Clinical Microbiology and Infection xxx (2017) 1–6

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Narrative review

Mycobacterium tuberculosis and whole-genome sequencing: how close are we to unleashing its full potential?

G. Satta [1, 2, *], M. Lipman [3, 4], G.P. Smith [5, 6], C. Arnold [1, 7], O.M. Kon [2, 8], T.D. McHugh [1]

[1] *UCL-TB and UCL Centre for Clinical Microbiology, Department of Infection, University College London, UK*
[2] *Imperial College Healthcare NHS Trust, London, UK*
[3] *UCL-TB and UCL Respiratory, University College London, UK*
[4] *Royal Free London NHS Foundation Trust, London, UK*
[5] *National Mycobacterium Reference Laboratory, Public Health England, UK*
[6] *Heart of England NHS Foundation Trust, Birmingham, UK*
[7] *Genomic Services and Development Unit, Public Health England, UK*
[8] *National Heart and Lung Institute, Imperial College London, UK*

ARTICLE INFO

Article history:
Received 17 September 2017
Received in revised form
21 October 2017
Accepted 30 October 2017
Available online xxx

Editor: L. Leibovici

Keywords:
Applications
Drug resistance
Mycobacterium tuberculosis
Whole genome sequencing

ABSTRACT

Background: Nearly two decades after completion of the genome sequence of *Mycobacterium tuberculosis* (MTB), and with the advent of next generation sequencing technologies (NGS), whole-genome sequencing (WGS) has been applied to a wide range of clinical scenarios. Starting in 2017, England is the first country in the world to pioneer its use on a national scale for the diagnosis of tuberculosis, detection of drug resistance, and typing of MTB.
Aims: This narrative review critically analyses the current applications of WGS for MTB and explains how close we are to realizing its full potential as a diagnostic, epidemiologic, and research tool.
Sources: We searched for reports (both original articles and reviews) published in English up to 31 May 2017, with combinations of the following keywords: whole-genome sequencing, Mycobacterium, and tuberculosis. MEDLINE, Embase, and Scopus were used as search engines. We included articles that covered different aspects of whole-genome sequencing in relation to MTB.
Content: This review focuses on three main themes: the role of WGS for the prediction of drug susceptibility, MTB outbreak investigation and genetic diversity, and research applications of NGS.
Implications: Many of the original expectations have been accomplished, and we believe that with its unprecedented sensitivity and power, WGS has the potential to address many unanswered questions in the near future. However, caution is still needed when interpreting WGS data as there are some important limitations to be aware of, from correct interpretation of drug susceptibilities to the bioinformatic support needed. **G. Satta, Clin Microbiol Infect 2017;▪:1**

Introduction

The complete genome sequence of *Mycobacterium tuberculosis* (MTB) was described in 1998 [1]. Since then, whole-genome sequencing (WGS) has been applied to a wide range of clinical scenarios, with the potential to revolutionize tuberculosis (TB) diagnosis, outbreak investigation, drug and vaccine development, plus assist in understanding MTB evolution and pathogenicity. In 2017, Public Health England (PHE) introduced routine WGS in the clinical setting of the National Health Service (NHS); and England will be the first country in the world to pioneer its use at a population level for the diagnosis, detection of drug resistance, and typing of MTB [2] (Scotland, Wales, and Northern Ireland are not included yet).

The main aim of this narrative review is to summarize and critically analyse the current literature relating to WGS and MTB as a diagnostic, epidemiologic, and research tool; and determine to what extent we have been able to realize its full potential (search criteria are specified in Box 1). This review will provide the reader

* Corresponding author. G. Satta, University College London, Centre for Clinical Microbiology, Department of Infection, Consultant in Medical Microbiology and Infectious Diseases – Imperial College Healthcare NHS Trust, London, UK.
E-mail address: Giovanni.satta@nhs.net (G. Satta).

https://doi.org/10.1016/j.cmi.2017.10.030
1198-743X/

Box 1
Search strategy and selection criteria

We searched for reports (both original articles and reviews) published in English up to 31 May 2017, with combinations of the following keywords: whole-genome sequencing, Mycobacterium, and tuberculosis. MEDLINE, Embase, and Scopus were used as search engines. We reviewed the articles resulting from these searches and the relevant references cited in them. Additional search using an internet search engine was performed to potentially identify some additional papers. As this is a narrative review, we considered it most appropriate to focus on three main themes (prediction of susceptibility, outbreak investigation, and research applications) and included articles that reflected this. Inevitably, we did not include some papers that were not relevant to the main themes and that did not provide any additional information.

with an overview of the current applications (and limitations) of WGS, with a specific focus on three main themes: prediction of drug susceptibility, outbreak investigation and genetic diversity, and research applications (pathogenesis, drug discovery, and vaccine development) (Table 1). It will interpret the present situation and provide guidance on the future direction of development.

Prediction of drug susceptibility and resistance

Extensive work on WGS and susceptibility testing has been led by the Wellcome Trust/University of Oxford (and collaborators), with the demonstration that this approach can be integrated into routine diagnostic workflows, with data generated in 9 days (compared with an average of 3 weeks for conventional susceptibility methods) (Fig. 1) and at a similar or cheaper price than the current diagnostic workflows [3,4]. A recent systematic review [5] on the use of WGS for the detection of drug resistance in MTB highlighted its role as a promising alternative to existing phenotypic and molecular drug susceptibility testing methods. In particular, this high sensitivity and specificity has been demonstrated for the first-line drugs isoniazid (INH) and rifampicin (RIF). However, there is still significant variation with the remaining first- and second-line drugs. This results from the complexity of genetic mechanisms conferring resistance, plus in some cases resistance driven by non-specific mechanisms such as efflux pumps [6]. This may be of considerable importance when using WGS to guide clinical decisions. The lack of standardization for some phenotypic tests, in particular pyrazinamide, also confounds the identification of resistance in clinical isolates and mutants [7].

The implementation of WGS in the clinical setting is hindered by some significant limitations. Not least of these is the requirement for a culture sample before DNA isolation and sequencing. Direct WGS from clinical samples has been successfully demonstrated [8–10]; although all current protocols for clinical practice still require a culture step. Additionally, before switching completely to WGS testing, a long transition period with both WGS and conventional culture susceptibility methods used in parallel will be needed to clarify the current discrepancies between genotype and phenotype and to satisfy the need for large datasets to better understand the role of rare resistance mechanisms, and the level of resistance conferred by different mutations.

On the other hand, major progress has already been made in terms of data interpretation and standardization. Several online software tools, all-in-one and easy to use, are available for rapid interpretation of WGS data in MTB [11] and various online databases (ReSeqTB [12], TB Portal [13], GenTB [14]) offer support to share, interpret, and link genetic results to drug resistance phenotype and other epidemiological variables, providing international harmonization of such data. At the same time, these and other databases also allow the automatic surveillance of drug resistance at a global level, influencing public health interventions.

WGS has the potential to revolutionize the definition of drug susceptibility testing (DST) of MTB in both high- and low-income settings, and a growing knowledge of the genetic mechanisms of resistance, combined with an improved IT infrastructure, will facilitate its adoption and enhance its clinical utility for drug testing. A key challenge is to demonstrate that its use in the routine diagnostic service will have an impact on patient outcomes. Treatment guidelines have evolved over decades to reflect the role of clinical suspicion and not to over-rely on microbiological positive results, especially in the setting of culture-negative samples or in case of extra-pulmonary disease. Recently, rapid PCR methods have allowed the rapid detection of resistance within the same day. However, data from developing countries do not seem to support any improvement in outcome [15,16] and it is currently unknown if this is also relevant to WGS.

Table 1
Current applications of whole genome sequencing and *Mycobacterium tuberculosis*, with achievements and limitations

Achievements	Limitations
1. Prediction of drug susceptibility and resistance	
- Diagnostic workflow with data generated in 9 days and at a price 7% cheaper	- Significant variation for the remaining first-line and other drugs
- First line drugs (rifampicin and isoniazid): strong performance with high sensitivity and specificity	- Culture still needed for DNA extraction and WGS
- Potential for WGS directly from clinical samples	- Bioinformatics support and IT infrastructure needed to download and analyse data
- Online tools available for rapid data interpretation	- Lack of accreditation (ISO 15189 and others)
2. Epidemiological analysis	
- Higher resolution compared with MIRU-VNTR typing, IS6110 RFLP typing, and spoligotyping methods	- Still insufficient to resolve transmission networks in tuberculosis outbreaks
- Ability to distinguish relapse from reinfection	- Clinical benefits and cost-effectiveness not demonstrated
- Better understanding of evolution, lineages, and genomic variation	- Bioinformatics support and IT infrastructure needed to download and analyse data
3. Research	
- Demonstration of specific deletions and SNPs peculiar to clinical strains	- Further studies and techniques still needed to confirm gene function
- Discovery of mechanism of action of new drugs	- Mutations may be non-specific
- Demonstration of more extensive genetic variability than original expected	- Need for large database to compare data at international level
- Potential in vaccine development	

Please cite this article in press as: Satta G, et al., *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential?, Clinical Microbiology and Infection (2017), https://doi.org/10.1016/j.cmi.2017.10.030