Review

Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing

T.M. Walker 1,*, M. Merker 2, T.A. Kohl 2, D.W. Crook 1,3, S. Niemann 2,4, T.E.A. Peto 1,3

1) Department of Microbiology and Infectious Diseases, Nuffield Department of Medicine, University of Oxford, Oxford, UK
2) Molecular Mycobacteriology, Forschungszentrum Borstel, Leibniz-Zentrum für Medizin und Biowissenschaften, Borstel, Germany
3) National Institute of Health Oxford Biomedical Research Centre, University of Oxford, John Radcliffe Hospital, Oxford, UK
4) German Center for Infection Research, Borstel Site, Borstel, Germany

ARTICLE INFO

Article history:
Received 5 August 2016
Received in revised form
11 October 2016
Accepted 12 October 2016
Available online xxx

Editor: F. Allerberger

Keywords:
Drug resistance
Extensive drug resistance
Multi drug resistance
Surveillance
Susceptibility
Tuberculosis
Whole genome sequencing

ABSTRACT

Whole genome sequencing (WGS) can help to relate Mycobacterium tuberculosis genomes to one another to assess genetic relatedness and infer the likelihood of transmission between cases. The same sequence data are now increasingly being used to predict drug resistance and susceptibility. Controlling the spread of tuberculosis and providing patients with the correct treatment are central to the World Health Organization’s target to ‘End TB’ by 2035, for which the global prevalence of drug-resistant tuberculosis remains one of the main obstacles to success. So far, WGS has been applied largely to drug-susceptible strains for the purposes of understanding transmission, leaving a number of analytical considerations before transferring what has been learnt from drug-susceptible disease to drug-resistant tuberculosis. We discuss these potential problems here, alongside some of the challenges to characterizing the Mycobacterium tuberculosis ‘resistome’—the optimal knowledge-base required for WGS-based assays to successfully direct individualized treatment regimens through the prediction of drug resistance and susceptibility in the future. T.M. Walker, CMI 2016;1

Introduction

In 2014 the World Health Organization announced its ‘End TB’ strategy. A target was set to reduce global tuberculosis incidence by 90% on 2015 rates by 2035, corresponding to fewer than ten new cases per 100 000 population per year. ‘Intensified research and innovation’ was pronounced one of three pillars upon which to build success. There was a call for ‘new tools’ to be ready by 2025, including a vaccine, better prophylaxis and treatment regimens, and a point of care test. The target was endorsed by all member states at the 2014 World Health Assembly [1].

As a ‘new tool’, whole genome sequencing (WGS) is one of the most exciting and disruptive technologies to take to the stage for decades [2], with advances in nanotechnology now promising to deliver affordable and accessible sequencing platforms that are simple to use, near the patient [3]. These developments are perhaps more important to the management of tuberculosis than any other infectious disease due to the diagnostic delays stemming from the slow growth rate of the Mycobacterium tuberculosis complex. The introduction of targeted nucleic acid amplification-based technologies has only partially compensated for this slow growth rate, notifying clinicians which drugs to avoid, but not which drugs to give to the patient [4]. Consequently, the culture-derived phenotype remains the reference standard where practicable and affordable. Historically, this has left many low-income settings relying on smear microscopy as the only available diagnostic test, although the roll-out of the Xpert MTB/RIF more recently has allowed a putative, but helpful, distinction between drug-susceptible and multidrug-resistant (MDR, defined as resistance to at least isoniazid and rifampicin) disease to be drawn [5]. The consequences of empirical or misguided tuberculosis treatment can be disastrous both at a population, and at an individual level. Under-recognizing and under-treating MDR or extensively drug-resistant (XDR) tuberculosis at a programmatic level risks...

* Corresponding author. Timothy M. Walker, Nuffield Department of Medicine, University of Oxford, Level 7, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK.
E-mail address: timothy.walker@ndm.ox.ac.uk (T.M. Walker).

http://dx.doi.org/10.1016/j.cmi.2016.10.014
1198-743X/© 2016 European Society of Clinical Microbiology and Infectious Diseases. Published by Elsevier Ltd. All rights reserved.

Please cite this article in press as: Walker TM, et al., Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing, Clinical Microbiology and Infection (2016), http://dx.doi.org/10.1016/j.cmi.2016.10.014
sustaining it in a population through transmission, whereas treating individual tuberculosis patients with an insufficient number of effective drugs risks selecting for resistant subpopulations, creating resistant strains *de novo*, or further amplifying the number of drugs ineffective against an already M/XDR strain. With 480 000 cases of MDR tuberculosis in 2014 already [6], work towards a tool for the rapid and accurate diagnosis and control of drug resistance is an absolute priority. Here we review the prospects of WGS-based technologies providing solutions to both challenges.

**Tuberculosis surveillance**

The 1994 WHO DOTS (Directly Observed Treatment, Short course) strategy called for the gathering and quarterly reporting of data on cases and outcomes to help optimize control measures and treatment programmes as epidemiological trends change [7]. Because it is widely considered that the early detection of outbreaks, and the identification, isolation and treatment of incident active cases are of value both to the individual concerned and their wider community, many high-income countries have simultaneously engaged in enhanced surveillance around active cases with the aim of offering chemoprophylaxis to latently infected contacts and treatment to those with active disease [8]. Although the WHO specifically recommends investigating household contacts of patients with infectious MDR tuberculosis in low-income and middle-income countries as well, this is largely on an *a priori* basis, as the WHO admits the supporting evidence is poor [9]. With resources often scarce, these recommendations are infrequently followed.

Contact investigations have been aided by classical molecular fingerprinting techniques in high-income countries over the past 25 years. These have included insertion sequence (IS)6110 typing, spoligotyping and, latterly, mycobacterial interspersed repetitive unit—variable number tandem repeat typing, each of which quantifies the copy number of relatively stable repetitive segments of DNA in the *M. tuberculosis* complex genome [10]. By dividing culture-confirmed cases into discrete groups, typing techniques have allowed cluster investigations to search for epidemiological links between cases with similar fingerprints, even though similar fingerprints by themselves are insufficient grounds to expect a link. The well-documented advantages of WGS data are that they provide greater resolution than traditional techniques [11—13], allowing contact investigations to be directed with greater specificity. Despite the obvious potential for cost saving, there has been no study to date to formally assess the impact of WGS on tuberculosis control in general, or on M/XDR disease in particular. Moreover, options for public-health interventions in M/XDR outbreaks are potentially fewer than for drug-susceptible disease, with the precise role for chemoprophylaxis remaining controversial, not least because M/XDR tuberculosis is a heterogeneous entity that can be resistant to many different drugs [14].

How WGS might help in low- and medium-income settings has also still to be defined and assessed. Intuitively, constrained resources would most sensibly be targeted at transmission of M/XDR disease with a view to active case finding. Because drug resistance can be predicted from the genome sequence before phenotypic results becoming available, transmission of M/XDR strains could be detected early [15], potentially facilitating more rapid public health responses, including active case finding. However, although it is expected that WGS will help to focus cluster investigations on the most relevant cases, it remains an assumption that this will lead to less transmission of disease and improved health outcomes.

To maximize the potential benefit of WGS for surveillance of M/XDR tuberculosis, what has been learned from drug-susceptible disease must be accurately applied to and adjusted for the peculiarities of M/XDR disease. This includes the interpretation of genomic distances, and the understanding of mutation rates, lineage effects and host-dependent factors such as human immuno-deficiency virus (HIV) infection.

Approaches to WGS data analysis will no doubt be further honed for maximum epidemiological gain, but to date there has been a remarkable convergence of results emanating from different research groups, using a diversity of analytical techniques. In the vast majority of cases direct or recent transmission has been associated with only a handful of nucleotide differences separating samples [11,12,16—18]. Pragmatic single nucleotide polymorphism (SNP) thresholds have therefore been proposed to help guide cluster investigations [16], helping to accommodate some variability that has been observed in the molecular clock over short time periods [17]. It has also been argued in multiple independent reports that phylogenetic signals could be used to infer the direction of transmission between patients [16,19—21], providing further advantage over one-dimensional fingerprinting results by helping to target investigations at the contacts of the most infectious patients. Each of these findings has contributed to the decision to implement WGS into routine practice in England, and towards similar plans in other high-income countries [22]. The contribution towards tuberculosis control will be assessed as a matter of process in these countries where the incidence and rates of M/XDR tuberculosis tend to be low, but from which lessons for the control of M/XDR disease may be learnt.

The idea of a consistent molecular clock is a convenient generalization across genomic loci and across lineages within the *M. tuberculosis* complex, but there are reasons why M/XDR disease might be considered separately due to factors potentially accelerating the clock. Genomic loci under positive selection pressure, including those conferring drug resistance or compensating for the fitness cost of resistance, have a higher than background mutation rate [4]. The phenomenon of ‘hitch-hiking’ SNPs has also been described, where SNPs involved in drug resistance co-emerge with others that play no ostensive role in resistance [23]. Loci like these with higher mutation rates may result in a greater than expected number of SNPs being observed between patients where resistance is acquired. The problem may be compounded where patients suffering prolonged infection develop divergent microevolution at different anatomical sites, possibly influenced by exposure to treatment, and potentially resulting in greater within-host pathogen diversity than has come to be expected (Fig. 1) [16,24]. Inferences about transmission must therefore also take patient sampling into account. Each of these factors has the potential to alter the expected number of SNPs between patients potentially linked by recent M/XDR transmission.

One approach to mitigating the effect of a faster molecular clock at loci that are under positive selection pressure has been to leave out sites associated with drug resistance when assessing transmission [11]. This could allow interpretative parameters derived from drug-susceptible disease to be applied to M/XDR disease. However, which sites to leave out is not obvious, as many mutations are only putatively associated with resistance [4], whereas other sites under selection may be compensatory mutations, ‘hitch-hiking’ SNPs, or those under selection from immune pressure. Studies focusing on the transmission of M/XDR strains are therefore urgently needed to explore whether the phenomena described here are widespread, and whether the experience from patients with drug-susceptible tuberculosis can be applied.

Lineage and host effects may prove to be further confounders. Differential rates of emergence of drug-resistance mutations have been documented between the East Asian (Beijing) and Euro-American lineages [25]. The Beijing lineage has been linked to the emergence of MDR disease in Eurasia over the past 30 years, as well as to highly resistant clones in South Africa where its success has
been linked to the emergence of the HIV pandemic [26–28]. How HIV has contributed to the success of MDR tuberculosis remains unclear. One hypothesis is that HIV treatment influences the pathogen mutation rate, but this is both unproven and could not account for events preceding the mass-treatment era or HIV pandemic [29]. Indeed, it has been argued that Latin-American-Mediterranean (LAM4) lineage strains from the Tugela Ferry outbreak in South Africa acquired drug resistance well before the HIV pandemic [29]. An alternative hypothesis argues that drug-resistant strains manage to successfully infect patients living with HIV despite potentially bearing some fitness costs, thereby reducing purifying selection and accelerating microevolution and the acquisition of drug resistance [30–32]. However, a recent study of a prolonged outbreak of MDR tuberculosis in South America did not demonstrate an HIV-effect on the pathogen mutation rate [33].

For the robust comparison of sequenced M/XDR strains, including across jurisdictions, standardized data analysis pipelines will be necessary. Consideration will have to be given to the acquisition of drug resistance, lineage, and maybe even to host factors if the time to the most recent common ancestor is to be accurately predicted for potentially clustered M/XDR strains. One approach to achieving standardization in general, rather than for M/XDR tuberculosis in particular, has been to construct an easily comparable genetic bar code based on gene alleles, at the expense of some resolution [21]. Achieving comparability of data without the loss of resolution, however, remains a preferable option and a research objective. Whatever the eventual solution, more work will be needed to assess the universality of previously defined criteria for identifying recent transmission, be that for outbreaks with evolving drug resistance or for outbreaks of circulating clones with established M/XDR patterns [28,34]. The future contribution of WGS towards the control of M/XDR tuberculosis, and thereby to the elimination of tuberculosis altogether, therefore remains uncertain, even though there are reasons to be confident.

**Predicting drug resistance and susceptibility**

It is estimated that only 25% of the 480 000 worldwide cases of MDR-TB were detected and reported in 2014, largely because the costs of laboratory-based phenotyping are beyond many low-income settings [6]. That proportion has been increasing since the WHO endorsed the Xpert MTB/RIF in 2011 [35], but as with all other molecular assays based on targeted amplification [5], the Xpert MTB/RIF predicts resistance, and not susceptibility, informing a clinician which drugs to avoid, rather than which drugs to give. With the WHO’s target to end tuberculosis by 2035, research activity into how WGS could form the basis of a new diagnostic test that predicts both resistance and susceptibility, while differentiating between newly evolved and transmitted M/XDR tuberculosis, is now increasing (Fig. 2) [36].

Much is already known about the molecular determinants of resistance to the first-line drugs, including isoniazid and rifampicin, whereas the determinants of resistance to pyrazinamide and third- and fourth-line drugs are less well described [4,37]. Even once the range of determinants of drug resistance is more completely understood, the contribution of WGS to achieving the WHO target will depend on whether personalized treatment regimens can deliver improved outcomes for M/XDR patients, at a programmatic level, when compared with ‘one-size-fits-all’ treatment protocols such as the recently WHO endorsed ‘Bangladesh regimen’ [38]. With current technologies, implementation of personalized treatment regimens will remain difficult in many low-income settings where phenotypic drug susceptibility testing remains unavailable due a lack of infrastructure and trained staff [6]. A paucity of doctors

---

**Fig. 1.** Factors influencing molecular clock and genetic diversity in multidrug/extensively drug-resistant (M/XDR) tuberculosis. (a, b) Population-based effects. (a) Phylogenetic tree of the *Mycobacterium tuberculosis* complex lineages with the ‘Beijing’ lineage, purported to have a higher mutation rate, coloured red. (b) Cartoon showing the emergence of the same example mutation independently on different branches of the tree (‘homoplasies’) due to positive selection. (c, d) Evolution within the individual. (c) Cartoon of the circular *M. tuberculosis* complex genome/ chromosome in which a single drug resistance mutation (in red) arises and is accompanied by compensatory or hitchhiking mutations (in blue), accelerating the rate of microevolution. (d) Cartoon of the evolution of *M. tuberculosis* complex strains within two hosts, (i) and (ii), infected by the same source at the same time point, 0. As patient (i) is not diagnosed until time point t, there has been opportunity for greater within-host evolution, and for the spontaneous emergence of M/XDR conferring mutations. Conclusions about the relatedness to case (ii) will be influenced by which strain is sampled from (i), and when.
further compounds the problem in these settings as bespoke regimens, by their very nature, require design. However, portable WGS platforms provide an opportunity to translate expanding knowledge of the molecular determinants of resistance into field technology that could circumvent the need for complex infrastructure [3]. It is, however, possible to foresee how data analysis platforms could be combined with suitable algorithms to define personalized regimens, reducing dependence on medically qualified staff.

Individualization of therapy based on WGS for the treatment of M/XDR disease will require an extensive knowledgebase characterizing the effect of genomic mutations on all relevant drugs. However, a clear understanding of what the molecular-level data are predicting is essential. WGS studies to date have attempted to link particular mutations to MICs [39], or to a probability of one of two outcomes—phenotypic resistance or susceptibility [37]. Analytical approaches to large data sets have ranged from the heuristic to the mathematically principled [4,37,40,41], whereas laboratory approaches have ranged from passage and selection of spontaneous mutations to targeted mutagenesis [42,43]. However, although some mutations have been identified as predictive of the in vitro phenotype [4], this is itself only a proxy for overall clinical outcome. The aspiration would be for molecular data to predict outcome directly, as challenging as this may prove to be (Fig. 3). The clinical implications of some mutations known to lead to high-level isoniazid (katG315T), rifampicin (rpoBS450L) and fluoroquinolone (gyrA94) resistance and treatment failure have recently been outlined in a consensus statement [44]. The clinical significance of other mutations conferring lower-level resistance, such as inhA promoter region mutations, and at rpoBH445L, is less clear [44,45]. However, although phenotypes themselves are established predictors of outcome in some cases [46], in many cases such a correlation remains to be established despite this being ostensibly a lesser challenge than linking individual mutations to outcome [47]. Not only have phenotypic data been collected over a longer time period, they are often presented in binary or ternary terms (susceptible/low-level resistance/high-level resistance), rather than as

---

**Table: Comparison of currently available diagnostic modalities to nanotechnology-based whole-genome sequencing.**

<table>
<thead>
<tr>
<th></th>
<th>Smear microscopy</th>
<th>Xpert MTB/RIF</th>
<th>Phenotyping</th>
<th>WGS (nanotechnology)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity</strong></td>
<td>Limit of detection 10^9 cfu/ml</td>
<td>Limit of detection 10^9 cfu/ml</td>
<td>Limit of detection 10 cfu/ml</td>
<td>Limit of detection falling</td>
</tr>
<tr>
<td><strong>Turnaround time</strong></td>
<td>Same day</td>
<td>Same day</td>
<td>Weeks</td>
<td>Promise of same day</td>
</tr>
<tr>
<td><strong>Clinical information</strong></td>
<td>First line treatment given if acid fast bacilli seen</td>
<td>First line treatment given on detection of Mtb DNA</td>
<td>Treatment individualised</td>
<td>Treatment individualised</td>
</tr>
<tr>
<td></td>
<td>Low cost per smear</td>
<td>Subsidised to ~ $10 in priority high-burden countries</td>
<td>High capital investment costs for laboratory and staff training costs</td>
<td>Potentially no need for a laboratory or specialist staff</td>
</tr>
<tr>
<td><strong>Skills base required</strong></td>
<td>Operator dependent</td>
<td>Operator independent</td>
<td>Operator dependent</td>
<td>Operator independent</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Comparison of currently available diagnostic modalities to nanotechnology-based whole-genome sequencing.

**Fig. 3.** Advantages and remaining challenges of whole genome sequencing for surveillance and susceptibility testing.

---

Please cite this article in press as: Walker TM, et al., Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing, Clinical Microbiology and Infection (2016), http://dx.doi.org/10.1016/j.cmi.2016.10.014
quantitative data [48]. Linking the individual or combined role of a plethora of mutations to clinical outcome will be more complicated. A significant amount of resistance to anti-tuberculosis drugs can be reduced to the effects of single mutations in the form of SNPs and small insertions or deletions [4]. This holds true for other members of the *M. tuberculosis* complex such as *Mycobacterium africanum*, although similar mutations have been observed to arise with different frequencies [49]. More complex mechanisms of resistance have also been described, with compensatory mutations mitigating the fitness costs of resistance-conferring mutations [50], and the effects of multiple mutations cumulatively leading to stepwise increases in MIC [51]. Mutations can also be antagonistic (epistatic), where the pre-existence of one mutation cancels out the fitness costs of resistance-conferring mutations [50], even as essential research and development remains in progress over the coming years (Fig. 3). However, the community will have to eventually converge on a standardized approach to data analysis to ensure maximum comparability. The likely impact on preventing transmission of M/XDR tuberculosis, and on improving diagnostic speed and precision, needs to be widely assessed. However, the experience of the Xpert MTB/RIF assay roll-out underlines the need to strengthen health systems in parallel with the implementation of any new test if the 2035 target is to be met [59].

**Transparency declaration**

None of the authors have any conflict of interest. TMW is a National Institute of Health Research (NIHR) Academic Clinical Lecturer and DWC and TEAP are NIHR Senior Investigators supported by the NIHR Oxford Biomedical Research Centre (BRC) programme. SN is a consultant to FIND. All authors contributed to the design and writing of this review.

**References**


Please cite this article in press as: Walker TM, et al., Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing, Clinical Microbiology and Infection (2016) 1–6, http://dx.doi.org/10.1016/j.cmi.2016.10.014.