model (assuming that after acquiring infection, susceptible individuals move to exposed or latent state before showing disease symptoms) simulation. The additional of exposed (E) state is important as it indicates a transition or incubation period from infection to disease, which reflect the real world of infectious disease. Limitations and advantages of the Eldholm et al method is similar to those described in Transpyloric method as both depend on a time calibrated phylogeny from other software, such as BEAST as input and use similar assumptions, including within-host diversity and other input parameter values. Compared to other traditional methods, these two use transmission trees with branch colours and vertical arrows to better clarify TB transmission from person-to-person. A major advantage of these methods is that they infer the direction of transmission from the data.

Stimson et al [31], developed a new method, to identify whether the genomes of two MTB isolates are part of a cluster. The method is based on a probabilistic approach that uses the molecular clock rate (SNPs per genome per year), transmission rate (transmissions per year) and transmission cut-off (defined

as the cut-off level for the transmission method) to define clusters. Clock rate is the substitution rate, defined as the rate of accumulation of changes in a lineage which depends on both the mutation rate and effects of natural selection [31]. In the development of the method, various assumptions were made. For example, the authors assumed that the population from which samples are drawn is homogeneous unless it is otherwise stated, and that transmission is equally likely between hosts irrespective of factors such as HIV and other co-morbidities. In their approach, they compared their tribalistic method and the SNP threshold method by forming clusters using a range of SNP thresholds and transmission cut-offs. In the comparison of the two methods, authors suggested that the method developed in their study is at least as good at identifying direct transmissions within an outbreak as the SNP threshold method, and typically performs at least slightly better. An advantage of Stimson et al method compared to other methods is that it is more flexible to handle SNPs with different substitution process, variability in the substitution and transmission processes, and it is capable to handle more epidemiological data, such as spatial [31]. One of the weaknesses of the probabilistic method developed by Stimson et al [31], as compared to Bayesian method by Diderot et al [29], is that it does not consider within-host diversity to capture pathogen heterogeneity.

Marker et al [32], and colleagues assessed TB transmission based on SNP distances by determining the number of isolates per isolate that were in a range of ≤ 10 SNPs referred to as transmission index. They used 10 SNP threshold to infer the number of recently linked cases that considered within a 10-year time period. In the implementation of their method, authors aimed to link each isolate with a continuous parameter which reflects the number of recently linked cases as transmission networks. The networks reflected with a minimum spanning tree which allows the visualization of super-spreaders. They assumed that an isolate with a high transmission index might well be linked to a patient that infected multiple secondary cases. The main difference from SNP threshold approach is that transmission index uses a number of isolates within a maximum range of 10 SNPs to infer the number of recently linked cases, assuming that an isolate with a high SNP distance might well be linked to a super-spreader. Advantage of the transmission index-based method by Marker et al is that it has the potential to indicate transmission hotspots within an outbreak scenario and it is independent from a phylogenetic clade definition, which might be difficult to assign due to the close genetic relationship in TB outbreaks and low bootstrap values for small sub-groups at the tips of a tree. Limitation of the transmission index approach is that it is not well stated which software or packages used for SNP distances identification and how spanning tree for transmission networks were constructed.

Discussion

It is generally believed that the genetic distance between paired isolates (number of SNPs that differ between two sequences) can be used to assess TB transmission [13,19,31]. The method assumes that a genetic distance between strains that is below a specified number of SNPs suggests recent transmission. Although it has been widely applied to measure recent TB transmission, SNP thresholds vary substantially across studies and it is not well understood which value or range of SNP thresholds should be used to infer TB transmission and how such thresholds might vary across settings [31,42]. While SNP-threshold based methods are technically easy to apply

and make intuitive sense, they could suggest transmission incorrectly in high incidence settings with endemic strains but no epidemiological links [53]. For example, one study [7], from a high incidence TB setting found that several TB patients with genetic distances of ≤ 5 SNPs lacked epidemiological links, indicating casual transmission or missing source cases.

While most studies described here applied the SNP threshold method without including drug resistance conferring mutations, studies [19-26] included drug resistance mutations in as an additional criterion to infer TB transmission. In these studies, it was assumed that strains that emerged from the same monophyletic group would have smaller genetic distance and the same drug resistance conferring mutations. However, mixed infection, antibiotic treatment pressure, and the convergent evolution of common drug resistance mutations complicate the use of drug resistance mutations to infer transmission, particularly in high TB burden settings. Strains that are genetically similar enough to be included in the same cluster may have both common and different drug resistance conferring mutations, implying the presence of both primary and acquired resistance [35,53]. For example, studies indicated the presence of common mutations conferring resistance to first-line drugs amongst all clustered isolates but different mutations resistance to second-line drugs, suggesting the occurrence of both primary and acquisition resistance [8,69].

Newer methods based on Bayesian methods together with epidemiological modelling [29,30], probabilistic approaches [31], or transmission index [32], have been applied to assess TB transmission. While these methods provide a more robust assessment of TB transmission, and in the case of Bayesian methods, take genetic heterogeneity into account, they are more difficult technically to implement. Moreover, the Bayesian and epidemiological methods rely on a BEAST analysis, which is dependent on having a large enough genetic diversity within the sample set. Two studies discussed here applied both SNP-threshold and Bayesian methods to assess transmission, both indicating the SNP threshold method performed better. While the newly developed Bayesian [29], and probabilistic methods [31], construct transmission trees and clusters to describe recent TB transmission, some of these methods make specific assumptions that might not be relevant to all settings and moreover, are technically difficult to implement. Compared to other methods, the probabilistic method by Stimson et al is more flexible to handle SNPs with different substitution process, and it is capable to handle more epidemiological data, such as spatial. Of newly developed methods, Transpyloric method has been applied in several studies, such as [27,28], as it incorporates within-host diversity and produces informative transmission trees with stars and vertical arrows, indicating the occurrence of transmission events from person-to-person. The main limitation of the Transpyloric based methods is that they assume that all outbreak cases have been sampled and sequenced and that the outbreak has reached its end [33]. In reality, all outbreak cases cannot be sampled as some cases may not be reported to health care. The transmission index-based method by Marker et al [32], aimed to identify the number of recently linked cases as transmission networks and visualisation of super-spreaders in a spanning tree. This method assumed that an isolate with a high transmission index (≤ 10 SNP distances) might well be linked to a patient that infected multiple secondary cases. Advantage of the transmission index-based method is that it has the potential to indicate transmission hotspots within an outbreak. However, software

or packages used in this method for SNP distances identification and construction of spanning tree for transmission networks are not clarified for utilization in other studies.

In addition to the specific approach used to assess clustering and transmission, underlying bioinformatics pipelines can impact inference of transmission [61]. For example, almost all studies discussed here used a single MTB H37Rv strain as a reference genome for read mapping prior to variant calling. However, one study [18] suggested that since the mutation rate of MTB is very low and stable, this approach may result in limited resolution because it does not take every detectable difference into account, and suggests using a pangenome approach that may detect more variants. In addition to the choice of reference genome, another study suggested that the choice of variant calling can also influence the number of SNPs detected, leading to conflicting transmission inferences, and concluded that measurements of genetic distance and phylogenetic structure depend on variant calling [61]. Here, moves to standardise the TB WGS pipeline could be useful.

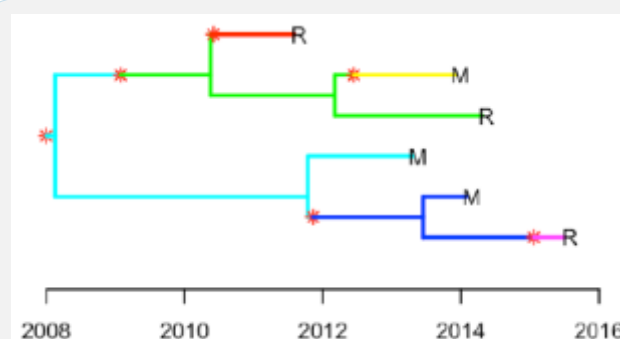


Figure 2: Transmission tree based on MCMC output using Transpyloric. The star and the change of colour represent the occurrence of transmission events. The letters M and R indicate TB cases.



Figure 3: Transmission chains annotated in the timed phylogenetic tree. Red colour highlights strains linked by transmission events from a common ancestor. Branches in magenta indicate subsequent transmission from a secondary case. Figures 2 and 3 are illustrative example of transmission tree outputs by Transpyloric and epidemiological approaches.

In addition to bioinformatics, all WGS based approaches to inferring transmission could face the common biases, including how to disentangling genetic heterogeneity due to different sampling culturing processes prior to WGS and data analysis pipelines in different environments also remains a challenge [31,42,54]. The relatively low genetic diversity (compared to other microbes) is also problematic for transmission models for MTB [68].

Epidemiological contact tracing, based on patient interviews is an important component of investigating TB transmission, and could be combined with WGS [17], or used to validate genotypic methods of assessing transmission. However, the majority of studies reviewed used WGS to infer the possibility of TB transmission without contact tracing, particularly those conducted in high TB burden settings. While contact tracing might be of limited use in high burden settings to interrupt TB transmission, specially designed studies that apply this technique to develop and verify tools to apply to WGS to infer transmission may be useful.

Competing interests: The author declares no competing interests.

References

- Chatterjee A, Nilgiriwala K, Saranat D, Rodrigues C, Mistry N. Whole genome sequencing of clinical strains of Mycobacterium tuberculosis from Mumbai, India: A potential tool for determining drug-resistance and strain lineage. *Tuberculosis*. 2017; 107: 63-72.
- Guerra-Assuno JA, Crampin AC, Houben RMGJ, Mzembe T, Mallard K, et al. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. *Elife*. 2015; 4: 05166.
- Casali N, Broda A, Harris SR, Parkhill J, Brown T, et al. Whole genome sequence analysis of a large isoniazid-resistant tuberculosis outbreak in London: A retrospective observational study. *PLOS medicine*. 2016; 13(10): 1002137.
- Shah NS, Auld SC, Brust JC, Mathema B, Ismail N, et al. Transmission of extensively drug resistant tuberculosis in South Africa. *New England Journal of Medicine*. 2017; 376(3): 243-253.
- Nelson KN, Shah NS, Mathema B, Ismail N, Brust JC, et al. Spatial Patterns of Extensively Drug-Resistant Tuberculosis Transmission in KwaZulu-Natal, South Africa. *The Journal of infectious diseases*. 2018; 218(12): 1964-1973.
- Glynn JR, Guerra-Assuno JA, Houben RM, Sichali L, Mzembe T, et al. Whole genome sequencing shows a low proportion of tuberculosis disease is attributable to known close contacts in rural Malawi. *PLOS one*. 2015; 10(7): 0132840.
- Luo T, Yang C, Peng Y, Lu L, Sun G, et al. Whole-genome sequencing to detect recent transmission of Mycobacterium tuberculosis in settings with a high burden of tuberculosis. *Tuberculosis*. 2014; 94(4): 434-440.
- Wollenberg KR, Desjardins CA, Zalutskaya A, Slodovnikova V, Oler AJ, et al. Whole genome sequencing of Mycobacterium tuberculosis provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. *Journal of clinical microbiology*. 2017; 55(2): 457-469.
- Bjorn-Mortensen K, Soborg B, Koch A, Ladefoged K, Merker M, et al. Tracing Mycobacterium tuberculosis transmission by whole genome sequencing in a high incidence setting: A retrospective population-based study in East Greenland. *Scientific reports*. 2016; 6: 33180.
- Roetzer A, Diel R, Kohl TA, Rckert C, Nbel U, et al. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: A longitudinal molecular epidemiological study. *PLoS medicine*. 2013; 10(2): 1001387.
- Dheda K, Limberis JD, Pietersen E, Phelan J, Esmail A, et al. Outcomes, infectiousness, and transmission dynamics of patients with extensively drug-resistant tuberculosis and home-discharged patients with programmatically incurable tuberculosis: A prospective cohort study. *The lancet Respiratory medicine*. 2017; 5(4): 269-281.
- Senghore M, Otu J, Witney A, Gehre F, Doughty EL, et al. Whole-genome sequencing illuminates the evolution and spread of multidrug-resistant tuberculosis in Southwest Nigeria. *PLOS one*. 2017; 12(9): 0184510.
- Lalor MK, Casali N, Walker TM, Anderson LF, Davidson JA, et al. The use of whole-genome sequencing in cluster investigation of a multidrug-resistant tuberculosis outbreak. *European Respiratory Journal*. 2018; 51(6): 1702313.
- Wyllie DH, Davidson JA, Smith EG, Rathod P, Crook DW, et al. A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying Mycobacterium tuberculosis transmission: A prospective observational cohort study. *Biomedicine*. 2018; 34: 122-130.
- Holt KE, McAdam P, Thai PVK, Truong NTT, Ha DTM, et al. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the Saw Beijing variant in Vietnam. *Nature genetics*. 2018; 50(6): 849-856.
- Phelan JE, Lim DR, Mitarai S, de Sessions PF, Tujan MAA, et al. Mycobacterium tuberculosis whole genome sequencing provides insights into the Manila strain and drug-resistance mutations in the Philippines. *Scientific reports*. 2019; 9(1): 1-6.
- Alaridah N, Hallbck ET, Tngrot J, Winqvist N, Sturegrd E, et al. Transmission dynamics study of tuberculosis isolates with whole genome sequencing in southern Sweden. *Scientific reports*. 2019; 9(1): 1-9.
- Jandrasits C, Krger S, Haas W, Renard BY. Computational Pan-genome Mapping and pairwise SNP-distance improve Detection of Mycobacterium tuberculosis Transmission Clusters. *PLOS Computational Biology*. 2019; 15(12).
- Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study. *The Lancet infectious diseases*. 2018; 13(2): 137-146.
- Walker TM, Lalor MK, Broda A, Ortega LS, Morgan M, et al. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: An observational study. *The Lancet Respiratory Medicine*. 2014; 2(4): 285-292.
- Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nature genetics*. 2014; 46(3): 279.
- Yang C, Luo T, Shen X, Wu J, Gan M, et al. Transmission of multidrug-resistant Mycobacterium tuberculosis in Shanghai, China: A retrospective observational study using whole-genome sequencing and epidemiological investigation. *The Lancet Infectious Diseases*. 2017; 17(3): 275-284.
- Clark TG, Mallard K, Coll F, Preston M, Assefa S, et al. Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. *PLOS one*. 2013; 8(12): 83012.

24. Tyler AD, Randell E, Baikie M, Antonation K, Janella D, et al. Application of whole genome sequence analysis to the study of *Mycobacterium tuberculosis* in Nunavut, Canada. *PLOS one*. 2017; 12(10): 0185656.
25. Meehan CJ, Moris P, Kohl TA, Peerska J, Akter S, et al. The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology. *EbioMedicine*. 2018; 37: 410-416.
26. Walker TM, Merker M, Knoblauch AM, Helbling P, Schoch OD, et al. A cluster of multidrug resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: A molecular epidemiological study. *The Lancet Infectious Diseases*. 2018; 18(4): 431-440.
27. Yang C, Lu L, Warren JL, Wu J, Jiang Q, et al. Internal migration and transmission dynamics of tuberculosis in Shanghai, China: An epidemiological, spatial, genomic analysis. *The Lancet Infectious Diseases*. 2018; 18(7): 788-795.
28. Ayabina D, Ronning JO, Alfsnes K, Debech N, Brynildsrud OB, et al. Genome-based transmission modelling separates imported tuberculosis from recent transmission within an immigrant population. *Microbial genomics*. 2018; 4(10).
29. Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Molecular biology and evolution*. 2014; 31(7): 1869-1879.
30. Eldholm V, Rieux A, Monteserin J, Lopez JM, Palmero D, et al. Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. *Elife*. 2016; 5: 16644.
31. Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, et al. Beyond the SNP threshold: Identifying outbreak clusters using inferred transmissions. *Molecular biology and evolution*. 2019; 36(3): 587-603.
32. Merker M, Barbier M, Cox H, Rasigade JP, Feuerriegel S, et al. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *Elife*. 2018; 7: 38200.
33. Didelot X, Fraser C, Gardy J, Colijn C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. *Molecular biology and evolution*. 2017; 34(4): 997-1007.
34. World Health Organisation. Global tuberculosis report. 2018. WHO/CDS/TB/2018.20
35. Poon AF. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. *Virus evolution*. 2016; 2(2): 031.
36. Gagneux S, Burgos MV, DeRiemer K, Enciso A, Munoz S, et al. Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. *PLoS pathogens*. 2016; 2(6): 61.
37. Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, et al. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science*. 2006; 312(5782): 1944-1946.
38. Kendall EA, Fofana MO, Dowdy DW. Burden of transmitted multidrug resistance in epidemics of tuberculosis: A transmission modelling analysis. *The Lancet Respiratory Medicine*. 2015; 3(12): 963-972.
39. Coscolla M, Gagneux S. December. Consequences of genomic diversity in *Mycobacterium tuberculosis*. In *Seminars in immunology*. 2014; 26(6): 431-444.
40. Bryant JM, Harris SR, Parkhill J, Dawson R, Diacon AH, et al. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: A retrospective observational study. *The Lancet Respiratory Medicine*. 2013; 1(10): 786-792.
41. Guerra-Assuno JA, et al. Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: A whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. *J. Infect. Dis*. 2015; 211: 11541163.
42. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, et al. Whole genome sequencing of *Mycobacterium tuberculosis*: Current standards and open issues. *Nature Reviews Microbiology*. 2019; 17(9): 533-545.
43. Jajou R, de Neeling A, van Hunen R, de Vries G, Schimmel H, et al. Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing: A population-based study. *PLOS one*. 2018; 13(4): 0195413.
44. Trauner A, Borrell S, Reither K, Gagneux S. Evolution of drug resistance in tuberculosis: Recent progress and implications for diagnosis and therapy. *Drugs*. 2014; 74: 1063-1072.
45. Borrell S, Gagneux S. Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis*. 2009; 13: 1456 1466.
46. Cox HS, McDermid C, Azevedo V, et al. Epidemic levels of Drug Resistant Tuberculosis (MDR and XDR-TB) in a high HIV prevalence setting in Khayelitsha, South Africa. *PLOS ONE*. 2010; 5: 13901.
47. Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, et al. Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. *Nature communications*. 2015; 6: 7119.
48. Takiff HE, Feo O. Clinical value of whole-genome sequencing of *Mycobacterium tuberculosis*. *The Lancet Infectious Diseases*. 2015; 15(9): 1077-1090.
49. Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *The Lancet Respiratory Medicine*. 2016; 4(1): 49-58.
50. Nikolayevskyy V, Niemann S, Anthony R, van Soolingen D, Tagliani E, et al. Role and value of whole genome sequencing in studying tuberculosis transmission. *Clinical Microbiology and Infection*. 2019; 25(11): 1377-1382.
51. Yang C, Gao Q. Recent transmission of *Mycobacterium tuberculosis* in China: The implication of molecular epidemiology for tuberculosis control. *Frontiers of medicine*. 2018; 12(1): 76-83.
52. Robinson ER, Walker TM, Pallen MJ. Genomics and outbreak investigation: From sequence to consequence. *Genome medicine*. 2013; 5(4): 36.
53. Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, et al. Interpreting whole genome sequencing for investigating tuberculosis transmission: A systematic review. *BMC medicine*. 2016; 14(1): 21.
54. Nikolayevskyy V, Niemann S, Anthony R, van Soolingen D, Tagliani E, et al. Role and value of whole genome sequencing in studying tuberculosis transmission. *Clinical Microbiology and Infection*. 2019; 25(11): 1377-1382.
55. Khnert D, Coscolla M, Brites D, Stucki D, Metcalfe J, et al. Tuberculosis outbreak investigation using phylodynamic analysis. *Epidemics*. 2018; 25: 47-53.

56. Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah L, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine*. 2011; 364(8): 730-739.
57. Gurjav U, Outhred AC, Jelfs P, McCallum N, Wang Q, et al. Whole genome sequencing demonstrates limited transmission within identified *Mycobacterium tuberculosis* clusters in New South Wales, Australia. *PLOS one*. 2016; 11(10).
58. Mehaffy C, Guthrie JL, Alexander DC, Stuart R, Rea E, et al. Marked microevolution of a unique *Mycobacterium tuberculosis* strain in 17 years of ongoing transmission in a high-risk population. *PLOS one*. 2014; 9(11).
59. Hatherell HA, Didelot X, Pollock SL, Tang P, Crisan A, et al. Declaring a tuberculosis outbreak over with genomic epidemiology. *Microbial genomics*. 2016; 2(5).
60. Sekkides O. Understanding tuberculosis transmission might be the gamechanger we need. *The Lancet Infectious Diseases*. 2019; 19(3): 63.
61. Walter KS, Colijn C, Cohen T, Mathema B, Liu Q, et al. Genomic variant identification methods alter *Mycobacterium tuberculosis* transmission inference. *Bio Rix*. 2019: 733642.
62. Menardo F, Duchêne S, Brites D, Gagneux S. The molecular clock of *Mycobacterium tuberculosis*. *PLOS pathogens*. 2019; 15(9): 1008067.
63. Black PA, De Vos M, Louw GE, Van der Merwe RG, Dippenaar A, et al. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC genomics*. 2015; 16(1): 857.
64. Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, et al. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome biology*. 2014; 15(11): 490.
65. Lieberman TD, Wilson D, Misra R, Xiong LL, Moodley P, et al. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated *Mycobacterium tuberculosis*. *Nature medicine*. 2016; 22(12): 1470.
66. Trauner A, Liu Q, Via LE, Liu X, Ruan X, et al. The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome biology*. 2017; 18(1): 71.
67. Nimmo C, Brien K, Millard J, Grant AD, Padayatchi N, et al. Dynamics of within-host *Mycobacterium tuberculosis* diversity and hetero resistance during treatment. *Biomedicine*. 2020; 55: 102747.
68. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLOS one*. 2009; 4(11).
69. Iøerger TR, Feng Y, Chen X, Dobos KM, Victor TC, et al. The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa. *BMC genomics*. 2010; 11(1): 670.