Review

CellPress

Whole-genome sequencing to control antimicrobial resistance

Claudio U. Köser[1], Matthew J. Ellington[2], and Sharon J. Peacock[1,2,3,4]

[1] Department of Medicine, University of Cambridge, Cambridge, UK
[2] Clinical Microbiology and Public Health Laboratory, Public Health England, Cambridge, UK
[3] Cambridge University Hospitals National Health Service Foundation Trust, Cambridge, UK
[4] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK

Following recent improvements in sequencing technologies, whole-genome sequencing (WGS) is positioned to become an essential tool in the control of antibiotic resistance, a major threat in modern healthcare. WGS has already found numerous applications in this area, ranging from the development of novel antibiotics and diagnostic tests through to antibiotic stewardship of currently available drugs via surveillance and the elucidation of the factors that allow the emergence and persistence of resistance. Numerous proof-of-principle studies have also highlighted the value of WGS as a tool for day-to-day infection control and, for some pathogens, as a primary diagnostic tool to detect antibiotic resistance. However, appropriate data analysis platforms will need to be developed before routine WGS can be introduced on a large scale.

A modern tool for a growing challenge

The spread of antibiotic resistance constitutes one of the most serious threats to human health [1]. If left unchecked, even 'minor surgery and routine operations could become high risk procedures' [2]. As a result, the UK Chief Medical Officer has called for antibiotic resistance to be included in the National Risk Register [2]. The causes of the spread of antibiotic resistance are complex, as are the strategies to combat this threat [3–5]. These include developing novel agents as well as maximising the utility of currently available antibiotics by preventing infections and improving diagnostics. As a result of recent technological advances, microbial WGS has emerged as a tool that could underpin the success of such goals.

A detailed discussion of the various sequencing technologies and their advantages and disadvantages for pathogen WGS can be found elsewhere [6–8]. For the purposes of this review it is merely necessary to appreciate that two classes of sequencers exist. The first can be compared to oil tankers in that they are relatively slow but have a high throughput and low cost (currently approximately $65 per bacterial genome) [6,9]. By contrast, benchtop sequencers are akin to speedboats. They have a lower throughput and higher cost (approximately $150 per genome in the case of the Illumina MiSeq) but can sequence multiple genomes in less than 24 h [6,10]. In fact, recent advances in sample preparation have enabled WGS directly from single bacterial colonies grown under routine diagnostic conditions without the need for subculture to isolate sufficient DNA for sequencing [11]. Together, these sequencers have allowed researchers to interrogate the evolution of

Glossary

Deep sequencing: WGS of a pure sample (e.g., DNA extracted from liquid culture that was inoculated from a single bacterial colony) is usually sequenced to at least a 10- to 15-fold depth (i.e., each position in a genome is sequenced 10–15 times). However, when sequencing samples that might be heteroresistant (see below), or might contain multiple strains of the same species, greater sequencing depth is required to identify minority variants or to separate the strains [49]. For example, deep sequencing, albeit of sections of the genome rather than its entirety, is used routinely in clinical applications for genotypic HIV DST (see below) to detect resistances that occur below 15–20% of the total population, which is the limit of detection for Sanger sequencing [9].

Drug-susceptibility testing (DST): the process of classifying an isolate or a mix of isolates from the same species (e.g., in the case of HIV) as susceptible (high likelihood of clinical response) or resistant (low likelihood of clinical response) to an antibiotic or antiviral. In some cases, an intermediate category (clinical success uncertain) is employed [97]. For most pathogens this is achieved using phenotypic methods, whereas for slow-growing organisms, such as HIV, genotypic DST is the standard of care [9,98]. WGS will not replace phenotypic DST for fast-growing organisms given that it would be too expensive, but will enable rapid genotypic DST for MTBC, for which phenotypic alternatives are slow [9,49].

DST errors: when comparing two DST methods, a very major error occurs if the assay under evaluation gives a susceptible result compared with a resistant result by the reference method. The misclassification of a susceptible reference result as resistant is referred to as a major error. The classification of a susceptible or resistant reference result as intermediate, or vice versa, is known as a minor error [98].

Heteroresistance: a mixture of susceptible and resistant populations, which may arise as a result of a mixed infection with strains with different susceptibility patterns, or the evolution of a resistant subpopulation from a susceptible parental strain during the infection [99]. The ability of different DST methods to detect resistant isolates at a low percentage of the total population varies (see deep sequencing) [54].

Metagenomics: unbiased sequencing from a primary sample (i.e., without isolating or culturing a specific organism). Depending on the type of sample, it may be highly mixed with several bacterial and viral species as well as host nucleic acid. Provided that the mixed DNA is sequenced to sufficient depth, the genome of an organism can be recovered directly from a clinical sample [70].

Whole-genome sequencing (WGS): modern WGS relies on one or a combination of second- or third-generation sequencing technologies [6]. Although the entire genome is sequenced, it cannot be assembled completely if a genome contains repeated stretches of DNA that are longer than the length of DNA that can be sequenced in a single read by the sequencing technology employed [100]. Additional parts of the genome may be excluded for some analyses. For example, only the core genome that is present in all isolates in question is used for phylogenetic analyses, which represents about 80% of the entire genome for *Clostridium difficile* [62].

Corresponding author: Köser, C.U. (cuk21@cam.ac.uk).
Keywords: whole-genome sequencing; antibiotic resistance; surveillance; diagnostics.

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2014.07.003

antibiotic resistance at an unprecedented scale. For example, a recent study sequenced the genomes of more than 3000 pneumococcal isolates and another sequenced more than 3600 group A streptococci [12,13]. In addition, benchtop sequencers have become sufficiently fast to become primary diagnostic tools rather than tools for research or retrospective surveillance [9,14]. We review here the insights that WGS has provided with regards to antimicrobial resistance. Owing to the rapid expansion of this field, this endeavour must inevitably be a selective one.

WGS to develop novel antibiotics

WGS has become an essential tool for drug development by enabling the rapid identification of resistance mechanisms, particularly in the context of tuberculosis (TB), which remains a global public health emergency [15,16]. In 2005 the first published use of 454 pyrosequencing (the first second-generation WGS technology) was to identify the F0 subunit of the ATP synthase as the target of bedaquiline, which subsequently became the first representative of a novel class of anti-TB agents to be approved in 40 years [16,17]. This has enabled researchers to sequence this gene in phylogenetically diverse reference collections to ensure that it is conserved across *Mycobacterium canettii* as well as the various lineages and species that comprise the *Mycobacterium tuberculosis* complex (MTBC), the causative agents of TB [18]. This represents an important step because drug candidates are usually only tested against a small number of isolates during the early phases of drug development. Similarly, only a limited number of MTBC genotypes are sampled in clinical trials, depending on where these are conducted [19]. As a result, intrinsically resistant strains might be missed, as has been the case for PA-824, an anti-TB agent in Phase III trials [19–21].

The early elucidation of resistance mechanisms using WGS also has implications for the design of clinical trials. If resistance mechanisms are discovered that only result in marginally increased minimal inhibitory concentrations (MICs) compared with the wild type MIC distributions, more frequent dosing or higher doses could be employed in clinical trials to overcome this level of resistance. Moreover, the discovery of cross-resistance between agents using WGS can influence the choice of antibiotics that are included in novel regimens. TB is always treated with multiple antibiotics to minimise the chance of treatment failure as a result of the emergence of resistance during treatment [22]. Regimens that contain agents to which a single mutation confers cross-resistance should be avoided if these mutations arise frequently *in vivo*. WGS has recently highlighted that this may be the case with three Phase II trial regimens that contain bedaquiline and clofazimine because the mutational upregulation of an efflux pump confers cross-resistance to both drugs [23].

In addition to being a tool to design clinical trials, WGS has become an increasingly important tool during clinical trials. Specifically, it is increasingly being used to distinguish exogenous reinfection from relapse of the primary infection, which is crucial in assessing the efficacy of the drug or regimens under investigation [24,25]. Traditional epidemiological tools do not always provide the necessary resolution for this purpose. This is due to the fact that they only interrogate minute parts of the genome {e.g., multilocus sequence typing (MLST) of *Pseudomonas aeruginosa* analyses only 0.18% of the genome [26]}. By contrast, WGS interrogates the complete (or near-complete) genetic repertoire of an organism. Therefore, the resolution of WGS is only limited by the rate of evolution of the pathogen and will become the gold standard for clinical trials of new anti-TB agents and other infectious diseases associated with recurrent disease [27,28].

WGS for surveillance

Surveillance is the cornerstone in the control of infectious diseases [3,29]. In this context, the ultimate molecular resolution delivered by WGS has provided unprecedented insight into the history of the emergence and spread of antibiotic resistance [30]. In the case of *Mycobacterium abscessus* this has resulted in a paradigm shift in the understanding of its transmission. WGS provided strong evidence that this non-tuberculous mycobacterium, which is difficult to treat owing to its intrinsic resistance to many antibiotics, is transmissible between cystic fibrosis (CF) patients in the hospital setting [31]. Transmission occurred despite strict infection control policies that were designed to prevent the acquisition of other pathogens, such as epidemic *P. aeruginosa* (to be discussed below), that are known to be transmissible between CF patients. Importantly, this included the transmission of macrolide-resistant isolates whereas environmental *M. abscessus* isolates are susceptible to the drug [31]. As a result, the national infection control guidelines have been revised to minimise the possibility of further transmission [32]. Moreover, a much larger UK-wide and international study is currently underway to investigate the extent of transmission elsewhere [33].

WGS has highlighted the importance of the international spread of antibiotic-resistant organisms [34]. For example, the epidemic *P. aeruginosa* Liverpool strain has been shown to transmit between CF clinics in the UK and North America [26]. The development of fluoroquinolone resistance in the EMRSA-15 lineage of methicillin-resistant *Staphylococcus aureus* (MRSA) in the mid-1980s preceded the pandemic spread of this variant [35]. Similarly, the acquisition of fluoroquinolone resistance has been identified as a key genetic change linked to the emergence of the two lineages of epidemic *Clostridium difficile* (027/BI/NAP1) during the early 2000s [36].

Large WGS studies have also started to shed light on the interface between humans and animals. Throughout the 1990s, multidrug-resistant *Salmonella* Typhimurium DT104 was responsible for a global epidemic [37]. Contrary to prior tenets, a comparison of longitudinal human and livestock isolates from Scotland demonstrated that human infections were not primarily due to transmission from local animals. Instead, the human and animal populations of Typhimurium DT104 in Scotland were largely distinct, which suggests that other sources of infection, such as imported food, were responsible for the human disease [37]. By contrast, WGS demonstrated that livestock-associated CC398 MRSA originated from human methicillin-susceptible *S. aureus* and acquired methicillin resistance