International Journal of Infectious Diseases 56 (2017) 130–135

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

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

Review

Removing the bottleneck in whole genome sequencing of *Mycobacterium tuberculosis* for rapid drug resistance analysis: a call to action

Ruth McNerney [a,*], Taane G. Clark [b], Susana Campino [b], Camilla Rodrigues [c], David Dolinger [d], Liezel Smith [a], Andrea M. Cabibbe [e], Keertan Dheda [a], Marco Schito [f]

[a] Lung Infection and Immunity Unit, Department of Medicine, Division of Pulmonology and UCT Lung Institute, University of Cape Town, Old Main Building, Groote Schuur Hospital, Observatory, Cape Town, South Africa
[b] Faculty of Infectious and Tropical Diseases and Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK
[c] Department of Microbiology, P.D. Hinduja National Hospital and Medical Research Centre, Mumbai, India
[d] FIND, Campus Biotech, Geneva, Switzerland
[e] Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy
[f] Critical Path Institute, Tucson, Arizona, USA

ARTICLE INFO

Article history:
Received 4 October 2016
Received in revised form 30 November 2016
Accepted 30 November 2016

Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
Tuberculosis
Next-generation sequencing
Drug resistance
Clinical isolate
Sputum

SUMMARY

Whole genome sequencing (WGS) can provide a comprehensive analysis of *Mycobacterium tuberculosis* mutations that cause resistance to anti-tuberculosis drugs. With the deployment of bench-top sequencers and rapid analytical software, WGS is poised to become a useful tool to guide treatment. However, direct sequencing from clinical specimens to provide a full drug resistance profile remains a serious challenge. This article reviews current practices for extracting *M. tuberculosis* DNA and possible solutions for sampling sputum. Techniques under consideration include enzymatic digestion, physical disruption, chemical degradation, detergent solubilization, solvent extraction, ligand-coated magnetic beads, silica columns, and oligonucleotide pull-down baits. Selective amplification of genomic bacterial DNA in sputum prior to WGS may provide a solution, and differential lysis to reduce the levels of contaminating human DNA is also being explored. To remove this bottleneck and accelerate access to WGS for patients with suspected drug-resistant tuberculosis, it is suggested that a coordinated and collaborative approach be taken to more rapidly optimize, compare, and validate methodologies for sequencing from patient samples.

1. Introduction

The publication of the first *Mycobacterium tuberculosis* genome sequence in 1998 heralded a new era in tuberculosis (TB) research.[1] The bacterium was found to have a relatively small circular genome of approximately 4.5 million base pairs and was estimated to have around 4000 potential genes. Sequencing of the *M. tuberculosis* H37Rv reference strain was soon followed by other strains, and expectations were high that mechanisms of pathogenesis and virulence were about to be revealed that would enable the development of novel drugs and improved vaccines to assist in TB control.[2,3] However, the anticipated advances have been slow to materialize and such work continues, accompanied by the realization of the sophistication of *M. tuberculosis* as a highly successful human pathogen.

The early genomes were deciphered using various cloning and shotgun sequencing approaches, followed by assembly and annotation with an assortment of analytical tools – a painstaking process with each genome taking years to complete. Advances in technology over the past 15 years have greatly reduced the complexity, cost, and time of sequencing. Thousands of *M. tuberculosis* strains have since been sequenced, contributing to studies on evolution, transmission, and drug resistance.[4–10] Several Web-based tools have been developed to assist TB sequence analysis, and software is now freely available for rapid genotypic analysis and the identification of drug resistance-associated mutations.[11–15]

* Corresponding author.
E-mail address: Ruth.Mcnerney@uct.ac.za (R. McNerney).

http://dx.doi.org/10.1016/j.ijid.2016.11.422

Next-generation sequencing (NGS) is already used in clinical practice for characterizing cancers and hereditary diseases. It may also be used to assist TB case management by detecting resistance to anti-TB drugs.[16–18] However, the quality of the DNA template is critical for successful whole genome sequencing (WGS), and *M. tuberculosis* presents considerable technical challenges in this respect. Firstly, the scarcity of bacilli in clinical samples limits the availability of *M. tuberculosis* genomic DNA (gDNA). Secondly, the remarkable hardiness of the lipid-rich *M. tuberculosis* cell wall makes disruption of the *M. tuberculosis* bacterium difficult and can affect the quality and the yield of gDNA. In addition, the *M. tuberculosis* genome itself is unusually robust, with a guanine/cytosine average content of 65% across the genome, and with some regions exceeding 80%.[1] Thus careful consideration must be given to the choice of sample, bacterial lysis, DNA extraction methodology, library preparation, and sequencing platform if *M. tuberculosis* sequencing is to enter routine clinical practice.[19]

This article explores sample collection and processing as crucial factors for WGS and NGS analysis of patient-derived *M. tuberculosis*. Although some questions remain unanswered, advice is offered on current best practices and pitfalls to avoid.

2. Next-generation sequencing platforms

NGS technologies analyse whole genomes without recourse to cloning, as was previously required for Sanger sequencing. High throughput platforms have been developed that can analyse millions of DNA fragments in parallel, and sequencing of bacterial genomes that previously took years to complete can now be achieved in hours. Sophisticated labelling systems allow multiplexing, where multiple samples are combined within a single run, greatly reducing costs. The early NGS platforms required specialized laboratories and highly trained personnel, but newer bench-top instruments suitable for clinical laboratories are now available. In addition to WGS, targeted sequencing, where multiple targets are amplified and sequenced in parallel, offers a rapid detection of resistance, albeit on a limited scale compared to whole genome approaches.[17,20]

Several NGS systems are commercially available, providing a range of platforms to choose from, some of which have gone through regulatory registration and clearance. Two NGS systems have been used to detect mutations associated with *M. tuberculosis* drug resistance.[21] Illumina sequencing (Illumina, San Diego, CA, USA) is based on reversible dye-terminators.[22] First a library is prepared whereby pure DNA is chopped into smaller fragments that are modified prior to amplification on a specialized chip holding hundreds of thousands of oligonucleotides. New fragments are built one nucleotide at a time, with fluorescent tags indicating which nucleotide has been incorporated. Sequential rounds of nucleotide additions build a new strand of DNA, with thousands of positions throughout the genome being sequenced at the same time in a process called massive parallel sequencing. The result is millions of short DNA fragments replicating the entire genome, with multiple fragments covering each stretch of gDNA. In most cases fragments are aligned and assembled by computation against a predetermined reference genome, thus allowing for the identification of polymorphisms. The depth of coverage (number of fragments that represents a specific nucleotide) is an indication of sequencing quality for each nucleotide sequenced, because if hundreds of fragments give the same signal for a specific position there is high confidence in the specific nucleotide call. Depth of coverage varies across genomes and is adversely affected by regions of high GC content and repetitive elements. Typically, when calling drug resistance mutations, coverage of at least 10-fold is required. However, other studies suggest at least 30-fold coverage is needed, else the result is considered of low confidence.[19]

Ion Torrent or Ion semiconductor sequencing (Ion Torrent Inc., USA, marketed by Thermo Fisher Scientific) uses different chemistry to monitor nucleotides incorporated during the creation by polymerization reaction of the new strand of DNA.[23] Sequencing is performed on semiconductor chips that detect changes in pH caused by the release of H^+ ions during the polymerization of DNA and incorporation of deoxyribonucleoside triphosphate (dNTP). dNTPs are added sequentially, with washing between each step to remove unbound nucleotide molecules. As with Illumina sequencing, assembly by computation and comparison with reference genomes provides a readout of the whole genome and identifies polymorphisms. Labelling dyes and optical detection are not employed in this platform, a factor that the manufacturer claims increases the speed of sequencing and reduces running costs. Ion Torrent sequencing has been used successfully to detect mutations associated with *M. tuberculosis* drug resistance.[16,21]

PacBio sequencing, or single molecule real-time sequencing (Pacific Biosciences, USA), is primarily being utilized as a research tool.[24] Rather than short fragments, this technology generates long strands of DNA and it is possible to assemble whole bacterial genomes from a single reaction. Assembly problems for repetitive regions encountered with short-read sequencing methods are overcome, for example the *M. tuberculosis* PE and PPE gene families, which are often excluded from Illumina sequencing data analysis. The technology may also be used to investigate DNA methylation, DNA damage, and other epigenetic information. PacBio sequencing is rapid, but has a limited capacity to multiplex, which translates to a moderate throughput. A drawback when considering clinical applications is the requirement for large amounts of pure high molecular mass gDNA, which for *M. tuberculosis* necessitates lengthy culture and extraction processes.

The MinION sequencing platform (Oxford Nanopore Technologies, UK) is a rapid technology with a portable sequencing instrument that also offers relatively long read lengths.[25] It was recently shown to detect antibiotic resistance genes of *Escherichia coli* in DNA purified from urine following processing to remove contaminating human DNA, with results available in approximately 4 h.[26] However, successful sequencing of the *M. tuberculosis* genome with this platform has not yet been reported.

The choice of sequencing strategy is influenced by the reasoning for sequencing. Identifying novel polymorphisms of biological significance ideally requires a depth of coverage of at least 100-fold, and for discovering rare polymorphisms up to a 1000-fold coverage may be preferred. Similarly, to investigate transmission chains and differentiate between closely related bacteria with confidence requires high coverage. Lower coverage (20- to 30-fold) may be sufficient to observe specific well-defined single nucleotide polymorphisms (SNPs). Thus, while high quality DNA with few strand breaks may be crucial for some research applications, DNA of lower quality may be tolerated when looking for known SNPs associated with drug resistance. For sequencing to be optimally effective for managing patients with drug-resistant TB, results should be available in days rather than weeks and ideally sequencing should be performed directly from primary clinical specimens. It should be noted, however, that when dealing with patient samples, mixed infections (more than one *M. tuberculosis* strain) and heteroresistance (more than one polymorphism at the same point) may require special consideration, and higher depth of reads are desirable. There is as yet no consensus on the depth of coverage required for NGS drug resistance screening for *M. tuberculosis*.

3. Sputum collection and processing

The sample of choice for the diagnosis of pulmonary TB is expectorated sputum, which is subjected to analysis by smear

Table 1
Reported DNA extraction methodologies for whole genome sequencing approaches for detecting drug resistance in *Mycobacterium tuberculosis*

Sample	Sequencing platform	Extraction methods	Reference
Liquid culture 4×10^6 to 10×10^6 CFU/ml	Ion PGM, Ion Torrent	Heat-killed (90 °C) followed by DNA purification with UCP Pathogen Minikit (Qiagen), following mechanical disruption using Pathogen Lysis Tubes with small beads (Qiagen)	Witney et al., 2015[16]
M. tuberculosis-positive liquid culture	MiSeq, Illumina	Sonication followed by heat (95 °C); removal of eukaryote DNA and other inhibitors with MolYsis Basic5 (Molzym Life Science); ethanol precipitation or commercial kit (QIAamp or QuickGene-Mini80); further purification using AMPure XP SPRI beads (Beckman Coulter)	Votintseva et al., 2015[59]
Smear-positive sputum	MiSeq, Illumina	*N*-acetyl-L-cysteine/NaOH decontamination, heat (80 °C), freeze/thaw, vortexing with glass beads, DNA purification with DNeasy Blood and Tissue DNA extraction kit (Qiagen); SureSelect^{XT} target enrichment	Brown et al. 2015[18]
Smear-positive sputum	MiSeq, Illumina (shotgun sequencing)	Liquefied with *N*-acetyl-L-cysteine and sodium citrate; diluted in PBS before pelleting (3220*g* for 20 min); two rounds of differential osmotic lysis of human cells using sterile water; pellet treated with the RNase-Free DNase (Qiagen), and washed before DNA extraction using NucleoSpin Tissue-Kit (Macherey-Nagel, Duren, Germany)	Doughty et al., 2014[41]

PBS, phosphate buffered saline.

microscopy, culture, or diagnostic PCR. Currently WGS is performed following the isolation and culture of the bacteria, which may take weeks; thus, it would be of considerable advantage to sequence directly from the clinical sample. Sputum may be obtained spontaneously (by instructing the patient to cough[27]) or by induction following the inhalation of a fine spray of saline.[28] *M. tuberculosis* is an aerosol-spread pathogen and as such stringent safety precautions must be implemented whenever live bacteria may be present. This includes during sample collection. *M. tuberculosis* bacilli are rarely found in saliva, and the quality of specimens received should be monitored to avoid processing of substandard samples.[29] The number of *M. tuberculosis* bacilli in a TB-positive sputum sample may vary between a handful to millions.[30] In the case of smear-positive sputum, it is assumed that the bacillary load is at least 10 000 bacilli per millilitre of sputum.

WGS and targeted NGS directly from sputum has been demonstrated, but to date no systematic investigations of sample preparation procedures with an output of providing WGS or targeted NGS quality gDNA suitable for detailed analysis of drug resistance-associated mutations has been reported (see Table 1). The quality of sequence data obtained (coverage and depth of reads) is dependent on the purity and integrity of DNA submitted. Sputum is a complex matrix containing mucus, human cells, cell debris, various bacteria and viruses, and sometimes blood and pus. It is often viscous with semi-solid lumps, which require thinning to release the bacteria. Three approaches are possible: heating, chemical treatment, and physical disruption, such as vortexing with 3–4-mm diameter glass beads.[31] However, all of these approaches risk damaging the bacterial cell wall, permitting direct exposure of the gDNA to extreme conditions, as well as the potential for the release of gDNA, which may be lost during subsequent washing and concentration steps. In addition, traditional techniques such as bead-beating and sonication may fragment the DNA and are not recommended for sequencing protocols where DNA of high integrity is needed.[32]

In order to amplify the number of *M. tuberculosis* bacteria from the sample, clinical specimens are cultured. Culture typically requires 3–4 weeks, but may take longer since some *M. tuberculosis* may be associated with a slower growing phenotype, particularly strains with high-grade drug resistance. It may be possible to sequence from early cultures, where the number of *M. tuberculosis* bacilli has been enriched, but prior to microbiological culture positivity. In such cases the efficiency of WGS may be influenced by the sample processing required for culture. The slow growth rate of *M. tuberculosis* makes it vulnerable to overgrowth by other microorganisms; to prevent this unwanted contamination, clinical specimens are treated with sodium hydroxide or acid to preferentially kill the unwanted microorganisms.[30] Due to their robust cell walls, some *M. tuberculosis* bacilli survive this process and following neutralization and concentration by centrifugation the samples are placed in broth or solid media for culturing.[33] Other reagents that assist sputum processing are *N*-acetyl-L-cysteine, which disrupts disulfide bonds,[34] and detergents such as sodium laurylsulphate.[30] Unfortunately, not all of the *M. tuberculosis* bacteria survive the decontamination process and considerable losses may occur at this stage if the process is not well controlled.[35] In addition, overgrowth by contaminants or no growth due to harsh sample pre-treatment may require the entire process to be repeated.

Additional methods and adjuncts have been used post-culture to enhance the isolation of *M. tuberculosis* gDNA. Bactericidal reagents such as isopropanol can assist by their ability to denature proteins and damage bacterial cell walls. Similarly, guanidine thiocyanate may be used where gDNA released is captured using activated silica.[36] Further purification and removal of proteins can be effected by phenol and chloroform extraction when, at the appropriate pH, DNA will partition into the aqueous layer.[37] It should be noted that bleach (sodium hypochlorite) should not be used, as although it will effectively thin sputum, it is highly detrimental to DNA.[38]

An extraction procedure optimized for maximum sensitivity has been developed by scientists investigating *M. tuberculosis* in archaeological specimens.[39] This incorporates enzyme digestion with Proteinase K and extended incubation in guanidium thiocyanate for up to 3 days at 56 °C. *N*-phenacylthiazolium bromide (PTB) is then used to cleave covalent cross-links to enable strand separation and amplification. Although effective for ancient samples, the technique also recommends using bead-beating and freeze/thaw steps that are less suitable for an analysis that requires DNA of high integrity.

4. Enrichment methods

A particular challenge for NGS sequencing from sputum is the presence of low numbers of bacteria amongst high levels of other materials. This includes large amounts of DNA of human origin and from other bacteria, which will react with sequencing reagents and may quench the reaction by out-competing the *M. tuberculosis* DNA. To enhance sensitivity, it is necessary to increase the proportion of *M. tuberculosis* DNA by a process of enrichment. However, any enrichment technique may run the risk of selection bias, which can alter the results. A potential way to reduce levels of human DNA is enzymatic degradation following differential lysis, in which the human cells are lysed but the bacilli remain intact.[40]

Doughty et al. evaluated water-induced osmotic lysis of human cells in *N*-acetyl-L-cysteine liquefied sputum. Following treatment with DNAase the samples were washed and heat-treated to remove the enzyme before the extraction of *M. tuberculosis* DNA was initiated.[41] Samples were sequenced using a shotgun approach, but success was limited. Although characteristic *M. tuberculosis* signals were identified, the depth of coverage obtained was insufficient for the detection of drug resistance-associated mutations with a high confidence of the nucleotide call.

An alternative approach is to capture either the bacteria or the bacterial DNA and remove them from the sample matrix. A simple approach is preferential binding to magnetic beads, and products are commercially available that utilize ligand-coated beads to bind bacteria.[42,43] A more sophisticated approach involves specific pull-down of *M. tuberculosis* DNA using complementary oligonucleotides. Brown et al. succeeded in sequencing *M. tuberculosis* from 24 smear-positive sputum specimens using a number of 120mer RNA baits that complement and span the *M. tuberculosis* genome.[18] The baits were labelled with biotin, thus enabling library preparation using the SureSelect^{XT} system (Agilent Technologies) followed by Illumina paired-end sequencing. Although successful, the authors point out that the requirement for highly specialized skills and equipment and estimated cost per sample of approximately US$350 may be a disadvantage, therefore other strategies need to be explored. A potential, less expensive approach is whole genome amplification, which has been applied to other bacteria with some success.[44,45] Whole genome multiple displacement amplification uses small random primers and high-fidelity DNA polymerase (phi29) to amplify large segments of DNA.[46] The technique requires good quality DNA, but its capacity to amplify the GC-rich regions of the *M. tuberculosis* genome has yet to be reported.

5. Sequencing from cultured bacteria

DNA extracted from cultured bacteria may be used in genome-wide association studies (GWAS) to identify putative resistance mutations. In addition, isolation and culturing of the bacteria from sputum samples serves as an effective enrichment strategy, where sequencing from cultures may provide information of resistance more rapidly than phenotypic testing methods.[47–49,59] To ensure sequence data of satisfactory quality with sufficient depth of reads, the quality and quantity of the extracted DNA should be checked before samples are submitted for library preparation and sequencing. Simple spectroscopic methods to measure absorption are not adequate as they frequently give misleading results due to small nucleic acids and contaminating chemicals. Quantity should be checked using targeted DNA-specific tests. To examine the integrity of the DNA, a sample should be examined by agarose electrophoresis (see **Supplementary Material** for further details).

Lessons learned from extracting cultured organisms may assist in the development of sputum processing methods. *M. tuberculosis* bacilli have a characteristic lipid-rich cell wall that is highly protective of the cell contents. Layers of peptidoglycan (cross-linked for additional stability), lipoglycans, mycolic acids, and large waxes form a robust barrier that must be overcome to allow the release of the gDNA. Methods to break open the cell wall include enzymatic, mechanical, and chemical approaches.[50,51] Enzymes may be used sequentially, with lysozymes to degrade the cell wall followed by proteinases, combined with solvent extraction to remove extracellular proteins.[52] Mechanical methods including bead-beating, sonication, freeze-fracturing, homogenizing, grinding (frozen samples), and high pressure have also been employed.[50,52] Kaser et al. reported optimum results using a bead-beater with beads of a small diameter (0.1 mm).[32] Unfortunately, these physical methods risk shearing the gDNA, which may affect sequencing quality.[32,50,53] Chemical methods include the use of detergents, and chaotropic agents such as guanidine thiocyanate have also been investigated.[32,36] Once released and purified, the bacterial DNA is highly vulnerable to degradation from naturally occurring enzymes and should be protected by use of DNAase-free reagents and the addition of metal chelating reagents such as ethylenediaminetetraacetic acid (EDTA). (Note: EDTA may be detrimental to some enzyme-based sequencing reactions.) gDNA may also be damaged by physical and environmental forces such as shearing, UV light, and high temperatures.

A DNA extraction protocol previously developed for restriction fragment length polymorphism (RFLP) analysis has been shown to consistently provide DNA of high molecular weight and sufficient quality for sequencing[54] (see **Supplementary Material** for full protocol). In this method, cells are heated to 80 °C before cooling and treating with lysozyme. A Proteinase K/detergent solution (sodium dodecyl sulphate) is then added and the solution incubated at 65 °C prior to treatment with cetyl trimethylammonium bromide (CTAB) to precipitate lipoglycans and polysaccharides,[55] thus assisting their removal along with unwanted proteins during organic solvent extraction (chloroform/isoamyl alcohol). This method of DNA purification has been used successfully for *M. tuberculosis* with the Illumina, Ion Torrent, and PacBio sequencing platforms. However, for use in a clinical laboratory, the DNA extraction method selected should be robust and time-efficient, and some improvement on the above protocol in this respect may expedite the uptake of WGS in a routine laboratory setting.

6. Conclusions

The emergence of resistance to multiple anti-TB drugs is an urgent public health problem.[56] The World Health Organization recommends at least five effective medicines be used for patients with rifampicin-resistant or multidrug-resistant TB (resistance to at least rifampicin and isoniazid).[57] Poor access to susceptibility tests often results in empiric treatment with a drug regimen that may not be effective if the bacteria have already developed resistance to some, or all of the drugs.

Considerable progress has been made towards determining resistance to anti-TB drugs by WGS, which would enable personalized and optimized treatment for patients. Sequencing should ideally be done directly from clinical samples to prevent the lengthy delays incurred during culture-based phenotypic testing, thus accelerating access to effective treatment. Although new sequencing platforms and analytical software are now available, there is as yet no consensus on the critical issue of sample preparation. The literature is rich with potential techniques and methodologies, but an inclusive systematic review of the topic is lacking. Traditional techniques such as bead-beating and sonication that are used successfully in conjunction with PCR-based detection may not be appropriate for sequencing protocols in which high molecular mass gDNA is required.[32] There have been few attempts to optimize extraction for *M. tuberculosis* WGS or targeted NGS for detecting mutations that are presumptively associated with drug resistance, and evidence of the impact of extraction tools on sequence quality is currently insufficient.

Whilst other aspects of WGS, such as determining phylogenetic strain lineage, virulence, etc., are essential in understanding the transmission and pathogenesis of TB, it is detecting drug resistance directly from clinical samples that warrants priority. Given the scope of the topic and the urgency of the clinical and public health need, consideration should be given to a coordinated collaborative approach to reduce duplication and fast-track progress. Precedence for such an initiative has previously been established through the creation of the ReSeq TB Consortium by which a comprehensive database of drug resistance-associated mutations

is being developed though international collaboration.[58] The TB research community is urged to follow this example and work together to identify methodological algorithms that will allow rapid comprehensive testing for drug resistance by WGS. Accelerated access to effective treatment for patients with multiple and extensively drug-resistant disease would reduce morbidity and mortality, improve patient care and the familial burden, and may also reduce opportunities for transmission. Thus a consolidated approach that is affordable to the regions where drug-resistant TB is endemic, towards the development of standardized, optimized DNA extraction protocols, will benefit both patients and efforts to control this deadly disease.

Acknowledgements

T.G.C. is supported by the Medical Research Council UK (grant number MR/K000551/1, MR/M01360X/1, MR/N010469/1, MC_PC_15103). SC is funded by the Medical Research Council UK (grant number MR/M01360X/1, MC_PC_15103). K.D. is supported by the EDCTP, South African MRC, and the NIH. The study sponsors had no role in the study design, in the collection, analysis, and interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication.

Ethical approval: Ethical approval was not required for this work.

Conflict of interest: All authors declare no conflict of interest.

Author contributions: RM and MS conceived the study. RM, TC, SC, CR, DD, LS, AC, KD, and MS contributed to the writing of the review and reviewed the final draft.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijid.2016.11.422.

References

1. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998;**393**:537–44.
2. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. Alignment of whole genomes. *Nucleic Acids Res* 1999;**27**:2369–76.
3. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* 2002;**184**:5479–90.
4. Brites D, Gagneux S. Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunol Rev* 2015;**264**:6–24.
5. Phelan J, Coll F, McNerney R, Ascher DB, Pires DE, Furnham N, et al. *Mycobacterium tuberculosis* whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance. *BMC Med* 2016;**14**:31.
6. Guerra-Assuncao JA, Crampin AC, Houben RM, Mzembe T, Mallard K, Coll F, et al. Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife* 2015;**4**:e05166.
7. Glynn JR, Guerra-Assuncao JA, Houben RM, Sichali L, Mzembe T, Mwaungulu LK, 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**:e0132840.
8. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 2014;**46**:279–86.
9. Farhat MR, Sultana R, Iartchouk O, Bozeman S, Galagan J, Sisk P, et al. Genetic determinants of drug resistance in *Mycobacterium tuberculosis* and their diagnostic value. *Am J Respir Crit Care Med* 2016;**194**:621–30.
10. Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Med* 2016;**14**:21.
11. Faksri K, Tan JH, Chaiprasert A, Teo YY, Ong RT. Bioinformatics tools and databases for whole genome sequence analysis of *Mycobacterium tuberculosis*. *Infect Genet Evol* 2016;**45**:359–68.
12. Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* 2015;**7**:51.
13. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat Commun* 2015;**6**:10063.
14. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web tool delineating Mycobacterium tuberculosis antibiotic resistance and lineage from whole-genome sequencing data. *J Clin Microbiol* 2015;**53**:1908–14.
15. Iwai H, Kato-Miyazawa M, Kirikae T, Miyoshi-Akiyama T. CASTB (the comprehensive analysis server for the Mycobacterium tuberculosis complex): a publicly accessible Web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. *Tuberculosis (Edinb)* 2015;**95**:843–4.
16. Witney AA, Gould KA, Arnold A, Coleman D, Delgado R, Dhillon J. Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases. *J Clin Microbiol* 2015;**53**:1473–83.
17. Colman RE, Anderson J, Lemmer D, Lehmkuhl E, Georghiou SB, Heaton H, et al. Rapid drug susceptibility testing of drug-resistant Mycobacterium tuberculosis isolates directly from clinical samples by use of amplicon sequencing: a proof-of-concept study. *J Clin Microbiol* 2016;**54**:2058–67.
18. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZ, et al. Rapid whole-genome sequencing of Mycobacterium tuberculosis isolates directly from clinical samples. *J Clin Microbiol* 2015;**53**:2230–7.
19. Tyler AD, Christianson S, Knox NC, Mabon P, Wolfe J, Van Domselaar G, et al. Comparison of sample preparation methods used for the next-generation sequencing of Mycobacterium tuberculosis. *PLoS One* 2016;**11**:e0148676.
20. Daum LT, Rodriguez JD, Worthy SA, Ismail NA, Omar SV, Dreyer AW, et al. Next-generation ion torrent sequencing of drug resistance mutations in Mycobacterium tuberculosis strains. *J Clin Microbiol* 2012;**50**:3831–7.
21. Phelan J, O'Sullivan DM, Machado D, Ramos J, Whale AS, O'Grady J, et al. The reliability of whole genome sequencing technology for detecting resistance to anti-tuberculous drugs. *Genome Med* 2016:
22. Illumina. Solexa technology. Illumina; 2016. Available at: http://www.illumina.com/technology/next-generation-sequencing/solexa-technology.html (accessed September 20, 2016).
23. Thermo Fisher Scientific. Ion Torrent™ next-generation sequencing technology. Available at: https://www.thermofisher.com/za/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html (accessed September 20, 2016).
24. Pacific Biosciences. SMRT sequencing. Available at: http://www.pacb.com/smrt-science/smrt-sequencing/ (accessed September 20, 2016).
25. Feng Y, Zhang Y, Ying C, Wang D, Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* 2015;**13**:4–16.
26. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 2016;**72**:104–14.
27. McNerney R. Advice on samples for tuberculosis testing. You Tube; 2010.
28. Weiszhar Z, Horvath I. Induced sputum analysis: step by step. *Breathe* 2013;**9**:300–6.
29. Hernandez P, Punchak M, Camacho M, Hepple P, McNerney R. Investigating the quality of expectorated sputum for tuberculosis diagnosis in Bolivia. *Int J Tuberc Lung Dis* 2015;**19**:1065–7.
30. Groothuis DG, Yates MD. Diagnostic and public health mycobacteriology, 2nd ed, London: Bureau of Hygiene and tropical Diseases; 1991.
31. Hadad DJ, Morais CG, Vinhas SA, Fennelly KP, Dietze R, Nascimento CP, et al. Evaluation of processing methods to equitably aliquot sputa for mycobacterial testing. *J Clin Microbiol* 2012;**50**:1440–2.
32. Kaser M, Ruf MT, Hauser J, Marsollier L, Pluschke G. Optimized method for preparation of DNA from pathogenic and environmental mycobacteria. *Appl Environ Microbiol* 2009;**75**:414–8.
33. Stinson KW, Eisenach KD, Kayes S, Matsumoto M, Siddiqi S, Nakashima S, et al. Mycobacteriology laboratory manual, 1st ed, Stop TB Partnership; 2014.
34. Kubica GP, Dye WE, Cohn ML, Middlebrook G. Sputum digestion and decontamination with *N*-acetyl-L-cysteine–sodium hydroxide for culture of mycobacteria. *Am Rev Respir Dis* 1963;**87**:775–9.
35. Chatterjee M, Bhattacharya S, Karak K, Dastidar SG. Effects of different methods of decontamination for successful cultivation of *Mycobacterium tuberculosis*. *Indian J Med Res* 2013;**138**:541–8.
36. Wang TY, Wang L, Zhang JH, Dong WH. A simplified universal genomic DNA extraction protocol suitable for PCR. *Genet Mol Res* 2011;**10**:519–25.
37. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press; 1989.
38. Hayatsu H, Pan S, Ukita T. Reaction of sodium hypochlorite with nucleic acids and their constituents. *Chem Pharm Bull (Tokyo)* 1971;**19**:2189–92.
39. Donoghue HD, Spigelman M, O'Grady J, Szikossy I, Pap I, Lee OY, et al. Ancient DNA analysis—an established technique in charting the evolution of tuberculosis and leprosy. *Tuberculosis (Edinb)* 2015;**95**(Suppl 1):S140–4.
40. Lim YW, Schmieder R, Haynes M, Willner D, Furlan M, Youle M, et al. Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J Cyst Fibros* 2013;**12**:154–64.
41. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ* 2014;**2**:e585.
42. DiaSorin Ireland Ltd. Bugs'n Beads Kit—instructions for use. DiaSorin Ireland Ltd; 2012. Available at: http://uk.werfen.com/~/media/il%20uk/docs/diasorin/

ifu%20120102%20%20%20bugs%20n%20beads%20english.pdf (accessed December 10, 2016).
43. Wilson S, Lane A, Rosedale R, Stanley C. Concentration of Mycobacterium tuberculosis from sputum using ligand-coated magnetic beads. *Int J Tuberc Lung Dis* 2010;**14**:1164–8.
44. Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS One* 2009;**4**:e6864.
45. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo MJ, Dupont CL, Badger JH, et al. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotechnol* 2011;**29**:915–21.
46. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem A, Liebaers I, et al. Whole-genome multiple displacement amplification from single cells. *Nat Protoc* 2006;**1**:1965–70.
47. Martinez E, Bustamante A, Menon R, Wang Q, Jelfs P, Marais B, et al. Whole-genome sequencing of *Mycobacterium tuberculosis* for rapid diagnostics: feasibility of a decentralised model. *Lancet Respir Med* 2016;**4**:e13–4.
48. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med* 2016;**4**:49–58.
49. Koser CU, Bryant JM, Becq J, Torok ME, Ellington MJ, Marti-Renom MA, et al. Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *N Engl J Med* 2013;**369**:290–2.
50. Belisle JT, Sonnenberg MG. Isolation of genomic DNA from mycobacteria. In: Parish T, Stoker NG, editors. *Mycobacteria Protocols*. New Jersey, USA: Humana Press Inc; 1998. p. 31–44.
51. Zhang ZQ, Ishaque M. Evaluation of methods for isolation of DNA from slowly and rapidly growing mycobacteria. *Int J Lepr Other Mycobact Dis* 1997;**65**:469–76.
52. van Helden PD, Victor TC, Warren RM, van Helden EG. Isolation of DNA from *Mycobacterium tuberculosis*. *Methods Mol Med* 2001;**54**:19–30.
53. Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 2012;**7**:e33865.
54. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;**31**:406–9.
55. Baess I. Isolation and purification of deoxyribonucleic acid from mycobacteria. *Acta Pathol Microbiol Scand B Microbiol Immunol* 1974;**82**:780–4.
56. Dheda K, Gumbo T, Gandhi NR, Murray M, Theron G, Udwadia Z, et al. Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis. *Lancet Respir Medicine* 2014;**2**:321–38.
57. World Health Organization. WHO treatment guidelines for drug-resistant tuberculosis 2016 update. Geneva, Switzerland: WHO; 2016.
58. Schito M, Dolinger DL. A collaborative approach for "ReSeq-ing" Mycobacterium tuberculosis drug resistance: convergence for drug and diagnostic developers. *EBioMedicine* 2015;**2**:1262–5.
59. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, et al. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. *J Clin Microbiol* 2015;**53**:1137–43.