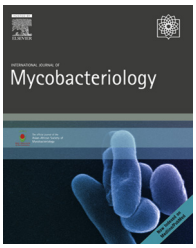
International Journal of Mycobacteriology 5 (2016) 384–391

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/IJMYCO



Review

Diagnosing tuberculosis in the 21st century – Dawn of a genomics revolution?

Christopher Jeanes *, *Justin O'Grady*

Norwich Medical School, University of East Anglia, Norwich Research Park, Norfolk NR4 7TJ, United Kingdom

ARTICLE INFO

Keywords:
Tuberculosis
Diagnostics
Whole genome sequencing
Next generation sequencing
Genomics

ABSTRACT

Tuberculosis (TB) ranks alongside HIV as the leading cause of death worldwide, killing 1.5 million people in 2014. Traditional laboratory techniques do not provide sufficiently rapid results to inform clinicians on appropriate treatment, especially in the face of increasingly prevalent drug-resistant TB. Rapid molecular methods such as PCR and LAMP are vital tools in the fight against TB, however, rapid advances in next generation sequencing (NGS) technology are allowing increasingly rapid and accurate sequencing of entire bacterial genomes at ever decreasing cost, providing unprecedented depth of information. These advances mean NGS stands to revolutionise the diagnosis and epidemiological study of *Mycobacterium tuberculosis* infection. This review focuses on current applications of NGS for TB diagnosis including sequencing cultured isolates to predict drug resistance and, more desirably, direct diagnostic metagenomic sequencing of clinical samples. Also discussed is the potential impact of NGS on the epidemiological study of TB and some of the key challenges that need to be overcome to enable this promising technology to be translated into routine use.

Contents

* Corresponding author.
E-mail addresses: chrisjeanes01@gmail.com (C. Jeanes), Justin.OGrady@uea.ac.uk (J. O'Grady).

Peer review under responsibility of Asian African Society for Mycobacteriology.
http://dx.doi.org/10.1016/j.ijmyco.2016.11.028

Introduction

Tuberculosis (TB) remains an important global health problem with an estimated 9.6 million TB cases and 1.5 million associated deaths worldwide in 2014 [1]. Improved diagnostics and ongoing research are central tenets to the World Health Organisation (WHO) End TB Strategy and the Foundation for Innovative New Diagnostics (FIND) Strategy for Tuberculosis and Lower Respiratory Tract Infections [2,3]. Of significant concern is the rapidly increasing prevalence of multi-drug resistant (MDR-TB, TB resistant to at least rifampicin and isoniazid) and extensively drug resistant TB (XDR-TB, MDR-TB with additional resistance to any fluoroquinolone and one second line injectable agent (amikacin/kanamycin/capreomycin). Annually there are 500,000 new MDR-TB cases including 7.4–12% XDR-TB. Current culture and PCR based diagnostics tests for XDR-TB are not sufficiently rapid and/or accurate to effectively manage patients [2,3]. Whilst molecular methods such as Xpert MTB/RIF® (Cephid, U.S.) have provided significant steps forward in achieving the goals set out by the WHO and FIND, next generation sequencing (NGS) approaches hold great promise for the future diagnosis and management of TB on an individual and population basis [4,5].

This review will outline how NGS, through fast and accurate whole genome sequencing (WGS) and/or shotgun metagenomics sequencing (SMS), holds the potential to transform the diagnosis and management of TB both for the individual patient and as an epidemiological tool.

The MTB genome and drug resistance

Mycobacterium tuberculosis (MTB) belongs to a complex of related mycobacteria species (*M. tuberculosis* complex or MTBC) including *M. tuberculosis*, *Mycobacterium africanum*, *Mycobacterium bovis*, *Mycobacterium microti*, *Mycobacterium canetti*, *Mycobacterium pinnipedii* and *Mycobacterium caprae* [6]. The MTB genome was first sequenced in 1998 from the H37Rv strain and comprises of around 4.4 million base pairs, codes for around 4000 genes (many of which are involved in lipogenesis and lipolysis) and has a high (65.6%) guanine and cytosine (G + C) content [7]. It has long been thought that there is minimal genetic variation between MTB strains, and what variation exists is of minimal clinical significance. WGS has played an important role in challenging this dogma, suggesting that there is a greater than previously appreciated genomic variance between MTB strains which can give rise to clinically significant phenotypic differences (for example the increased virulence observed in the Beijing family of strains) [8–13]. Unlike other bacterial species, drug resistance in MTB is largely attributable to small genetic events such as single nucleotide polymorphisms (SNPs) as opposed to other sources of variation (i.e. horizontal transfer by mobile genetic elements such as plasmids) [14]. As such, novel mutations which could potentially confer drug resistance need to be carefully screened for, hence sensitive methods which can look at the whole MTB genome are desirable for understanding the mechanisms of drug resistance and the evolution of drug resistance over time.

There are a number of established and well characterised drug resistance mechanisms within the MTB genome, mainly arising through point mutations in genes encoding for the target protein of the drug, or proteins required for drug activation. Rifampicin resistance commonly arises through mutations in an 81 base pair region in the *rpoB* gene known as the rifampicin resistance determining region (RRDR) – the target of the Xpert MTP/Rif® assay. The *rpoB* gene codes for the β-subunit of RNA polymerase which is the target of rifampicin. Mutations in this region account for approximately 96% of phenotypic rifampicin resistance and as more than 90% of rifampicin resistant strains are also resistant to isoniazid, rifampicin resistance is used as a surrogate marker for MDR-TB [15,16]. Isoniazid resistance arises through mutations in catalase peroxidase encoding genes (*kat*G) (which prevents activation of isoniazid resulting in high level resistance) or mutations in *inh*A, the target of the activated form of isoniazid (low level resistance). The majority of pyrazinamide resistance is conferred through the prevention of drug activation through a diverse set of mutations in the *pnc*A gene and ethambutol resistance through over expression of the target protein (Emb) or mutations in the *embB* gene in the majority of resistant strains. Mutations associated with resistance to second line anti-TB agents resulting in XDR-TB are also well described. The majority of fluoroquinolone resistance results through mutations in subunits A and B of DNA gyrase (topoisomerase II) – the target protein of the fluoroquinolones, and resistance to the second line injectable agents amikacin and kanamycin resulting from 16s rRNA mutations [17,18]. Predictable and well characterised mutations such as these make them targets for molecular based tests, however a clear mechanism of drug resistance is not established in 10–40% of resistant strains. The genomic basis of resistance in these cases are likely to be SNPs occurring outside established resistance genes and the mechanism of resistance uncertain [19]. WGS can be used to accurately detect these SNPs [20] and can help define the mechanism of resistance and importantly allow development over time of a reference library of resistance mutations against which an index strain can be compared.

Current molecular methods in TB diagnostics and molecular epidemiology

Sputum microscopy and microbiological culture is the 'gold-standard' for the diagnosis of TB. Whilst direct microscopy offers the advantages of being a rapid and inexpensive test with high specificity for TB (in endemic areas), the method suffers from low (and variable) sensitivity (20–60%) [21]. Depending on the setting and resources available, samples will be cultured in liquid medium over several weeks or on

solid medium (for example Löwenstein-Jensen medium) for 8 (up to 12 weeks) [22]. Liquid culture platforms, for example the Mycobacterium Growth Indicator Tube (MGIT) system (Becton-Dickenson, US) have been able to reduce the time to positivity compared to solid media from 21 days to 6–12 days [23,24]. Direct nucleic acid amplification tests (NAAT) are playing an increasingly important role in the diagnosis of TB, with the WHO endorsing the Xpert MTB/RIF® test in 2010. The Xpert MTB/Rif® assay is a sample-in-answer-out test run on the GeneXpert® system which extracts nucleic acid, amplifies the RRDR of the MTBC, probes for mutations using molecular beacons and analyses using the GeneXpert® software [25]. This system can provide identification of the presence of MTBC as well as identifying potential resistance to rifampicin at or near the point-of-care with 2 h with 89% (85–92%) sensitivity and 99% (98–99%) specificity [4,26–28]. Loop mediated isothermal amplification (LAMP) is an alternative NAAT – an isothermal amplification method using 4 primers to target 6 distinct DNA sequences [29]. The key advantage of LAMP over the Xpert MTB/RIF® assay is that there is no requirement for a thermal-cycler. A recent meta-analysis suggests a pooled sensitivity of 93% (92–95%) and specificity of 94% (92–95%) compared to culture, however there was a high degree of heterogeneity between studies ($I^2 > 85\%$) [30] .

Methods currently used for species/strain typing in molecular epidemiology of MTBC include IS6110 restriction fragment length polymorphisms (RFLP), spoligotyping, variable-number-tandem repeat (VNTR) and multispacer sequence typing. IS6110 is a mobile genetic element found in MTBC and is found in multiple copies in the MTB genome, the number of copies varies between strains so can be used as a tool to discriminate between strains of MTB [31–33]. Spoligotyping (spacer-oligotyping) utilises the DNA polymorphism within the Direct Repeat (DR) locus of the MTB genome. The DR locus consists of 36 bp DRs with variable 'spacer' regions interspersed between them. Species within the MTBC vary in the number of DR repeats and presence of specific variable spacers allowing some separation of species [34]. Similarly, VNTR is a molecular typing tool based upon 24 loci within the MTB genome which contain variable numbers of tandem repeats called mycobacterial interspersed repetitive units (MIRU-VNTRs). The technique uses PCR primers designed to amplify these with subsequent determination of the size of PCR product (amplicon) using gel electrophoresis which can be used in turn to determine the number of MIRU-VNTR copies [35]. Whilst these tools are useful in epidemiological studies, there are some disadvantages, for example IS6110 RFLP requires a high volume of starting DNA and whilst spoligotyping can differentiate species within the MTBC, it lacks the resolution to resolve closely related strains of MTB. These techniques have been reviewed in more detail elsewhere [9,36,37].

Next generation sequencing

The past decade has seen a revolution in the way in which genetic sequencing is performed. The original chain termination method of sequencing deoxyribonucleic acid (DNA) described by Sanger has been surpassed by NGS technology [38,39]. NGS can provide whole bacterial genome sequences in a short timeframe (hours–days), at a relatively low cost and, increasingly, without the need for culture [40]. There are a number of different NGS platforms available which can be broadly classified into those which result in DNA sequences which are relatively short (<1000 base pairs), so-called 'short reads', and those providing longer DNA sequences (long reads >5000 base pair average) both of which offer theoretical advantages. The depth of information provided through NGS has the potential to improve the speed and accuracy of MDR/XDR-TB diagnosis, and improve our understanding the epidemiology of TB.

NGS technologies produce thousands-millions of reads of varying length, which can then be reassembled into longer or whole genome sequences using bioinformatics. The next generation sequencing technologies which provide long sequence reads include PacBio® single molecule-real time (SMRT®), (Pacific Biosciences, US) and MinION® nanopore sequencing (Oxford Nanopore Technologies, UK). Sequencing by synthesis® (Illumina Inc., US) and Ion Torrent semiconductor sequencing (ThermoFisher Scientific, US) both provide short read lengths. Illumina are currently the NGS market leaders, offering high accuracy (~99%) short-read (⩽300 bases) sequence data. Current infectious diseases applications include species identification, isolate sequencing and typing for public health surveillance [41,42], for example, the use of WGS for Salmonella surveillance [43,44], outbreak surveillance and transmission studies (for example monitoring methicillin resistant *Staphylococcus aureus* outbreaks) [45,46]. PacBio provides long read sequencing and has been used extensively in microbiology research including mycobacterial research [47–50]. The MinION platform offers long read sequencing and is a highly transportable sequencing device at a much reduced cost (1000 USD) can be plugged into a laptop computer and has been used for a number of epidemiology and diagnostic applications including the recent Ebola outbreak in West Africa and for the diagnosis of pathogens and antimicrobial resistance directly from urine [51,52].

Whilst NGS is a powerful tool, it does have associated challenges which need consideration. One challenge is being able to determine whether differences between sequences occur because of sequence variance or because of sequencing error. Error rates are different for the different sequencing technologies (e.g. indels, G/C bias, homopolymer errors) and types of error are also different. In order to overcome this, multiple reads of the same sequence are required (sequencing depth), which can be compared to a reference dataset or used to generate consensus data for SNP calling (not possible if errors are systematic). Also, genomic DNA contains repeat elements which can be hundreds to thousands of base pairs in length. Short reads are unable to 'bridge' these repeats and so a complete genome cannot be assembled. Long reads can bridge these repeat elements and can be *de novo* assembled relatively easily, resulting in complete genomes [53,54].

Despite some challenges, the potential benefits NGS holds over current methods for the diagnosis and epidemiological monitoring of TB include time to diagnosis, depth of information provided and convenience:

1: Turnaround time to results can be greatly reduced (from weeks to hours) using NGS for TB diagnosis directly from sputum or determining the antibiotic resistance of an isolate [55,56].
2: The depth of information provided by NGS can be used to identify whether there are multiple co-infecting strains within the same individual (and has demonstrated the presence of mixed strain infection in historical TB isolates [57,58]) and to help differentiate relapse from re-infection [59]. Epidemiological information on circulating strains, drug resistance and chains of transmission can be collected in real-time.
3: Convenience: NGS has the potential to combine diagnosis, drug resistance profiling and epidemiological analysis into one tests [9].

NGS can be performed from isolates of MTB following culture, or, directly from clinical isolates. The ability to detect and characterise MTB directly from clinical samples holds the greatest potential in terms of improving the management of MTB but also presents the greatest challenge. The remainder of this review will examine the current use of NGS in detecting MTB from cultured isolates and directly from clinical samples.

Whole genome sequencing from isolates

Several retrospective studies have looked into the role of WGS in detecting drug resistant TB from cultured isolates and have found a high level of genotypic-phenotypic concordance with regards to antimicrobial susceptibility [55,56]. One study in particular [56], developed a training set of 120 resistance determining and 772 benign mutations and successfully predicted susceptibility in 89.2% of 1552 validation-set phenotypes (449 isolates phenotypically resistant to at least one drug, 284 MDR-TB isolates and 3 XDR-TB isolates). This level of concordance is promising suggesting a clear role for WGS in predicting drug susceptibility in MTB, however further study is required as the genetic basis for many phenotypically resistant strains is not fully understood. This gap in knowledge was highlighted in one study comparing the sensitivity of techniques for predicting antimicrobial resistance from WGS, achieving 99.1% concordance with phenotypic results for *S. aureus* but only 82.6% concordance for MTB [60].

Recent studies have examined the potential role of WGS in informing the clinical management of TB [61,62]. One prospective study [61] has utilised the Ion Torrent system to explore the role of WGS in informing the clinical management of XDR-TB in 16 clinical specimens from 6 patients. The authors demonstrated that resistance phenotype could, in most cases, be reasonably predicted from the genome sequence and WGS was useful in improving confidence in clinical decision making. The authors point out that there is a need for consensus agreement for genotype-phenotype standards. In order to help address this issue, a number of databases have been set-up containing information on mutations conferring drug resistance such as TBDreamDB and MUBII-TB-DB [56,63–65]. One group has produced a collated database, containing information on resistance to 11 anti-TB agents through 1325 mutations and has developed a software tool, '*TB profiler*', to allow raw sequencing data to be screened and drug resistance predicted *in silico* [66]. More recently a large, multi-centred prospective study has begun to challenge some of the perceived barriers to WGS by comparing real-time WGS of MTB isolates from positive culture with traditional laboratory workflow. The study was conducted across 8 laboratories in Europe and North America with local sequencing (Illumina) and centralised bioinformatics. The study demonstrated that WGS predicted both species and phenotypic drug resistance with 93% accuracy compared to traditional workflow with a potential 7% cost improvement [67].

The role of WGS in investigating the transmission of TB is the subject of much interest given the potential to be able to differentiate between strains of TB that are indistinguishable based on current molecular epidemiological techniques, so providing a tool with improved resolution [68,69]. The principal is that mutations which arise in the MTB genome from one individual will be passed onto all the other people that are infected by the original host [36]. A recent systematic review of the role of WGS in TB transmission studies revealed that the use of WGS data to infer transmission and direction of transmission of TB varies between studies and settings [70]. One key area of difference was the use of SNP thresholds to define transmission – the thresholds used for transmission studies varied from 0 to 2 SNPs whilst others studies used $\leqslant$10. The authors point out that the particular SNP threshold is context dependent with factors including the strain diversity, within-host diversity, read quality and number of amplification steps making comparisons across studies challenging. A further systematic review designed to examine how WGS compared to conventional epidemiology (i.e. contact tracing and current molecular methods) compared 12 studies. The study revealed that WGS had high sensitivity (78.5–100%) compared to conventional epidemiological tools and had a higher discriminatory power than conventional genotyping to differentiate clusters of MTB strains [71].

Whole genome sequencing directly from clinical samples

Given the need for reduced turnaround time to diagnosis and rapid recognition of drug-resistance, one of the most desirable applications of NGS is to be able to sequence pathogen DNA directly from clinical samples (shotgun metagenomics) without the need for time consuming culture. One of the key challenges in achieving this goal is the need for a sufficient proportion of pathogen DNA in the metagenome in order to generate enough breadth and depth of coverage of the pathogen genome. However, the numbers of MTBC in clinical samples tends to be low (pauci-bacillary) meaning a low ratio of pathogen:bacterial and pathogen:human DNA – this is especially challenging in both the HIV population and the paediatric population (from whom obtaining sputum can be difficult) [72]. One study has demonstrated that SMS can be applied directly to sputum samples which have undergone a relatively straightforward differential lysis and DNA extraction process. The authors were able to correctly detect and characterise MTB sequences in 7 out of 8 samples taken

from smear positive patients. Although the organisms were correctly identified to species/lineage/clade level providing useful information, there was a high number of non-TB reads in the sequencing data and both the breadth of coverage of the H37Rv reference genome (0.002X to 0.7X) and the depth of sequencing (1.2–1.9X) were too low to be able to confidently predict resistance [73]. This study highlights the promising potential for SMS in diagnosing TB but also the need for development of enrichment processes which efficiently increases the ratio of MTB:non-MTB (human, bacterial, viral, fungal) DNA. A recent study describes one such pathogen specific enrichment method that yields enough DNA for SMS directly from clinical samples [74]. The enrichment process uses custom designed RNA baits (SureSelect^{XT}, Agilent, US) to specifically capture MTB DNA from total DNA extracted from sputum. When applied to clinical samples this method improved the number of MTB reads from 0.3% in non-enriched samples to 82% in enriched samples and the mean depth of sequencing from 4.6x to 200x [74].

Challenges

There are a number of key issues which make the translation of NGS methods in under-resourced settings a challenge. NGS hasn't yet penetrated most low and middle income countries (LMICs) where TB is most prevalent and until recently it hasn't been feasible to apply it to infectious diseases diagnosis. Short-read highly multiplexed NGS is becoming affordable (<$50 per bacterial genome) but waiting to batch sufficient samples erodes the time saving advantage of this approach. Also, the cost of the machines/maintenance, the turnaround time for sequencing and the complexity of the informatics make short-read technology unlikely to be utilised for infectious diseases diagnostics in LMICs. Long-read technologies have lower read-accuracy (by 5–10% for single reads) but data analysis is easier. The PacBio RSII and Sequel devices suffer many of the same disadvantages as short-read sequencing devices but MinION sequencing demonstrates great potential for infectious diseases diagnostics in LMICs. A combination of portability, low capital cost ($1000 USD), no maintenance cost, automated rapid library preparation (Voltrax) and sequencing and automated real-time analysis makes MinION ideal for TB diagnosis (from culture for DST and/or directly from sputum) in LMICs. Cost per flow cell ($500 USD) is an issue, but current sequence yields (R9.4 – up to 10 Gb) make it feasible to multiplex (up to 12 samples) and it may be possible to reuse flow cells in the future. In a recent study, we demonstrated the potential of MinION sequencing for diagnostics, achieving pathogen identification and antibiotic resistance profiling directly from clinical samples (urine) within 4 h. This included pathogen DNA enrichment, DNA extraction, rapid library preparation (10 min), 1 h MinION sequencing and data analysis [52].

Conclusion

NGS holds great potential for improving the diagnosis and management of TB and drug-resistant TB. WGS of MTBC isolates is a rapid and cost effective method for molecular epidemiology and drug resistance detection, although more work needs to be done to characterise resistance associated mutations so that better concordance is achieved between resistance genotype and phenotype. Shotgun metagenomics sequencing directly from sputum for rapid diagnosis and resistance detection is the ultimate application of NGS for TB but is hampered by low genome coverage caused by paucibacillary samples and competing bacterial and human DNA. Targeted approaches, such as specifically capturing the *M. tuberculosis* complex cells or DNA, and/or efficient host DNA depletion strategies are necessary to achieve the sensitivity of detection and genome coverage required.

Conflict of interest

None.

REFERENCES

[1] WHO, Global Tuberculosis Report, Switzerland, Geneva, 2015, p. 2015.
[2] WHO, The World Health Organisation end TB Strategy, WHO, Geneva, Switzerland, 2015, p. 2015.
[3] FIND, Foundation for Innovative New Diagnostics: Strategy for Tuberculosis (and Lower Respiratory Tract Infections) 2015–2020, FIND Communications, Geneva, Switzerland, 2014.
[4] M. Bates, A. Zumla, The development, evaluation and performance of molecular diagnostics for detection of *Mycobacterium tuberculosis*, Expert Rev. Mol. Diagn. (2016).
[5] WHO, Tuberculosis Diagnostic Technology and Market Landscape, UNITAID, WHO, Geneva, 2015.
[6] K. Reddington, A. Zumla, M. Bates, D. van Soolingen, S. Niemann, T. Barry, J. O'Grady, T.B. Seek, A two-stage multiplex real-time-PCR-based method for differentiation of the *Mycobacterium tuberculosis* complex, J. Clin. Microbiol. 50 (7) (2012) 2203–2206.
[7] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (6685) (1998) 537–544.
[8] S. Niemann, P. Supply, Diversity and evolution of *Mycobacterium tuberculosis*: moving to whole-genome-based approaches, Cold Spring Harb. Perspect. Med. 4 (12) (2014) a021188.
[9] J.E. Galagan, Genomic insights into tuberculosis, Nat. Rev. Genet. 15 (5) (2014) 307–320.
[10] S. Gagneux, Genetic diversity in *Mycobacterium tuberculosis*, Curr. Top. Microbiol. Immunol. 374 (2013) 1–25.
[11] M. Coscolla, S. Gagneux, Does *M. tuberculosis* genomic diversity explain disease diversity?, Drug Discovery Today 7 (1) (2010) e43–e59
[12] P.A. Black, M. de Vos, G.E. Louw, R.G. van der Merwe, A. Dippenaar, E.M. Streicher, A.M. Abdallah, S.L. Sampson, T.C. Victor, T. Dolby, J.A. Simpson, P.D. van Helden, R.M. Warren, A. Pain, Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates, BMC Genomics 16 (1) (2015) 857.

[13] S. Gagneux, P.M. Small, Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development, Lancet Infect. Dis. 7 (5) (2007) 328–337.
[14] P.E. Almeida Da Silva, J.C. Palomino, Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs, J. Antimicrob. Chemother. 66 (7) (2011) 1417–1430.
[15] J.C. Palomino, A. Martin, Drug resistance mechanisms in *Mycobacterium tuberculosis*, Antibiotics (Basel) 3 (3) (2014) 317–340.
[16] A.S. Kalokhe, M. Shafiq, J.C. Lee, S.M. Ray, Y.F. Wang, B. Metchock, A.M. Anderson, M.L. Nguyen, Multidrug-resistant tuberculosis drug susceptibility and molecular diagnostic testing, Am. J. Med. Sci. 345 (2) (2013) 143–148.
[17] D. Laurenzo, S.A. Mousa, Mechanisms of drug resistance in *Mycobacterium tuberculosis* and current status of rapid molecular diagnostic testing, Acta Trop. 119 (1) (2011) 5–10.
[18] Y. Zhang, W.W. Yew, Mechanisms of drug resistance in *Mycobacterium tuberculosis*, Int. J. Tuberc. Lung Dis. 13 (11) (2009) 1320–1330.
[19] M.R. Farhat, B.J. Shapiro, K.J. Kieser, R. Sultana, K.R. Jacobson, T.C. Victor, R.M. Warren, E.M. Streicher, A. Calver, A. Sloutsky, D. Kaur, J.E. Posey, B. Plikaytis, M.R. Oggioni, J.L. Gardy, J.C. Johnston, M. Rodrigues, P.K. Tang, M. Kato-Maeda, M.L. Borowsky, B. Muddukrishna, B.N. Kreiswirth, N. Kurepina, J. Galagan, S. Gagneux, B. Birren, E.J. Rubin, E.S. Lander, P.C. Sabeti, M. Murray, Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*, Nat. Genet. 45 (10) (2013) 1183–1189.
[20] H. Zhang, D. Li, L. Zhao, J. Fleming, N. Lin, T. Wang, Z. Liu, C. Li, N. Galwey, J. Deng, Y. Zhou, Y. Zhu, Y. Gao, S. Wang, Y. Huang, M. Wang, Q. Zhong, L. Zhou, T. Chen, J. Zhou, R. Yang, G. Zhu, H. Hang, J. Zhang, F. Li, K. Wan, J. Wang, X.E. Zhang, L. Bi, Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance, Nat. Genet. 45 (10) (2013) 1255–1260.
[21] K.R. Steingart, V. Ng, M. Henry, P.C. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M.D. Perkins, M.A. Aziz, M. Pai, Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review, Lancet Infect. Dis. 6 (10) (2006) 664–674.
[22] Standards Unit, Microbiology Services, Public Health England, UK Standards for Microbiology Investigations: Investigation of Specimens for Mycobacterium species, Bacteriology (B40), Public Health England, London, UK, 2014 (Issue 6.1).
[23] Diagnostic standards and classification of tuberculosis in adults and children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999, Am. J. Respir. Crit. Care Med. 161 (4 Pt. 1) (2000) 1376–1395.
[24] J. Dinnes, J. Deeks, H. Kunst, A. Gibson, E. Cummins, N. Waugh, F. Drobniewski, A. Lalvani, A systematic review of rapid diagnostic tests for the detection of tuberculosis infection, Health Technol. Assess. 11 (3) (2007) 1–196.
[25] C.C. Boehme, P. Nabeta, D. Hillemann, M.P. Nicol, S. Shenai, F. Krapp, J. Allen, R. Tahirli, R. Blakemore, R. Rustomjee, A. Milovic, M. Jones, S.M. O'Brien, D.H. Persing, S. Ruesch-Gerdes, E. Gotuzzo, C. Rodrigues, D. Alland, M.D. Perkins, Rapid molecular detection of tuberculosis and rifampin resistance, N. Engl. J. Med. 363 (11) (2010) 1005–1015.
[26] A. Zumla, M. Raviglione, R. Hafner, C.F. von Reyn, Tuberculosis, N. Engl. J. Med. 368 (8) (2013) 745–755.
[27] J. O'Grady, M. Bates, L. Chilukutu, J. Mzyece, B. Cheelo, M. Chilufya, L. Mukonda, M. Mumba, J. Tembo, M. Chomba, N. Kapata, M. Maeurer, A. Rachow, P. Clowes, M. Hoelscher, P. Mwaba, A. Zumla, Evaluation of the Xpert MTB/RIF assay at a tertiary care referral hospital in a setting where tuberculosis and HIV infection are highly endemic, Clin. Infect. Dis. 55 (9) (2012) 1171–1178.
[28] K.R. Steingart, I. Schiller, D.J. Horne, M. Pai, C.C. Boehme, N. Dendukuri, Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults, Cochrane Database Syst. Rev. 1 (2014) CD009593.
[29] T. Notomi, H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, T. Hase, Loop-mediated isothermal amplification of DNA, Nucleic Acids Res. 28 (12) (2000) E63.
[30] L. Yan, H. Xiao, Q. Zhang, Systematic review: comparison of Xpert MTB/RIF, LAMP and SAT methods for the diagnosis of pulmonary tuberculosis, Tuberculosis (Edinburg) 96 (2016) 75–86.
[31] D. Thierry, P. Matsiota-Bernard, E. Pitsouni, C. Costopoulos, J. L. Guesdon, Use of the insertion element IS6110 for DNA fingerprinting of *Mycobacterium tuberculosis* isolates presenting various profiles of drug susceptibility, FEMS Immunol. Med. Microbiol. 6 (4) (1993) 287–297.
[32] D. Thierry, M.D. Cave, K.D. Eisenach, J.T. Crawford, J.H. Bates, B. Gicquel, J.L. Guesdon, IS6110, an IS-like element of *Mycobacterium tuberculosis* complex, Nucleic Acids Res. 18 (1) (1990) 188.
[33] D. Thierry, A. Brisson-Noël, V. Vincent-Lévy-Frébault, S. Nguyen, J.L. Guesdon, B. Gicquel, Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis, J. Clin. Microbiol. 28 (12) (1990) 2668–2673.
[34] J. Kamerbeek, L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, J. van Embden, Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology, J. Clin. Microbiol. 35 (4) (1997) 907–914.
[35] P. Supply, C. Allix, S. Lesjean, M. Cardoso-Oelemann, S. Rüsch-Gerdes, E. Willery, E. Savine, P. de Haas, H. van Deutekom, S. Roring, P. Bifani, N. Kurepina, B. Kreiswirth, C. Sola, N. Rastogi, V. Vatin, M.C. Gutierrez, M. Fauville, S. Niemann, R. Skuce, K. Kremer, C. Locht, D. van Soolingen, Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*, J. Clin. Microbiol. 44 (12) (2006) 4498–4510.
[36] M. Wlodarska, J.C. Johnston, J.L. Gardy, P. Tang, A microbiological revolution meets an ancient disease: improving the management of tuberculosis with genomics, Clin. Microbiol. Rev. 28 (2) (2015) 523–539.
[37] A.C. Schürch, D. van Soolingen, DNA fingerprinting of *Mycobacterium tuberculosis*: from phage typing to whole-genome sequencing, Infec. Genet. Evol. 12 (4) (2012) 602–609.
[38] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U.S.A. 74 (12) (1977) 5463–5467.
[39] C. Bertelli, G. Greub, Rapid bacterial genome sequencing: methods and applications in clinical microbiology, Clin. Microbiol. Infect. 19 (9) (2013) 803–813.
[40] K. Bjorn-Mortensen, J. Zallet, T. Lillebaek, A.B. Andersen, S. Niemann, E.M. Rasmussen, T.A. Kohl, Direct DNA extraction from *Mycobacterium tuberculosis* frozen stocks as a reculture-independent approach to whole-genome sequencing, J. Clin. Microbiol. 53 (8) (2015) 2716–2719.
[41] J.C. Kwong, N. McCallum, V. Sintchenko, B.P. Howden, Whole genome sequencing in clinical and public health microbiology, Pathology 47 (3) (2015) 199–210.
[42] X. Didelot, R. Bowden, D.J. Wilson, T.E. Peto, D.W. Crook, Transforming clinical microbiology with bacterial genome sequencing, Nat. Rev. Genet. 13 (9) (2012) 601–612.

[43] P.M. Ashton, S. Nair, T.M. Peters, J.A. Bale, D.G. Powell, A. Painset, R. Tewolde, U. Schaefer, C. Jenkins, T.J. Dallman, E.M. de Pinna, K.A. Grant, S.W.G.S.I. Group, Identification of Salmonella for public health surveillance using whole genome sequencing, PeerJ 4 (2016) e1752.
[44] S. Nair, P. Ashton, M. Doumith, S. Connell, A. Painset, S. Mwaigwisya, G. Langridge, E. de Pinna, G. Godbole, M. Day, WGS for surveillance of antimicrobial resistance: a pilot study to detect the prevalence and mechanism of resistance to azithromycin in a UK population of non-typhoidal Salmonella, J. Antimicrob. Chemother. (2016).
[45] D.W. Eyre, T. Golubchik, N.C. Gordon, R. Bowden, P. Piazza, E. M. Batty, C.L. Ip, D.J. Wilson, X. Didelot, L. O'Connor, R. Lay, D. Buck, A.M. Kearns, A. Shaw, J. Paul, M.H. Wilcox, P.J. Donnelly, T.E. Peto, A.S. Walker, D.W. Crook, A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance, BMJ Open 2 (3) (2012).
[46] S.R. Harris, E.J. Cartwright, M.E. Török, M.T. Holden, N.M. Brown, A.L. Ogilvy-Stuart, M.J. Ellington, M.A. Quail, S.D. Bentley, J. Parkhill, S.J. Peacock, Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study, Lancet Infect. Dis. 13 (2) (2013) 130–136.
[47] L. Zhu, J. Zhong, X. Jia, G. Liu, Y. Kang, M. Dong, X. Zhang, Q. Li, L. Yue, C. Li, J. Fu, J. Xiao, J. Yan, B. Zhang, M. Lei, S. Chen, L. Lv, B. Zhu, H. Huang, F. Chen, Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology, Nucleic Acids Res. 44 (2) (2016) 730–743.
[48] T. Kanda, Y. Furuse, H. Oshitani, T. Kiyono, Highly efficient CRISPR/Cas9-mediated cloning and functional characterization of gastric cancer-derived epstein-barr virus strains, J. Virol. 90 (9) (2016) 4383–4393.
[49] A.E. Sheppard, N. Stoesser, D.J. Wilson, R. Sebra, A. Kasarskis, L.W. Anson, A. Giess, L.J. Pankhurst, A. Vaughan, C.J. Grim, H. L. Cox, A.J. Yeh, C.D. Sifri, A.S. Walker, T.E. Peto, D.W. Crook, A. J. Mathers, M.M.M.M.I. Group, Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blaKPC, Antimicrob. Agents Chemother. 60 (6) (2016) 3767–3778.
[50] D.A. Rasko, D.R. Webster, J.W. Sahl, A. Bashir, N. Boisen, F. Scheutz, E.E. Paxinos, R. Sebra, C.S. Chin, D. Iliopoulos, A. Klammer, P. Peluso, L. Lee, A.O. Kislyuk, J. Bullard, A. Kasarskis, S. Wang, J. Eid, D. Rank, J.C. Redman, S.R. Steyert, J. Frimodt-Møller, C. Struve, A.M. Petersen, K.A. Krogfelt, J.P. Nataro, E.E. Schadt, M.K. Waldor, Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany, N. Engl. J. Med. 365 (8) (2011) 709–717.
[51] J. Quick, N.J. Loman, S. Duraffour, J.T. Simpson, E. Severi, L. Cowley, J.A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouédraogo, B. Afrough, A. Bah, J.H. Baum, B. Becker-Ziaja, J.P. Boettcher, M. Cabeza-Cabrerizo, Á. Camino-Sánchez, L.L. Carter, J. Doerrbecker, T. Enkirch, I. García-Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L.E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C.H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallasch, L.V. Patrono, J. Portmann, J.G. Repits, N.Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R.L. Yemanaberhan, E.G. Zekeng, T. Racine, A. Bello, A.A. Sall, O. Faye, N. Magassouba, C.V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, M. Gerard, G. Bado, B. Baillet, D. Delaune, K.Y. Nebie, A. Diarra, Y. Savane, R.B. Pallawo, G.J. Gutierrez, N. Milhano, I. Roger, C.J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D.J. Turner, G. Pollakis, J.A. Hiscox, D.A. Matthews, M.K. O'Shea, A.M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Wölfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keïta, A. Rambaut, P. Formenty, S. Günther, M.W. Carroll, Real-time, portable genome sequencing for Ebola surveillance, Nature 530 (7589) (2016) 228–232.
[52] K. Schmidt, S. Mwaigwisya, L. Crossman, M. Doumith, D. Munroe, C. Pires, A. Khan, N. Woodford, N. Saunders, J. Wain, J. O'Grady, D. Livermore, Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore based metagenomic sequencing, J. Antimicrob. Chemother. (2016) (in press).
[53] D. Sims, I. Sudbery, N.E. Ilott, A. Heger, C.P. Ponting, Sequencing depth and coverage: key considerations in genomic analyses, Nat. Rev. Genet. 15 (2) (2014) 121–132.
[54] P.M. Ashton, S. Nair, T. Dallman, S. Rubino, W. Rabsch, S. Mwaigwisya, J. Wain, J. O'Grady, MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island, Nat. Biotechnol. 33 (3) (2015) 296–300.
[55] C.U. Köser, J.M. Bryant, J. Becq, M.E. Török, M.J. Ellington, M.A. Marti-Renom, A.J. Carmichael, J. Parkhill, G.P. Smith, S.J. Peacock, Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*, N. Engl. J. Med. 369 (3) (2013) 290–292.
[56] T.M. Walker, T.A. Kohl, S.V. Omar, J. Hedge, C. Del Ojo Elias, P. Bradley, Z. Iqbal, S. Feuerriegel, K.E. Niehaus, D.J. Wilson, D.A. Clifton, G. Kapatai, C.L. Ip, R. Bowden, F.A. Drobniewski, C. Allix-Béguec, C. Gaudin, J. Parkhill, R. Diel, P. Supply, D.W. Crook, E.G. Smith, A.S. Walker, N. Ismail, S. Niemann, T.E. Peto, M.M.M.M.I. Group, Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study, Lancet Infect. Dis. 15 (10) (2015) 1193–1202.
[57] H.D. Donoghue, Insights into ancient leprosy and tuberculosis using metagenomics, Trends Microbiol. 21 (9) (2013) 448–450.
[58] G.L. Kay, M.J. Sergeant, Z. Zhou, J.Z. Chan, A. Millard, J. Quick, I. Szikossy, I. Pap, M. Spigelman, N.J. Loman, M. Achtman, H. D. Donoghue, M.J. Pallen, Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe, Nat. Commun. 6 (2015) 6717.
[59] J.M. Bryant, S.R. Harris, J. Parkhill, R. Dawson, A.H. Diacon, P. van Helden, A. Pym, A.A. Mahayiddin, C. Chuchottaworn, I.M. Sanne, C. Louw, M.J. Boeree, M. Hoelscher, T.D. McHugh, A.L. Bateson, R.D. Hunt, S. Mwaigwisya, L. Wright, S.H. Gillespie, S.D. Bentley, Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study, Lancet Respir. Med. 1 (10) (2013) 786–792.
[60] P. Bradley, N.C. Gordon, T.M. Walker, L. Dunn, S. Heys, B. Huang, S. Earle, L.J. Pankhurst, L. Anson, M. de Cesare, P. Piazza, A.A. Votintseva, T. Golubchik, D.J. Wilson, D.H. Wyllie, R. Diel, S. Niemann, S. Feuerriegel, T.A. Kohl, N. Ismail, S.V. Omar, E.G. Smith, D. Buck, G. McVean, A.S. Walker, T.E. Peto, D.W. Crook, Z. Iqbal, Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*, Nat. Commun. 6 (2015) 10063.
[61] A.A. Witney, K.A. Gould, A. Arnold, D. Coleman, R. Delgado, J. Dhillon, M.J. Pond, C.F. Pope, T.D. Planche, N.G. Stoker, C.A. Cosgrove, P.D. Butcher, T.S. Harrison, J. Hinds, Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases, J. Clin. Microbiol. 53 (5) (2015) 1473–1483.
[62] A.C. Outhred, P. Jelfs, B. Suliman, G.A. Hill-Cawthorne, A.B. Crawford, B.J. Marais, V. Sintchenko, Added value of whole-genome sequencing for management of highly drug-resistant TB, J. Antimicrob. Chemother. 70 (4) (2015) 1198–1202.

[63] J.L. Gardy, Towards genomic prediction of drug resistance in tuberculosis, Lancet Infect. Dis. 15 (10) (2015) 1124–1125.
[64] A. Sandgren, M. Strong, P. Muthukrishnan, B.K. Weiner, G.M. Church, M.B. Murray, Tuberculosis drug resistance mutation database, PLoS Med. 6 (2) (2009) e2.
[65] J.P. Flandrois, G. Lina, O. Dumitrescu, MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*, BMC Bioinformatics 15 (2014) 107.
[66] F. Coll, R. McNerney, M.D. Preston, J.A. Guerra-Assunção, A. Warry, G. Hill-Cawthorne, K. Mallard, M. Nair, A. Miranda, A. Alves, J. Perdigão, M. Viveiros, I. Portugal, Z. Hasan, R. Hasan, J.R. Glynn, N. Martin, A. Pain, T.G. Clark, Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences, Genome Med. 7 (1) (2015) 51.
[67] L.J. Pankhurst, C. Del Ojo Elias, A.A. Votintseva, T.M. Walker, K. Cole, J. Davies, J.M. Fermont, D.M. Gascoyne-Binzi, T.A. Kohl, C. Kong, N. Lemaitre, S. Niemann, J. Paul, T.R. Rogers, E. Roycroft, E.G. Smith, P. Supply, P. Tang, M.H. Wilcox, S. Wordsworth, D. Wyllie, L. Xu, D.W. Crook, C.-T.S. Group, Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study, Lancet Respir. Med. 4 (1) (2016) 49–58.
[68] A. Roetzer, R. Diel, T.A. Kohl, C. Rückert, U. Nübel, J. Blom, T. Wirth, S. Jaenicke, S. Schuback, S. Rüsch-Gerdes, P. Supply, J. Kalinowski, S. Niemann, Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study, PLoS Med 10 (2) (2013) e1001387.
[69] T. Luo, C. Yang, Y. Peng, L. Lu, G. Sun, J. Wu, X. Jin, J. Hong, F. Li, J. Mei, K. DeRiemer, Q. Gao, Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis, Tuberculosis (Edinburg) 94 (4) (2014) 434–440.
[70] H.A. Hatherell, C. Colijn, H.R. Stagg, C. Jackson, J.R. Winter, I. Abubakar, Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review, BMC Med. 14 (1) (2016) 21.
[71] V. Nikolayevskyy, K. Kranzer, S. Niemann, F. Drobniewski, Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: a systematic review, Tuberculosis (Edinburg) 98 (2016) 77–85.
[72] M. Bates, J. O'Grady, M. Maeurer, J. Tembo, L. Chilukutu, C. Chabala, R. Kasonde, P. Mulota, J. Mzyece, M. Chomba, L. Mukonda, M. Mumba, N. Kapata, A. Rachow, P. Clowes, M. Hoelscher, P. Mwaba, A. Zumla, Assessment of the Xpert MTB/RIF assay for diagnosis of tuberculosis with gastric lavage aspirates in children in sub-Saharan Africa: a prospective descriptive study, Lancet Infect. Dis. 13 (1) (2013) 36–42.
[73] E.L. Doughty, M.J. Sergeant, I. Adetifa, M. Antonio, 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 2 (2014) e585.
[74] A.C. Brown, J.M. Bryant, K. Einer-Jensen, J. Holdstock, D.T. Houniet, J.Z. Chan, D.P. Depledge, V. Nikolayevskyy, A. Broda, M.J. Stone, M.T. Christiansen, R. Williams, M.B. McAndrew, H. Tutill, J. Brown, M. Melzer, C. Rosmarin, T.D. McHugh, R.J. Shorten, F. Drobniewski, G. Speight, J. Breuer, Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples, J. Clin. Microbiol. 53 (7) (2015) 2230–2237.