Review Article

Mycobacterium tuberculosis Next-Generation Whole Genome Sequencing: Opportunities and Challenges

Thato Iketleng [1,2] **Richard Lessells,**[3] **Mlungisi Thabiso Dlamini,**[3] **Tuelo Mogashoa,**[2,4] **Lucy Mupfumi** [2,4] **Sikhulile Moyo,**[2,5] **Simani Gaseitsiwe,**[2,5] **and Tulio de Oliveira**[3]

[1]*College of Health Sciences, School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa*
[2]*Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana*
[3]*KwaZulu-Natal Research Innovation and Sequencing Platform, University of KwaZulu-Natal, Durban, South Africa*
[4]*Department of Medical Laboratory Sciences, University of Botswana, Gaborone, Botswana*
[5]*Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA*

Correspondence should be addressed to Thato Iketleng; iketlengt@gmail.com

Mycobacterium tuberculosis drug resistance is a threat to global tuberculosis (TB) control. Comprehensive and timely drug susceptibility determination is critical to inform appropriate treatment of drug-resistant tuberculosis (DR-TB). Phenotypic drug susceptibility testing (DST) is the gold standard for *M. tuberculosis* drug resistance determination. *M. tuberculosis* whole genome sequencing (WGS) has the potential to be a one-stop method for both comprehensive DST and epidemiological investigations. We discuss in this review the tremendous opportunities that next-generation WGS presents in terms of understanding the molecular epidemiology of tuberculosis and mechanisms of drug resistance. The potential clinical value and public health impact in the areas of DST for patient management and tracing of transmission chains for timely public health intervention are also discussed. We present the current challenges for the implementation of WGS in low and middle-income settings. WGS analysis has already been adapted routinely in laboratories to inform patient management and public health interventions in low burden high-income settings such as the United Kingdom. We predict that the technology will be adapted similarly in high burden settings where the impact on the epidemic will be greatest.

1. Introduction

To curb the emergence and spread of tuberculosis (TB) drug resistance, early detection and effective treatment informed by comprehensive drug susceptibility testing (DST) are vital. It is also important to monitor and understand the development, evolution, biology, and epidemiology of TB drug resistance to inform community level or public health interventions. Molecular methods such as Xpert MTB/RIF (Cepheid, Inc. Sunnyvale, CA, USA) and the line probe assays GenoType MTBDR*plus/sl* (Hain lifescience, GmbH, Nehren, Germany) have considerably increased access to DST and shortened turnaround time to results. However these methods provide resistance information for a limited number of drugs. *Mycobacterium tuberculosis* WGS is an attractive method for both DST to inform treatment decisions and surveillance of drug resistance in high burden settings where capacity for routine resistance testing for everyone with TB is inadequate.

Analysis of WGS data could also be used for epidemiological investigations such as tracing of transmission chains. In this manuscript, we review how *M. tuberculosis* next-generation WGS analysis could impact prediction of drug resistance and in turn clinical management of TB especially in high burden settings. We highlight how analysis of *M. tuberculosis* whole genome sequence data could routinely provide guidance for individual treatment, tracing of transmission chains and continuous drug resistance surveillance for public health interventions. We also look at challenges or barriers to application of *M. tuberculosis* WGS analysis for routine clinical use especially in resource limited TB high burden countries.

2. Opportunities Presented by *M. tuberculosis* Whole Genome Sequencing

2.1. Investigation of Transmission Chain. The advent of next-generation sequencing (NGS) has made WGS a faster, more affordable, and increasingly accessible alternative for molecular epidemiologic studies. The data generated from WGS allows for unparalleled ability to detect genetic variation in *M. tuberculosis*. Analysis of WGS data has led to the reconstruction of *M. tuberculosis* phylogeny and this has improved our understanding of the global distribution of *M. tuberculosis* [1].

WGS has been used to answer questions about TB transmission and will, in the near future, become the routine method for *M. tuberculosis* typing because it has superior resolution to the conventional typing methodologies. Studies have shown that relatedness of *M. tuberculosis* genomes can be estimated by comparing Single Nucleotide Polymorphism (SNPs) differences between the isolates. Isolates with the smallest number of SNPs differences or shortest SNP distance would be linked, possibly representing a transmission event or cluster [2–6]. A cut-off of five SNPs or fewer for linked transmission is widely used; however a maximum of three SNPs has been suggested to represent human-to-human transmission [2–6]. Walker *et al* (2012) looked at SNP differences between epidemiologically linked pairs of isolates in the United Kingdom and none of the linked pairs exceeded five SNP difference and thus came up with the five SNP cut-off [6].

The disadvantage of this approach is that it relies on epidemiologically linked pairs. In most high burden settings especially, epidemiological links are often not available and this would make this approach problematic. Another approach has been to look at SNP differences between unlikely transmission pairs. Using this approach, a SNP difference of 0-1 has been used to define clusters based on the fact that all unlikely transmission pairs had more than two SNP differences [3]. In instances where an estimate of how long ago the transmission occurred was known, mutation rates have been used to determine a SNP cut-off [7]. This approach assumes that the mutation rate is constant over time, which is not always the case. A mutation rate of 0.003 SNPs per day has been used to determine a cut-off of ≤10 SNPs as confirmatory of transmission [7].

Analysis of whole genome data, such as clustering of SNPs, has given us an insight into the transmission dynamics as well as intra- and interpatient variations at play during outbreaks of *M. tuberculosis* [2, 3, 5, 6].

Whole genome data used in conjunction with social network analysis enabled identification of socioenvironmental factors as a driver of an outbreak in British Columbia, Canada [2]. Through detailed whole genome data analysis of another outbreak in San Francisco, cases with no obvious epidemiological connection were linked, and microevolutionary events were identified that helped to define the likely chain of transmission [8]. The resolution of WGS is superior to mycobacterial interspersed repetitive units-variable number tandem repeat (MIRU-VNTR) as evidenced by the ability of WGS to differentiate lineages of *M. tuberculosis* with identical MIRU-VNTR genotypes [2, 3, 5, 6]. WGS also allows inference about direction of transmission between the cases to be made using SNP distances, even in the absence of epidemiological data [3, 6]. This is crucial because TB outbreaks often occur in communities where epidemiological data is difficult to collect. The superior resolution of WGS has enabled us to differentiate between relapse and reinfection [9]. This is crucial for accurately evaluating treatment and prevention programmes. WGS has been used to understand transmission dynamics in a high burden setting, proof that this method could indeed have the biggest impact yet in determining *M. tuberculosis* transmission dynamics in high burden settings [7, 10].

2.2. Identification of Mixed Infections. Mixed *M. tuberculosis* infections are described as TB disease caused by more than one distinct *M. tuberculosis* strain. Traditionally they are identified based on at least two distinct patterns on MIRU-VNTR results. *M. tuberculosis* next-generation WGS analysis using heterozygous base calls can provide better resolution of mixed infections. However, studies have used different definitions of mixed infections based on heterozygous base calls. Presence of more than 80 and 140 heterogeneous base calls in one sample has been used to define mixed infections [7, 9]. Other studies have used classifications such as mixed base call where 38% of reads support the variant as mixed infection [3]. Mixed infections are more common in high burden settings, with 10% to 20% reported among TB patients [11, 12].

Higher rates of mixed infections have been reported among retreatment cases compared to new cases [12]. They have also been associated with poor outcomes especially where the distinct strains have different drug resistance patterns and where there is an underlying immune-suppression caused by HIV infection [13, 14]. Mixed infections may have an impact on diagnosis as evidenced by the low sensitivity (80%) of the Xpert assay for rifampicin resistance on mixed infections compared to 93% on homogenous infections [11]. Identification of mixed infections is also critical for evaluating effectiveness of tuberculosis interventions. Changes in drug resistance patterns of isolates can sometimes be explained by existence of isolates with different patterns at the same time [15]. Instances of mixed infections by strains with different DST profiles may easily be misclassified as cases of acquired resistance upon unmasking of the resistance following treatment, misinforming interventions.

2.3. Prediction of Drug Resistance and Understanding of Mechanisms of Drug Resistance. Unlike other molecular methods that typically target specific genes for determination of drug resistance, WGS allows for the interrogation of the entire *M. tuberculosis* genome for mutations conferring drug resistance. Mutations occurring outside the genes known to be associated with drug resistance can be identified from TB whole genomes. The likelihood of finding novel drug resistance conferring mutation is thus increased. Compensatory mutations, that is, mutations not directly involved in drug resistance but rather compensate for the fitness cost of drug resistance mutations, can be identified from whole

genomes [16]. Strains with such compensatory mutations will have high fitness despite also harbouring drug resistance mutations [16]. It is through analysis of mutations in whole genomes of isolates from extensively drug-resistant tuberculosis (XDR-TB) outbreaks that we know that not all outbreaks are caused by clonal expansion of drug resistant strains, but rather some outbreaks are caused by acquired drug resistance that in many isolates appear to have been acquired independently [17, 18].

Although largely done retrospectively, WGS for determination of drug resistance has shown good concordance with conventional DST, with shorter turnaround times especially when done from early cultures [19, 20]. In one retrospective analysis, individualised drug regimens for multidrug-resistant tuberculosis (MDR-TB) and XDR-TB constructed on the basis of WGS were in close agreement with those constructed from phenotypic DST data [21]. Importantly, drug regimes constructed on the basis of WGS did not feature any drug to which phenotypic resistance was indicated, but rather WGS predicted more resistance to drugs such as ethambutol [21]. This matters because WGS would have ruled out drugs that might have been included in a treatment regimen, if decisions were based purely on information from phenotypic DST. By influencing the composition of the treatment regimen, this could lead to more effective regimens and could also reduce toxicity from unnecessary drugs.

There was strong evidence that WGS outperformed Xpert and line probe assays in terms of appropriate regimen selection [21]. Isolates with low-level resistance by WGS were susceptible by phenotypic DST probably because the critical concentrations set for DST were too high [21]. Novel or poorly defined mutations identified in phenotypically susceptible isolates were difficult to interpret [21]. This highlights the gap in knowledge of the genetic determinants of *M. tuberculosis* drug resistance that still need to be addressed. A comparison of WGS with Hain line probe assays and phenotypic DST for species identification and resistance determination for first-line drugs demonstrated comparable processing time for WGS [22]. The turnaround time for results was comparable to phenotypic DST when WGS was newly introduced into the laboratory workflow; however after successful incorporation into routine laboratory workflow, WGS results were available nine days earlier than phenotypic DST [22]. The relatively high rates of isolates with insufficient data for drug resistance determination highlight the need to further improve WGS to reduce these rates. The turnaround time for WGS results could potentially be reduced even further if efforts to apply WGS directly to clinical samples such as sputum without the need for culture are successful.

2.4. Whole Genome Sequencing for Continuous Drug Resistance Surveillance. In order to gauge the effectiveness of strategies to control *M. tuberculosis* drug resistance globally, accurate data on the occurrence of drug resistance is critical. WHO recommends routine DST for all TB tuberculosis patients to provide continuous surveillance of drug resistance. However most high burden countries still rely on epidemiological surveys conducted at best every five years. This is because of lack of capacity for routine DST for all TB

patients; in most instances molecular methods such as the Xpert MTB/RIF are the only available methods for routine DST for suspected MDR-TB cases. Some countries have the capacity to only carry out first-line DST and have to send out samples for second-line DST to laboratories outside of the country. WGS analysis for routine drug resistance surveillance for all TB patients is an attractive avenue. However, this would have a prohibitive cost at the moment given that most of the high burden countries struggle to even afford DST for suspected drug resistant cases.

The investment in routine DST for all TB patients using WGS may not seem cost-effective in the short term. However, studies on the impact and cost effectiveness of routine WGS in high burden setting are needed to determine the feasibility of WGS in this setting. WGS analysis yields susceptibility results for both first- and second-line drugs; the data could also be an invaluable resource for understanding the epidemiology of TB in the respective countries. WGS has been used successfully to complement a drug resistance survey in Uganda where a small proportion of the isolates that were phenotypically resistant to isoniazid and rifampicin were analysed using WGS to try and understand the extent of resistance [23].

3. Challenges for Implementation of Whole Genome Sequencing for Routine Clinical Use

3.1. DNA Extraction for WGS and Quality Control. Methods that have traditionally been used for extraction of DNA from *M. tuberculosis* can in principle be employed to extract DNA for WGS. The DNA would then need to be checked for quality and concentration usually using either quantitative PCR, a spectrophotometer such as the Qubit machine, or Agarose gel electrophoresis to determine whether the quality of the extracted DNA meets the minimum standard set for the particular instrument. However, before choosing a DNA extraction method it is important to take into consideration the library preparation method or kit being used as the kit usually prescribes the minimum input DNA required for library preparation. Table 1 summarises some of the available library preparation kits, NGS platforms with which they are compatible, and the minimum input DNA required for library preparation.

3.2. The Need for Culture for Whole Genome Sequencing. *M. tuberculosis* WGS has traditionally relied on growing the organism in culture. Culturing served two important purposes critical to the success of sequencing. Firstly, it ensured selective growth of *M. tuberculosis* and secondly, the growth allowed extraction of sufficient quantities of DNA for sequencing. However, this affects the turnaround time of WGS. Studies have also shown that culture methods may enrich certain strains of *M. tuberculosis* therefore affecting the population structure or clonal complexity of *M. tuberculosis* [34, 35]. The different kinds of culture media and growth conditions would therefore potentially affect the ability to detect mixed infections. This is important because, in terms of patient management, we might miss drug resistance

TABLE 1: Available library preparation kits, compatible platforms, and minimum input DNA required.

Library Preparation kit	Platform/System Compatibility	Minimum input DNA required
Nextera DNA Flex, Nextera XT, Nextera Mate Pair	All Illumina® platforms	1ng
Thermo Scientific MuSeek	Iron Personal Genome Machine™ and Ion Proton™ systems	100ng
Thermo Scientific MuSeek	Illumina® MiSeq and Illumina® HiSeq	50ng
Qiagen QIAseq	All Illumina® platforms	1ng
Ligation Sequencing Kit 1D, $1D^2$ Sequencing Kit,	All Oxford Nanopore Sequencing Technologies platforms	1000ng
Rapid Sequencing Kit	All Oxford Nanopore Sequencing Technologies platforms	400ng
Ion Xpress™ Plus Fragment	Ion OneTouch™ and Iron Personal Genome Machine™	100ng

and, in terms of molecular epidemiology, we might miss transmission events.

Recently researchers have started exploring ways to bypass the long process of growing the bacteria in culture as the starting point for WGS. Next-generation sequencing of *M. tuberculosis* from DNA extracted directly from sputum without culture, targeted amplification, or capture is challenging primarily because sputum is a complex mixture of human cells, mycobacterial cells, and oral/nasopharyngeal bacterial cells. DNA extracted directly from sputum without targeted amplification or capture will contain not only mycobacterial DNA, but also oral/nasopharyngeal bacterial DNA and large amounts of human/host DNA depending on the success of the decontamination of the sputum before the extraction. Owing to the nonspecific nature of the sequencing primers used in NGS, only a low number of reads will be from *M. tuberculosis* in cases where the concentration of *M. tuberculosis* DNA is low compared to either the host or oral bacterial DNA, as is often the case in sputum. Therefore, these approaches have been successful in sequencing *M. tuberculosis* albeit with very low coverage and depth limiting the utility of the data for downstream analysis (Table 2) [26].

To enable successful NGS with high coverage and depth, enrichment or capture methods capable of selectively capturing or enriching for *M. tuberculosis* DNA have been developed to get the DNA at concentrations and purity levels suitable for sequencing. SureSelect[XT] Target Enrichment System (Agilent Technologies, Inc. Santa Clara, USA) is one such system (Table 2). The system has allowed high quality *M. tuberculosis* WGS without the need for culture [24, 27]. Enrichment systems could play role in ensuring that high quality WGS can be done in time to inform patient management options. However, these systems are very costly and add to the already relatively expensive WGS process. The laboratory workflows involved in these enrichment systems are also very complex and thus may remain out of reach for most programmes in regions hardest hit by TB.

A somewhat simpler, low cost approach involves depletion of the human cells or DNA using a saline wash before continuing with *M. tuberculosis* DNA extraction using ethanol precipitation (Table 2) [25]. In principle this is similar

to the differential lysis protocol that has also been used successfully. The differential lysis protocol involves lysing human cells and then using DNase treatment to remove the human/host DNA followed by *M. tuberculosis* DNA extraction using a commercial kit [26]. These simpler methods tick all the boxes of an ideal extraction method; the workflow is simple and does not add much to the cost of sequencing and using the depletion of human cells approach at least, high quality sequences were generated in a clinically relevant time frame. The differential lysis protocol however has been less of a success because the proportion of reads mapping against the human genome was as high as 99% in some cases despite the attempt during DNA extraction to deplete human cells or DNA [26].

Despite the progress that has been made in developing all these methods to enable sequencing directly from sputum, we still only get good results in samples with high bacillary burden as depicted by decreasing depth and coverage with decreasing bacillary load [24]. Yet the people who could benefit most from rapid WGS-based DST are those at highest risk of morbidity and mortality, such as HIV-positive people, who are more likely to have paucibacillary disease making current methods unsuitable for such cases. Therefore much more work is needed to optimise protocols for DNA extraction and WGS directly from sputum, so that the technology can be widely applied where it is most needed.

3.3. Incompleteness of the Understanding of the Genetic Basis of M. tuberculosis Drug Resistance. The use of sequencing for *M. tuberculosis* drug resistance determination is limited by our current knowledge and understanding of the characterised resistance associated mutations. The current library of these characterised resistance mutations probably does not include all mutations potentially associated with the resistant phenotypes of *M. tuberculosis*. Data linking genotypic-phenotypic resistance is relatively complete for some first-line drugs such as rifampicin and isoniazid but still incomplete especially for new drugs (Table 3). In a large analysis of more than 10,000 isolates from six continents, phenotypic resistance to the first-line drugs rifampicin, isoniazid, ethambutol and pyrazinamide was predicted by WGS with sensitivities ranging from 91.3% to 97.5%, and susceptibility was predicted with specificities ranging from

TABLE 2: Summary of methods developed for DNA extraction directly from sputum samples for *M. tuberculosis* next-generation whole genome sequencing.

Reference	Sample	Sample preparation for sequencing	Sequencing Platform	Whole genome sequencing success rate (%)	% Coverage; Average Depth of coverage
Doyle et al., 2018 [24]	Sputum	SureSelectXT target enrichment system (Agilent technologies)	Illumina MiSeq or NextSeq sequencer	74.4%	>85%; 25X to 300X
Votinseva et al., 2017 [25]	Smear positive sputum	Saline wash, MolYsis Basic5 kit (Molzym, Germany) for depletion of human cells/DNA, Ethanol precipitation method.	Illumina MiSeq sequencer	62%	>90%; >12X
Doughty et al., 2014 [26]	Smear positive sputum	Differential lysis protocol followed by DNA extraction using the NucleoSpin Tissue-Kit (Machery-Nagel, Duren, Germany)	Illumina MiSeq sequencer	87%	20% to 99%; 0.002X to 0.7X
Brown et al., 2015 [27]	Smear positive sputum	SureSelectXT target enrichment system (Agilent technologies)	Illumina MiSeq sequencer	83%	90%; >20X

TABLE 3: Summary of the percentage phenotypic resistance that can be explained by genetic mutations at the locus of interest.

Drug	Locus	Percentage resistance accounted for by mutations	References
Rifampicin	*rpoB*	96%	[28]
Isoniazid	*katG, inhA, ahpC-oxyR*	84%	[29]
Ethambutol	*embB*	70%	[30]
Streptomycin	*rrs, rpsL*	60%-70%	[28]
Pyrazinamide	*pncA*	72%-99%	[31]
Fluoroquinolones	*gyrA*		[32]
		87% (Moxifloxacin)	
		83% (Ofloxacin)	
Aminoglycosides	*rrs*		[33]
		70%-80% (Capreomycin and Amikacin)	
		60% (Kanamycin)	

93.6% to 99.0% [36]. In specific analysis of the isolates with definite phenotypic susceptibility to all four first-line drugs, WGS correctly predicted pansusceptibility in 97.9%. [36]. However, it should be noted that these estimates of predictive accuracy were based on isolates with complete phenotypic and genotypic profiles.

The susceptible phenotypes with resistance conferring mutations appeared at least based on the determined predictive performances of the mutations they harboured, to be true resistant isolates [36]. Although resistance-conferring mutations were found to be good predictors of resistant phenotypes in this study (over 90% sensitivities for first-line drugs), the mutations were linked to unexpected phenotypes in a minority of cases highlighting the geno-pheno discrepancies that still need to be resolved [36].

The phenotypically resistant yet genetically susceptible discrepancy in isolates may be explained by the fact that we are looking for mutations in genes that have already been associated with resistance, yet these phenotypes may be explained by genetic changes outside of the loci known to be associated with resistance. WGS can be used to add to the body of knowledge and characterize some of the yet to be identified and poorly characterised genetic determinants of drug resistance. To achieve this, novel mutations identified by analysis of WGS data have to be validated with phenotypic and clinical outcome data.

In an effort to contribute towards closing these gaps, a genome wide analysis of 6464 multi and extensively drug resistant *M. tuberculosis* isolates from over 30 countries identified novel resistance associated mutations including small

Indels in *pncA* and large deletions in *katG* [37]. Although further functional characterisation is required to fully understand the role of these mutations in drug resistance, treatment failure, and ultimately their influence on treatment outcome, it is only through such efforts that we can one day understand the genetic determinants of *M. tuberculosis* dug resistance well enough for WGS to play a major role in TB patient management. Studies on the correlation of mutations, known and newly discovered, with treatment outcomes are urgently needed to determine the influence of the different mutations on treatment outcomes in a multidrug treatment regimen background.

3.4. Challenges with Handling the Large Amounts of Data from Whole Genome Sequencing. WGS produces huge amounts of data and therefore requires costly infrastructure for both storage and analysis. The analysis of WGS data also requires specialised bioinformaticians who are usually not available in clinical laboratories. However, the development of easy to use automated analysis pipelines and databases will in time allow people with minimal bioinformatic skills to analyse whole genome sequence data. Good examples of the progress that is being made in this regard are free open access online tools for rapid detection of *M. tuberculosis* drug resistance and lineage specific mutations from raw whole genome sequence data such as TB Profiler, Mykrobe Predictor TB, CASTB, KvarQ, and PhyResSE [38–40]. The tools generate results in an easy to understand format that clinicians can use for patient management [38–40]. Although still prototypes that need further testing and validation for clinical use, they are for now available for research purposes. It is a step in the right direction in terms of sorting out the analysis bottleneck that WGS data creates in clinical laboratories.

4. Conclusion

The future of *M. tuberculosis* WGS lies in the ability to apply the method directly to sputum, as this is the clinical material that is most commonly available. Sequencing directly from sputum samples without the need for culturing would provide a more accurate picture of the population structure of mixed infections. The relative representation of the different strains in mixed infections can be captured without the overgrowth of some strains over others due to favourable conditions of culture. This would better inform treatment and prevention interventions. *M. tuberculosis* WGS from sputum is possible and will be the way to go in the near future. This would significantly reduce the turnaround time for both resistance determinations and provide timely information about transmission dynamics.

The cost of WGS continues to go down with rapid advances to the technology. However, it remains expensive and inaccessible to high burden low-income settings, who would benefit most from the technology. We expect that in the near future *M. tuberculosis* WGS will be adapted in low-income high burden settings for periodic drug resistance surveillance to understand mechanisms of resistance and inform design of cheaper molecular diagnostics. However, if *M. tuberculosis* WGS is to really have an impact on the

epidemic in high burden settings, adaptation into routine laboratory algorithms needs to occur. This adaptation will need to be preceded by upgrades in both laboratory infrastructure and key competencies such as bioinformatics, databases, and software development to provide support and allow proper handling and interpretation of the massive amounts of data generated through WGS. In conclusion, *M. tuberculosis* WGS, especially directly from sputum, will play a crucial role in the fight against the spread of TB as this significantly shortens the turnaround time to results and enables the provision of effective treatment regimen to TB patients who might be harbouring drug resistant strains.

Conflicts of Interest

All authors declare no conflicts of interest.

References

[1] C. Ford, K. Yusim, T. Ioerger et al., "Mycobacterium tuberculosis - Heterogeneity revealed through whole genome sequencing," *Tuberculosis*, vol. 92, no. 3, pp. 194–201, 2012.

[2] J. L. Gardy et al., "Whole-genome sequencing and social-network analysis of a tuberculosis outbreak," *The New England Journal of Medicine*, vol. 364, no. 8, pp. 730–739, 2011.

[3] M. Kato-Maeda et al., "Use of Whole Genome Sequencing to Determine the Microevolution of *Mycobacterium tuberculosis* during an Outbreak," *PLoS ONE*, vol. 8, no. 3, Article ID e58235, 2013.

[4] C. J. Meehan et al., "The relationship between transmission time and clustering methods in Mycobacterium tuberculosis epidemiology," *bioRxiv*, 2018.

[5] A. Roetzer et al., "Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study," *PloS Medicine*, vol. 10, no. 2, Article ID e1001387, 2013.

[6] W. T. Fau, C. L. C. Ip et al., "Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study," *The Lancet Infectious Diseases*, vol. 13, no. 2, pp. 137–146, 2013.

[7] J. Guerra-Assunção, A. Crampin, R. Houben et al., "Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area," *eLife*, vol. 4, 2015.

[8] M. Kato-Maeda, C. Ho, B. Passarelli et al., "Use of whole genome sequencing to determine the microevolution of mycobacterium tuberculosis during an outbreak," *PLoS ONE*, vol. 8, no. 3, Article ID e58235, 2013.

[9] J. M. Bryant et al., "Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study," *The Lancet. Respiratory Medicine*, vol. 1, no. 10, pp. 786–792, 2013.

[10] J. R. Glynn, J. A. Guerra-Assunção, R. M. Houben et al., "Whole genome sequencing shows a low proportion of tuberculosis disease is attributable to known close contacts in rural Malawi," *PLoS ONE*, vol. 10, no. 7, Article ID e0132840, 2015.

[11] N. M. Zetola, S. S. Shin, K. A. Tumedi et al., "Mixed Mycobacterium tuberculosis complex infections and false-negative results for rifampin resistance by genexpert MTB/RIF are associated with poor clinical outcomes," *Journal of Clinical Microbiology*, vol. 52, no. 7, pp. 2422–2429, 2014.

[12] R. M. Warren, T. C. Victor, E. M. Streicher et al., "Patients with Active Tuberculosis often Have Different Strains in the Same Sputum Specimen," *American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 5, pp. 610–614, 2004.

[13] N. M. Zetola, C. Modongo, P. K. Moonan et al., "Clinical outcomes among persons with pulmonary tuberculosis caused by Mycobacterium tuberculosis isolates with phenotypic heterogeneity in results of drug-susceptibility tests," *The Journal of Infectious Diseases*, vol. 209, no. 11, pp. 1754–1763, 2014.

[14] S. S. Shin, C. Modongo, R. Ncube, E. Sepako, J. D. Klausner, and N. M. Zetola, "Advanced immune suppression is associated with increased prevalence of mixed-strain Mycobacterium tuberculosis infections among persons at high risk for drug-resistant TB in Botswana," *The Journal of Infectious Diseases*, vol. 211, no. 3, pp. 347–351, 2015.

[15] A. Van Rie, T. C. Victor, M. Richardson et al., "Reinfection and mixed infection cause changing Mycobacterium tuberculosis drug-resistance patterns," *American Journal of Respiratory and Critical Care Medicine*, vol. 172, no. 5, pp. 636–642, 2005.

[16] I. Comas, S. Borrell, A. Roetzer et al., "Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes," *Nature Genetics*, vol. 44, no. 1, pp. 106–110, 2012.

[17] T. R. Ioerger, Y. Feng, X. Chen et al., "The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa," *BMC Genomics*, vol. 11, no. 670, 2010.

[18] T. R. Ioerger, S. Koo, E. No et al., "Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa," *PLoS ONE*, vol. 4, no. 11, Article ID e7778, 2009.

[19] C. U. Köser, J. M. Bryant, J. Becq et al., "Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis," *The New England Journal of Medicine*, vol. 369, no. 3, pp. 290-291, 2013.

[20] K. Dheda, J. D. Limberis, E. Pietersen 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*, vol. 5, no. 4, pp. 269–281, 2017.

[21] J. Heyckendorf, S. Andres, C. U. Köser et al., "What is resistance? Impact of phenotypic versus molecular drug resistance testing on therapy for multi- and extensively drug-resistant tuberculosis," *Antimicrobial Agents and Chemotherapy*, vol. 62, no. 2, 2018.

[22] T. P. Quan, Z. Bawa, D. Foster et al., "Evaluation of whole-genome sequencing for mycobacterial species identification and drug susceptibility testing in a clinical setting: a large-scale prospective assessment of performance against line probe assays and phenotyping," *Journal of Clinical Microbiology*, vol. 56, no. 2, 2018.

[23] W. Ssengooba, C. J. Meehan, D. Lukoye et al., "Whole genome sequencing to complement tuberculosis drug resistance surveys in Uganda," *Infection, Genetics and Evolution*, vol. 40, pp. 8–16, 2016.

[24] R. M. Doyle, C. Burgess, R. Williams et al., "Direct whole genome sequencing of sputum accurately identifies drug resistant Mycobacterium tuberculosis faster than MGIT culture sequencing," *Journal of Clinical Microbiology*, vol. 56, no. 8, Article ID e00666-18, 2018.

[25] A. A. Votintseva et al., "Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples," *Journal of Clinical Microbiology*, vol. 55, no. 5, pp. 1285–1298, 2017.

[26] E. L. Doughty, M. J. Sergeant, I. Adetifa, M. Antonio, and M. J. Pallen, "Culture-independent detection and characterisation of Mycobacterium tuberculosis and M. africanum in sputum samples using shotgun metagenomics on a benchtop sequencer," *PeerJ*, vol. 2, 2014.

[27] A. C. Brown, J. M. Bryant, K. Einer-Jensen et al., "Rapid whole-genome sequencing of mycobacterium tuberculosis isolates directly from clinical samples," *Journal of Clinical Microbiology*, vol. 53, no. 7, pp. 2230–2237, 2015.

[28] S. H. Gillespie, "Evolution of drug resistance in *Mycobacterium tuberculosis*: clinical and molecular perspective," *Antimicrobial Agents and Chemotherapy*, vol. 46, no. 2, pp. 267–274, 2002.

[29] M. Gegia, N. Winters, A. Benedetti, D. van Soolingen, and D. Menzies, "Treatment of isoniazid-resistant tuberculosis with first-line drugs: a systematic review and meta-analysis," *The Lancet Infectious Diseases*, vol. 17, no. 2, pp. 223–234, 2017.

[30] P. E. A. Da Silva and J. C. Palomino, "Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs," *Journal of Antimicrobial Chemotherapy*, vol. 66, no. 7, pp. 1417–1430, 2011.

[31] M. G. Whitfield, H. M. Soeters, R. M. Warren et al., "A global perspective on pyrazinamide resistance: systematic review and meta-analysis," *PLoS ONE*, vol. 10, no. 7, Article ID e0133869, 2015.

[32] E. Avalos, D. Catanzaro, A. Catanzaro et al., "Frequency and geographic distribution of gyrA and gyrB mutations associated with fluoroquinolone resistance in clinical Mycobacterium tuberculosis isolates: A systematic review," *PLoS ONE*, vol. 10, no. 3, Article ID e0120470, 2015.

[33] S. B. Georghiou, M. Magana, R. S. Garfein, D. G. Catanzaro, A. Catanzaro, and T. C. Rodwell, "Evaluation of genetic mutations associated with *Mycobacterium tuberculosis* resistance to amikacin, kanamycin and capreomycin: a systematic review," *PLoS ONE*, vol. 7, no. 3, Article ID e33275, 2012.

[34] M. Hanekom et al., "Population structure of mixed *Mycobacterium tuberculosis* infection is strain genotype and culture medium dependent," *PLoS ONE*, vol. 8, no. 7, Article ID e70178, 2013.

[35] A. Martín, M. Herranz, M. J. Ruiz Serrano, E. Bouza, and D. García de Viedma, "The clonal composition of Mycobacterium tuberculosis in clinical specimens could be modified by culture," *Tuberculosis*, vol. 90, no. 3, pp. 201–207, 2010.

[36] The CRyPTIC Consortium, The 100.000 Genomes Project et al., "Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing," *The New England Journal of Medicine*, vol. 379, no. 15, pp. 1403–1415, 2018.

[37] F. Coll et al., "Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis," *Nature Genetics*, vol. 50, no. 2, pp. 307–316, 2018.

[38] F. Coll, R. McNerney, M. D. Preston et al., "Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences," *Genome Medicine*, vol. 7, no. 1, 2015.

[39] S. Feuerriegel, V. Schleusener, P. Beckert et al., "PhyResSE: A web tool delineating Mycobacterium tuberculosis antibiotic resistance and lineage from whole-genome sequencing data," *Journal of Clinical Microbiology*, vol. 53, no. 6, pp. 1908–1914, 2015.

[40] V. Schleusener, C. U. Köser, P. Beckert, S. Niemann, and S. Feuerriegel, "Mycobacterium tuberculosis resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools," *Scientific Reports*, vol. 7, no. 1, 2017.