The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee

Ellington, M J; Ekelund, O; Aarestrup, Frank Møller; Canton, R; Doumith, M; Giske, C; Grundman, H; Hasman, H; Holden, M T G; Hopkins, K L

Total number of authors: 21

Published in: Clinical Microbiology and Infection

Link to article, DOI: 10.1016/j.cmi.2016.11.012

Publication date: 2017

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee


2) Department of Clinical Microbiology and the EUCAST Development Laboratory, Kronoberg Region, Central Hospital, Växjö, Sweden
3) National Food Institute, Research Group for Genomic Epidemiology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
4) Servicio de Microbiología, Hospital Ramón y Cajal and Instituto Ramón y Cajal de Investigación Sanitaria, Madrid, Spain
5) Department of Laboratory Medicine, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden
6) Institute for Medical Microbiology and Infection Control, Copenhagen, Denmark
7) Institute of Medical Microbiology, and Research Centre, Warsaw, Poland
8) National Reference Centre for Antibiotic Resistance, Bacteriology-Hygiene unit, University of Freiburg, Freiburg, Germany
9) States Serum Institute, Medical & Biological Sciences, North Haugh, University of St Andrews, UK
10) University Medical Centre Freiburg, Infection Prevention and Hospital Hygiene, Faculty of Medicine, University of Freiburg, Freiburg, Germany
11) University Medical Centre, Department of Microbiology, Tampere University Hospital, Tampere, Finland
12) Department of Clinical Microbiology and the EUCAST Development Laboratory, Kronoberg Region, Central Hospital, Växjö, Sweden
13) Department of Medical Microbiology, North Bristol NHS Trust, Southmead Hospital, Bristol, UK
14) University Medical Centre Freiburg, Infection Prevention and Hospital Hygiene, Faculty of Medicine, University of Freiburg, Freiburg, Germany
15) Department of Infectious Diseases and Immunology, Utrecht University, Utrecht, The Netherlands
16) School of Medicine, Medical & Biological Sciences, North Haugh, University of St Andrews, UK
17) Norwegian National Advisory Unit on Detection of Antimicrobial Resistance, University Hospital of North Norway, Department of Microbiology and Infection Control, Tromsø, Norway

E-mail address: neil.woodford@phe.gov.uk (N. Woodford).

Matthew Ellington and Oskar Ekelund are joint first authors.

http://dx.doi.org/10.1016/j.cmi.2016.11.012

© 2016 Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases. All rights reserved.
Introduction

The remit of the European Committee on Antimicrobial Susceptibility Testing (EUCAST) Subcommittee on the Role of Whole Genome Sequencing (WGS) in Antimicrobial Susceptibility Testing of Bacteria was:

(a) to perform a review of the literature describing the role of WGS in antimicrobial susceptibility testing (AST) of bacteria;
(b) to assess the sensitivity and specificity of WGS compared with standard phenotypic AST;
(c) to consider how WGS for AST may be applied in clinical microbiology laboratories and the likely implications for phenotypic and other genotypic methods in use;
(d) to consider the epidemiological implications of using WGS;
(e) to consider the clinical implications of WGS for the selection of antimicrobial therapy;
(f) to consider the principles of how the results of WGS for AST could be presented to clinical users;
(g) to describe the drivers and barriers to routine use of WGS;
(h) to report at ECCMID 2016.

We chose to tackle this task on a ‘by organism’ basis (for Enterobacteriaceae (non-Salmonellae); Salmonella spp.; Staphylococcus aureus; Streptococcus pneumoniae; Neisseria gonorrhoeae; Mycobacterium tuberculosis; Clostridium difficile; Acinetobacter baumannii and Pseudomonas aeruginosa), with particular focus on the use of technology for characterizing cultured isolates of bacteria that have been identified as critical antimicrobial resistance (AMR) threats by the World Health Organization [1]. There are encouraging signs, but our report makes clear that more robust data are needed across these diverse ‘bug/drug’ combinations. Furthermore, work is needed to overcome problems currently posed by particular species and/or certain antimicrobial classes. We highlight these gaps, make recommendations (summarized below), and encourage others to use these to generate the analyses that will move this important topic forwards.

The committee met by teleconference between July 2015 and March 2016 to agree the organisms and detailed scope of the report and appoint specialists to undertake non-systematic reviews of the literature and write the corresponding sections of the report. The findings of the report were agreed via teleconference and/or email, and were presented at the 26th European Congress on Clinical Microbiology and Infectious Diseases (ECCMID) in April 2016. Shortly after that, a draft for public consultation was posted online (at www.eucast.org/documents/consultations). The Subcommittee considered and responded to comments and several were included in or affected the content of this print version.

Summary of the conclusions and recommendations

(a) For most bacteria considered in this report, the available evidence for using WGS as a tool to infer antimicrobial susceptibility (i.e. to rule-in as well as to rule-out resistance) accurately is either poor or non-existent. More focused studies and additional funding resources are needed as a priority to improve knowledge.
(b) The primary comparator for WGS-based prediction of antimicrobial susceptibility should, whenever possible, be the epidemiological cut-off value (ECOFF).
(c) Assessing genotypic data against clinical breakpoints represents a tougher challenge, but will be necessary if WGS-based testing is to guide clinical decision making. Clinical breakpoints should therefore be used as a secondary comparator, ideally using the same data sets as used for ECOFF-based assessments.
(d) Most published evidence does not currently support use of WGS-inferred susceptibility to guide clinical decision making (i.e. to replace routine phenotypic AST in most or all cases).
(e) There should be international agreement on the most appropriate and effective principles and quality control (QC) metrics to facilitate early standardization and harmonization of analytical approaches and interpretative criteria for WGS-based predictive AST. Only data sets passing agreed QC metrics should be used in antimicrobial susceptibility predictions, as resistance genes or mutations might be missed in sequences of poor quality.
(f) Different bioinformatics tools for predicting AST should perform to minimum standards and should be calibrated and shown to be equivalent in terms of the results generated.
(g) A single database of all known resistance genes/mutations should be established to ensure that there is parity of analysis and to facilitate measurement of comparative accuracies across different systems and bioinformatics tools. This database should be updated regularly, and must have strictly curated minimum standards for the inclusion of new resistance genes and mutations. An important function of a centralized database would be to control resistance gene nomenclature.
(h) Expansion of the evidence base is a critical priority if WGS is to be considered seriously as a rival to phenotypic AST.
(i) For most bacterial species and in most countries the current cost and speed of inferring antimicrobial susceptibility from WGS data remain prohibitive to wide adoption in routine clinical laboratories.
(j) This report should be considered as a baseline discussion document, which should be revisited and updated at regular intervals (likely every 18–24 months) as sequencing...
technologies become more affordable and more widely applied, and the available evidence increases.

Defining resistance

MIC distributions of wild-type bacteria, ECOFFs and their relationship to clinical breakpoints

In 2002, EUCAST introduced the concept of gathering large numbers of MIC values from many contributors to present on a web site as aggregated reference MIC distributions for every important combination of microorganisms and antimicrobial agents [2]. The original conditions of acceptance of individual MIC distributions were:

(a) that each contribution of MIC values must consist of a minimum number of isolates;
(b) that the species was well defined;
(c) that MIC determinations were performed using standardized methodology (or a method calibrated to a standardized method);
(d) that the concentrations tested were not truncated at the lower end of the concentration series.

A EUCAST Subcommittee is currently reviewing the rules for inclusion and exclusion of data sets in aggregated MIC distributions and the process by which ECOFF values are defined.

Contributors of MIC data are not informed about whether or not their contributions are accepted. There are currently more than 29,000 MIC distributions in the EUCAST database, which amounts to many millions of MIC values. The distributions are from breakpoint committees, individual researchers in human and veterinary medicine, programmes for the surveillance of AMR in humans and animals, EUCAST development projects, pharmaceutical companies as part of programmes for the development of new agents, and more. The distributions are freely available on www.eucast.org.

A typical aggregated MIC distribution, in this case for *Escherichia coli* and cefotaxime, is shown in Fig. 1. Of 72 cefotaxime MIC distributions for *E. coli* available in the database, 41 fulfilled the criteria for acceptance, were aggregated and the aggregated distribution displayed on the EUCAST web site as in Fig. 1.

MIC distributions are uni-modal or multi-modal. The left-hand, most often dominating and Gaussian-shaped part of the distribution, represents the isolates devoid of phenotypically detectable acquired resistance mechanisms, otherwise known as ‘the wild-type (WT) MIC distribution’. Furthermore, MIC values for bacteria from humans and animals are distributed in the same way. The WT is not affected by the geographical location where the isolates are collected, the specimen source (e.g. humans or animals [3], or healthy or sick individuals) or the era of collection (some of the distributions date from the 1950s whereas others are very recent).

There are several ways (biological or statistical) to assess the WT distribution to define the MIC value best representing the upper end of the WT distribution [4—6]. Despite the fact that there is often no absolute and distinguishing threshold to mark the upper end of the WT or beginning of the non-wild-type (NWT) distribution, it has been useful to define the ECOFF as the ‘highest MIC for organisms devoid of phenotypically detectable acquired resistance mechanisms’. It provides a means to distinguish between resistant and susceptible populations in a biological sense.

From a clinical point of view, categorizing isolates into WT and NWT informs the clinician of whether or not the isolate causing infection is devoid of acquired resistance mechanisms or not, irrespective of its clinical susceptibility categorization as susceptible (S), intermediate (I) or resistant (R).

There is no immediate relationship between the categorization of WT and NWT on one side and the clinical categorization ‘S’, ‘I’ and ‘R’ on the other. A WT microorganism can be categorized as ‘S’, ‘I’ or ‘R’ to a particular antimicrobial agent and a NWT organism may still be categorized as ‘S’. This means that if one wants to encompass both WT and NWT on one hand and clinical ‘S’, ‘I’ and ‘R’ on the other, the possible susceptibility categories for an isolate to
any antimicrobial agent are $S^{WT}$, $S^{NWT}$, $I^{WT}$, $I^{NWT}$, $R^{WT}$ and $R^{NWT}$. There are many examples of each of these categories in the EUCAST breakpoint tables. An *E. coli* isolate with a ciprofloxacin MIC of 0.25 mg/L exemplifies the $S^{NWT}$ category. Conversely, *Pseudomonas aeruginosa* and ticarcylene or *Stenotrophomonas maltophilia* and carbapenems both exemplify the $R^{WT}$ category. For ampicillin, TEM-1-producing *E. coli* represents $R^{NWT}$.

At the outset of the review of all breakpoints for all antimicrobial classes in 2002, EUCAST decided that clinical breakpoints should not divide WT MIC distributions. If breakpoints are allowed to bisect WT MIC (or zone diameter) distributions, the methodological variation would obviate reproducible ‘S’, ‘I’, and ‘R’ categorization [2], despite increased standardization of MIC determination [7]. We have not had reason to change this view. Following on from this, once it has been established that the species is a good clinical target for the agent in question, the ECOFF is the lowest possible susceptible breakpoint. The ECOFF is also the relevant ‘cut-off’ to screen for low-level resistance using phenotypic susceptibility testing.

ECOFFs provide an opportunity to compare AMR and resistance development when clinical breakpoints: (a) differ among committees and agencies, e.g. EUCAST, CLSI, US Food and Drug Administration and European Medicines Agency; (b) change over time; or (c) differ between humans and animals. There is no difference in principle between MIC distributions and ECOFFs for fast-growing non-fastidious and fastidious bacteria, and those for slow-growing bacteria, such as *Mycobacterium tuberculosis* [8], or fungi, such as *Candida* spp. and *Aspergillus* spp. [9].

More information on EUCAST in general and wild-type MIC distributions and ECOFFs in particular can be obtained from the EUCAST website (www.eucast.org) and the recently published review of EUCAST activities since 2001 [10].

This Subcommittee sought to assess whether available data are sufficient to test the hypothesis that the closest relationship between sequencing and phenotypic testing will be achieved by using the WT versus NWT categories.

*Molecular mechanisms of antimicrobial resistance*

For most of the clinically relevant bacterial pathogens, phenotypic analysis of bacterial susceptibility to antimicrobial agents is relatively straightforward and relies on well-proven methods, such as agar and broth microdilution (the latter being the reference standard) or disc diffusion, followed by interpretation according to agreed guidelines.

With the introduction of Sanger sequencing in the mid-1970s and PCR in the 1980s, it became possible to study some of the molecular mechanisms responsible for the observed non-susceptibility to various antimicrobial agents. Common examples of these molecular mechanisms are: (a) transferable AMR genes (e.g. extended-spectrum β-lactamases (ESBLs)); (b) upregulation of AMR gene expression by point mutations (e.g. *ampC* in *E. coli* or regulatory mutations effecting efflux in many taxa); (c) porin modification or loss (e.g. by deletion events and/or lack of expression); (d) point mutations in essential single-copy (e.g. *gyrA* and/or *parC* of *Enterobacteriaceae*) and multi-copy (e.g. mutations in one or more loci of the 23S rRNA gene) housekeeping genes. To complicate the matter, some bacterial species (or higher taxonomic orders) can be intrinsically resistant to a given antimicrobial agent [11]. This intrinsic resistance can be caused by some of the mechanisms listed above but can also be a result of the lack or unavailability of targets for the antimicrobial agent.

Traditional Sanger sequencing and rapid molecular methods, e.g. PCR, allow screening for a limited number of resistance genes, which are often selected because they confer resistance to key antibiotics (e.g. genes encoding ESBLs or carbapenemases). The data offer very limited opportunities to compare genotype with phenotype. By contrast, WGS has potential to yield data about any resistance gene or mutation present and the data might therefore be analysed to create a genotypically inferred antimicrobial resistance profile (or antibiogram) or, perhaps, to infer susceptibility.

**Using next-generation sequencing data for *in silico* (genotypic) detection of AMR**

Next-generation sequencing data producing WGS information can originate from a variety of different sequencing platforms. A review of these technologies is beyond the scope of this report, but can be found elsewhere [12]. Short read (e.g. 100 to 500 bp) output with high accuracy may be complemented by that produced by much longer reads. At the time of writing, these newer platforms come at significantly greater cost and higher error rates than the short-read technology. The dominant short-read technology produces single (raw) reads that are in most cases shorter than the gene(s) responsible for the reduced susceptibility to a given antimicrobial agent and either need to undergo de novo assembly to obtain larger fragments of the originally contiguous DNA (‘contigs’) or by reference ‘mapping’ to known genetic targets (in this case, AMR genetic determinants). Repeat regions in DNA fragments are particularly challenging and correct assembly may be problematic.

Using WGS data for detection of the many different molecular mechanisms leading to AMR yields far more information from a single physical test than other methods (e.g. PCR or microarrays) and, at its most fundamental level, does not require previous knowledge of the resistance phenotype of the isolate. Nevertheless, there is a need to understand the potential ‘added value’ of WGS with regard to the clinical implications of AMR and so the validity of data generated by these novel technologies must be challenged against phenotypic methods to differentiate WT isolates from NWT, or S isolates from R isolates. In this regard WGS is a genetic test that defines a genotype as WT or NWT and compares most directly with phenotypic criteria that do the same (ECOFFs).

Although more informative than conventional molecular techniques, WGS is no simple task, especially when the data have been generated by short-read (‘second generation’) technology. Detection of defined resistance genes can be achieved either by BLAST (Basic Local Alignment Search Tool) analysis of draft genomes towards a gene-based database or by mapping individual reads to the same type of database. Such solutions are already available as either downloadable tools, such as ARG-ANNOT [13], or web-based tools, such as RESFINDER for BLAST analysis and MEGAMAPPER for mapping of reads [14,15]. The gene-based solutions have the obvious requirement for full-length genes identical to already characterized (and preferably published) AMR genes. The bioinformatics solutions mentioned above are able to identify less-than-perfect hits (<100% nucleotide identity, truncated genes because of non-perfect de novo assembly), but such hits will always need to be subjected to some sort of assessment if they are to be translated into a predicted phenotype.

Accurate prediction of resistance by WGS can be complicated by insufficient knowledge about all genetic variation leading to reduced susceptibility for a given antimicrobial agent (such as colistin or daptomycin) as well as the emergence of new mechanisms and when resistance arises due to altered expression of intrinsic genes (e.g. those encoding efflux pumps). Also, shortcomings of second-generation sequencing technology may hamper accuracy. An example of the latter could be *Enterococcus faecium*, where a point mutation (G2576T) in two or three copies of the six 23S rRNA loci would lead to phenotypic resistance to linezolid [16].

De novo assembly of second-generation sequence data from the
same isolate would most likely lead to assembly into a WT version of the 23S gene due to the consensual nature of the assemblers, where only the most abundant base is reported in the draft genome data.

Quality metrics for WGS

Like any other test, the quality of WGS data can vary between individual test runs. Therefore before any actual bioinformatics analysis, QC steps are essential to assess whether the WGS data have reached a suitable standard. Only data sets passing these QC metrics should be used in antimicrobial susceptibility predictions, as resistance genes or mutations might be missed in sequences of poor quality. These QC steps include (a) assessing the quality and quantity of the raw reads to ensure sufficient coverage (e.g. >30 times coverage) of the bacterial genome, (b) assessing the quality of the de novo assembly (leading to a draft genome sequence) and (c) detecting possible contaminant DNA, originating either from upstream handling of the bacterial isolates and DNA purification or from the preparation and running of the DNA samples on the sequencer. Some of the different sequence QC parameters that have been used are listed in Table 1. The parameters most frequently used are highlighted in bold. There are currently no international standards for QC thresholds to use for assessing quality. This seems to be individually decided by researchers and may also depend on the purpose of the study or the methods used for sequencing. A high error rate for a sequencing method can be compensated to some extent by greater depth of coverage. The necessary QC threshold also depends on the species analysed. Hence, before WGS can be routinely implemented into accredited clinical practice there is a need to establish necessary minimum QC thresholds (e.g. by multiple sequencing of reference isolates) for identification of resistance genes and their variants.

The Global Microbial Identifier initiative [15] is currently collaborating with the US Food and Drug Administration and the COMPARE project [16] in proficiency testing of WGS data, and isolates have been distributed to 50 laboratories worldwide. This and similar initiatives are important first steps towards setting objective QC thresholds. There is, however, a need to expand this using more isolates as well as developing standard data sets of raw sequences to facilitate the assessment of performances across different laboratories.

The need for a standardized, open-access database

Most of the genomes released now are not closed, so there is a need for better standardization of annotation to facilitate detection of AMR genes because standard BLAST analysis will retrieve plenty of hits within annotated or raw sequences available in GenBank, and those hits will be inconsistently annotated even where the actual sequences are identical.

Currently several AMR databases exist, and they are either downloadable for use locally (e.g. ARG-ANNOT) [13] or are web-based solutions (ResFinder for BLAST analysis and MGmapper for mapping of reads [14,15] and a Comprehensive Antibiotic Research Database, CARD [17]). In addition to a fully curated database of accurately annotated genes that seeks to avoid the pitfalls posed by erroneously annotated genes, there also exists a need for a single, standardized ‘challenge database’ solution that contains all validated AMR genes as well as those point mutations in chromosomal target genes that are known to be associated with antimicrobial

Table 1

<table>
<thead>
<tr>
<th>QC parameter</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reads</td>
<td>Number of reads refers to sequence yield (the amount sequenced)</td>
</tr>
<tr>
<td>Average read length</td>
<td>The average length of all reads, measured in base pairs.</td>
</tr>
<tr>
<td>Number of reads mapped to reference sequence</td>
<td>The number of reads that map to a closed (finished) genome (same strain).</td>
</tr>
<tr>
<td>Proportion of reads mapped to reference sequence (%)</td>
<td>The proportion of reads that map to a closed genome (same strain).</td>
</tr>
<tr>
<td>Number of reads mapped to reference chromosome</td>
<td>The number of reads that map to a closed chromosome (same strain).</td>
</tr>
<tr>
<td>Proportion of reads mapped to reference chromosome (%)</td>
<td>The proportion of reads that map to a closed genome’s chromosome (same strain).</td>
</tr>
<tr>
<td>Reads mapped to reference plasmids</td>
<td>The number of reads that map to plasmids, if present.</td>
</tr>
<tr>
<td>Proportion of reads mapped to reference plasmids (%)</td>
<td>The proportion of reads that map to plasmids (if present) of the closed genomes.</td>
</tr>
<tr>
<td>Depth of coverage, total DNA sequence</td>
<td>Describes the number of times the sequenced base pairs cover the reference DNA.</td>
</tr>
<tr>
<td>Depth of coverage: chromosome</td>
<td>Number of base pairs sequenced divided by the total size (both chromosome and plasmids) of the closed genome (same strain), often expressed with an “x” (e.g. 30x). A minimum depth of 30x is usually preferred.</td>
</tr>
<tr>
<td>Depth of coverage: plasmid</td>
<td>As for total DNA coverage, but describes the number of base pairs sequenced divided by the total size of the closed plasmid (same strain).</td>
</tr>
<tr>
<td>Size of assembled genome</td>
<td>Often used to identify contamination.</td>
</tr>
<tr>
<td>Size of assembled genome per total size of DNA sequence (%)</td>
<td>The proportion of contigs that map directly to the closed genome (same strain).</td>
</tr>
<tr>
<td>Total number of contigs</td>
<td>Generally, the total number of contigs assembled, &lt;1000 contigs indicates good quality. For organisms with genomes 5–6 Mb in size then &lt;100 contigs is (generally) realistic. The total number of contigs assembled that have a sequence length &lt; 500 bp. This should correspond well to the total number of contigs.</td>
</tr>
<tr>
<td>Number of contigs &gt;500 bp</td>
<td></td>
</tr>
<tr>
<td>Longest contig length</td>
<td>The length of the longest contig.</td>
</tr>
<tr>
<td>Shortest contig length</td>
<td>The length of the shortest contig.</td>
</tr>
<tr>
<td>Mean, median and standard deviation N50</td>
<td>Mean, median and standard deviation of the contigs, used to evaluate quality.</td>
</tr>
<tr>
<td>N50</td>
<td>The length for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs, and for which the collection of all contigs of that length or shorter also contains at least half of the sum of the lengths of all contigs. N50 &gt; 15 000 normally indicates good quality, but minimum size of 30 000 bp is often preferred.</td>
</tr>
<tr>
<td>NG50</td>
<td>Helpful for comparisons between assemblies. As N50, except that 50% of the genome size must be of the NG50 length or longer. Where the assembly size ≥ the genome size then NG50 cannot exceed N50.</td>
</tr>
</tbody>
</table>
resistance. This can then be used as a standard reference data set for different bioinformatics analysis tools. Any such solution should be flexible so that stringency of detection can be changed to allow detection from partial gene sequences (length <100%) and/or AMR genes with identities <100%. Looking at conserved sequence motifs in gene families (e.g. β-lactamases) should also help in assessing the validity of newly detected genes and in particular those exhibiting low sequence identity matches.

This single web solution should be iterative and enhanced by regular, validated updates of newly identified gene sequences and point mutations at a frequency that remains to be decided. Machine learning should be explored to improve iteratively and automatically the detection algorithms for this purpose.

Evidence reports for in silico prediction of antimicrobial resistance

Enterobacteriaceae (other than Salmonella spp.)

Background

Multidrug-resistant Enterobacteriaceae are emerging as a serious infectious disease challenge. They can accumulate many antimicrobial resistance genes through horizontal transfer of genetic elements, those coding for β-lactamases (e.g. ESBLs and carbapenemases), fluoroquinolones and aminoglycosides being of particular concern.

Published studies

A small number of studies have assessed the feasibility of using WGS to infer AMR in E. coli and Klebsiella pneumoniae genomes; they are largely based on screening for known acquired AMR genes and a small number of known resistance-conferring mutations, such as those associated with ciprofloxacin resistance. In one study, Stoeasser et al. reported 95% concordance between phenotypic and WGS-predicted susceptibility for seven commonly used agents (amoxicillin, amoxicillin/clavulanic acid, ciprofloxacin, gentamicin, ceftiraxone, ceftazidime and meropenem) by querying 143 assembled genomes from E. coli and K. pneumoniae with a compiled database of acquired AMR sequences and mutations in the quinolone-resistance-determining regions of gyrA and parC [19]. An even higher level of concordance (99.74%) between phenotypic susceptibility testing and the WGS-predicted resistance to five agents or classes of agents (β-lactams, chloramphenicol, sulphonamides, tetracycline and trimethoprim) in E. coli genomes was reported in an earlier study using the same approach with a database of 1411 different AMR sequences, confirmed using simple blotting and PCR approaches to the expected genes [14,20]. A recent study investigating 76 E. coli isolates from farm cattle also showed good phenotype–genotype correlation (97.8%), with the majority of discordant results attributed to the prediction of streptomycin resistance [21].

There are important limitations to identifying the mode of transmission of these acquired genes in short-read sequences due to exclusion of repeat regions during the ‘cleaning’ stages while initial contigs are assembled. This typically results in inaccuracies in annotation. These are particularly marked in highly recombinant plasmids, which unfortunately carry most of the AMR genes that are relevant to β-lactam and aminoglycoside resistance in the Enterobacteriaceae. Nevertheless, using sequencing technologies with longer reads (and greater cost), such as PacBio and MinION, improved their detection [22–25]. These bioinformatics challenges also include the development of tools that can detect signature sequences of AMR determinants (e.g. β-lactamase motifs) to identify potentially new variants conferring acquired AMR, which can be explored in more detail.

Problems found or anticipated—gaps in the knowledge base

Chromosomal mutations that alter the cell membrane permeability due to modification in the structure or the levels of expression of outer membrane proteins, antimicrobial efflux due to efflux pumps such as resistance-nodulation-division and major facilitator superfamily pumps, or changes in the lipopolysaccharide structure have been linked to decreased susceptibility and resistance to β-lactams, quinolones, chloramphenicol, tetracyclines, tigecycline and colistin in Enterobacteriaceae, but have yet to be fully elucidated. This makes comprehensive phenotypic–genotypic comparisons difficult [26–32] by limiting the sensitivity of the WGS-based data. In particular, the relationships between chromosomal mutations and the related phenotypic changes responsible for resistance are not always well characterized and screening genome sequences for insertion sequences interrupting or modifying the expression of resistance-associated genes, including intrinsic β-lactamases, could be problematic due to constraints inherent in using short reads. Therefore, and in contrast to horizontally acquired resistance genes, the ability of WGS to predict resistance due to, or modulated by, chromosomal alterations is likely to be restricted by existing knowledge, as in the case of carbapenem non-susceptibility resulting from the combination of...
decreased permeability and AmpC or ESBL enzymes and also for antimicrobial agents for which the underlying genetic backgrounds of resistance are yet to be fully characterized (e.g. amoxicillin-clavulanic acid, nitrofurantoin, temocillin, colistin and tigecycline). Screening for the loss-of-function mutations via nonsense mutations, frameshifts or insertion elements that are meaningfully less complex than substitutions affecting the structure, dynamics and substrate specificity of resistance-conferring proteins is realistically achievable [33–35]. The effect of amino acid changes in the transmembrane β-strand loop 3 that constitutes the porin channel eyelet that was associated with diminished carbapenem uptake in the endemic K. pneumoniae carbapenemase-producing K. pneumoniae ST258 clone illustrates the complexity of interpreting amino acid substitutions identified by WGS in the absence of experimental evidence [36,37]. Recent studies have shown that the genetic basis of resistance to colistin in K. pneumoniae clinical isolates can be attributed in the majority of cases to alterations in the mgrB regulator or the two component systems promAB or phoE that regulate the expression of the biosynthesis pathway of lipid A [29,38]. However, incorrectly inferring susceptibility remains the risk if resistance is mediated by genuinely novel, undiscovered genetic factors, as evidenced by the description of mcr-1, the first known transferable colistin resistance determinant [18].

The gaps in the existing knowledge of genotype–phenotype relationships could be augmented in some cases by directly detecting the levels of gene expression by sequencing RNA extracts. This approach was successfully used as a proof-of-concept for the detection of omfp down-regulation associated with cephalosporin resistance and over-expression of the resistance-nodulation-division pump component acrB leading to decreased susceptibility to quinolones, tetracycline and chloramphenicol in an E. coli laboratory-selected mutant [39]. Although a similar approach should also be feasible for the detection of hyper-production of intrinsic chromosomal β-lactamases (e.g. AmpC in Enterobacter spp.) the use of such methods, for which accuracy is highly dependent on bacterial growth conditions, is likely to be confined to a sub-set of laboratories in the foreseeable future and will not be considered further for the purposes of this document, which seeks to examine widely used techniques only.

Summary

The relatively limited number of acquired resistance genes and resistance-associated mutations that dominate epidemiologically in the Enterobacteriaceae (compared with the large number of those that have been reported in the resistome) could explain the high levels of accuracy of genotype–phenotype correlation in published studies and means that well-informed screening approaches can be very accurate. However, susceptibility to some antimicrobial agents will be harder to predict than for others and understanding the full range of mechanisms and their interplay will require more study if improved levels of accuracy across large, genetically diverse data sets are to be achieved.

Salmonella spp.

Background

Molecular mechanisms conferring reduced susceptibility to antimicrobial agents are relatively well characterized in Salmonella spp. The majority of these are encoded by horizontally transferable genes, which potentially makes genotypic detection a reliable alternative to phenotypic testing as these genes generally assemble into full-length genes when using short-read sequencing data, as long as the quality and quantity of these are adequate to produce good assemblies. In addition to acquired gene-based AMR, mutationally acquired AMR also exists in Salmonella spp. The clinically most important examples currently are single or double mutations in the gyrA DNA gyrase and parC topoisomerase genes leading to reduced susceptibility to quinolones and fluoroquinolones. Resistance to third-generation cephalosporins due to acquired extended-spectrum and AmpC β-lactamases is also clinically important to consider.

Published studies

Very few comprehensive studies have been published where phenotypic susceptibility data have been compared with the underlying molecular mechanisms identified in WGS data sets from a collection of Salmonella isolates. Zankari et al. used a set of 50 Salmonella enterica serovar Typhimurium isolates originating from pigs and previously tested phenotypically against 17 different antimicrobial agents as part of the DANMAP surveillance program [20,40]. WGS was performed on these 50 isolates, of which 49 produced sufficient WGS data to create draft genomes for analysis with the ResFinder web-tool. Here, complete agreement (100% sensitivity and specificity) between tested and predicted susceptibility/resistance phenotypes (S/R) was observed [20]. However, this perfect correlation between phenotypes and genotypes generated by ResFinder was to some extent biased by (a) the fact that none of the isolates showed phenotypic resistant and monoresistant fluoroquinolones—which would, in most cases, have been unnoticed by ResFinder as it currently does not detect chromosomally acquired mutations leading to AMR; and (b) there was a relatively low level of diversity among the resistance phenotypes and hence resistance genes found.

In addition, a few studies exist where only a small number of isolates have been analysed using both phenotypic and genotypic methods. In a study of extremely drug-resistant Salmonella enterica serovar Senftenberg, two isolates from Zambia were analysed both by conventional phenotypic methods and by WGS analysis [41]. Here, genes conferring reduced susceptibility to nine drug classes including fluoroquinolones and extended-spectrum cephalosporins, were identified, again with the use of the ResFinder web-tool. These isolates also demonstrated high-level resistance to fluoroquinolones caused by mutations in GyrA (S83Y and D87G) and ParC (S80I), which were identified manually. An underlying molecular mechanism was identified for all the AMR phenotypes displayed by the two isolates. In a similar study, two ESBL-producing Salmonella enterica serovar Typhi isolates were tested phenotypically against 25 different antimicrobial agents belonging to ten different classes [42]. Here, ResFinder found seven different AMR genes, which in combination could explain the observed AMR phenotypes of the two isolates.

Problems found or anticipated—gaps in the knowledge base

The overall degree of resistance within Salmonella spp. varies depending on the serovar and phage type; some may be associated with resistance to particular antimicrobial agents, whereas others may have an increased propensity for multidrug resistance. Diversity within the isolate panel tested will therefore be likely to impact on the conclusions drawn regarding the utility of WGS for AMR prediction in Salmonella spp. WGS should therefore be applied to further isolate panels reflecting the diversity of Salmonella serovars (and their common resistance phenotypes) associated with clinical and veterinary infections. Included in these panels should be representatives of some of the particularly multiresistant clones that are currently circulating in human and animal populations, e.g. serovars Kentucky, Infantis and monophasic Typhimurium. Nevertheless, recent data from the European Centre for Disease Prevention and Control indicate that nearly 55% of Salmonella spp. from humans are susceptible to all antimicrobial classes tested, suggesting that resistance prediction in Salmonella spp. may
be more straightforward than in other organisms, where multidrug resistance is the norm [43].

As with the other Enterobacteriaceae, detection of chromosomal mutations leading to acquired antimicrobial resistance is a challenge that still needs to be addressed to be able to predict resistance phenotypes fully based on WGS data. Priority should be afforded to the detection of mutations leading to fluoroquinolone resistance, as this drug class has high clinical relevance and phenotypic resistance is commonly detected in salmonellae. Development of fluoroquinolone resistance can be a multifactorial process involving acquisition of mutations leading to amino acid substitutions within the topoisomerase genes and altered expression of outer membrane proteins and/or multidrug efflux pumps. Fortunately, the most common chromosomal mutations leading to acquired fluoroquinolone resistance in salmonellae are well characterized; decision rules to translate mutations into a predicted phenotype are therefore available and can, in principle, be incorporated into existing tools such as ResFinder. This, however, requires detection of amino acid variation rather than nucleotide variation, which is currently used to detect (transferable) resistance genes. Other possible candidates for detection of chromosomal mutations are pmrA and pmrB, leading to reduced susceptibility to colistin [44].

Summary

The relatively limited studies that are available on the feasibility of using WGS data to predict antimicrobial resistance in salmonellae are showing promising results, but the impact of the sample sets tested on genetic diversity needs to be explored in detail before further conclusions are drawn about the use of WGS data for AMR prediction in Salmonella spp.

Staphylococcus aureus

Background

Staphylococcus aureus exhibits intrinsic susceptibility to commonly used antimicrobial agents. Resistance is associated with mutations in core genes or the acquisition of specific antimicrobial resistance genes. Generally, the history of antimicrobial resistance in this species is associated with the evolution of resistance shortly after the first introductions of a new agent into clinical practice, for example resistance to methicillin was detected in clinical isolates of S. aureus within a year of introduction into the UK [45]. The problem of resistance in this species has driven extensive studies to identify the genetic basis of resistance, and as such there is a large body of literature documenting resistance mechanisms for most of the clinically relevant agents used. This has revealed a well-characterized spectrum of mechanisms bestowing resistance in S. aureus, and in the cases of some AMR determinants, in other staphylococcal species and also in other genera. This has proved to be a valuable resource for the in silico prediction of antimicrobial resistance and has contributed to the overall value of the results.

Published studies

To date, several studies have been published that demonstrate the ability to predict antimicrobial resistance from genome sequence data [46–50]. Initial studies examined the concordance between phenotype and genotype [46–48], and more latterly blinded studies have made predictions [49]. In addition, bioinformatics resources have been produced that will generate predictions of antimicrobial resistance from genome sequence data [51].

The initial demonstration of the potential of WGS data to predict AMR as a clinical tool came from proof-of-concept studies using the benchtop Illumina MiSeq platform to investigate suspected methicillin-resistant S. aureus (MRSA) outbreaks. Köser et al. sequenced 14 isolates belonging to four different clonal complexes of S. aureus, and demonstrated 100% concordance of the in silico resistance prediction with the phenotypic results for 13 different agents (cefoxitin, erythromycin, ciprofloxacin, gentamicin, tetracycline, rifampicin, fusidic acid, mupirocin, clindamycin, kanamycin, tobramycin, trimethoprim and linezolid) [46]. The authors used an in-house database of resistance determinants derived from literature mining and mapped sequence reads to a resistome pseudo-molecule (concatenated resistance genes in a single DNA sequence), followed by manual inspection to predict the resistance profile of each isolate. Investigating an MRSA cluster in an intensive care unit, Eyre et al. sequenced ten isolates belonging to the same spa-type (t5973) and conducted in silico predictions for penicillin and tetracycline [47]. The authors took a different bioinformatics path to investigate the antimicrobial resistance, using the de novo assemblies to look for the presence and absence of two genes, tetK and blaZ. In all cases the presence of these genes correlated with the phenotypic resistances to respective agents.

Examining 13 isolates belonging to USA300 clone, Lee et al. predicted the antimicrobial resistance profiles for nine agents (ciprofloxacin, clindamycin, doxycycline, erythromycin, gentamicin, oxacillin, tetracycline, trimethoprim-sulfamethoxazole and vancomycin) in complete concordance with the phenotypic results [50]. The authors generated in silico predictions from short-read data using SRST2, a mapping-based tool for fast and accurate detection of genes and alleles from WGS data from a user-defined database [52]. In the case of this study, details of the antimicrobial resistance database used were not provided.

The effectiveness of in silico prediction for antimicrobial resistance in S. aureus has been further demonstrated in larger studies, both in terms of the number of isolates and also the agents investigated. Using the genome data of 193 isolates belonging to a global collection of ST22, Holden et al. applied a mapping based approach, coupled with manual inspection, to identify molecular determinants that explained 99.8% of the measured phenotypic resistance traits [48]. In total, 847 resistance traits were tested for 18 agents (penicillin, oxacillin gentamicin, linezolid, erythromycin, clindamycin, ciprofloxacin, fusidic acid, mupirocin, moxifloxacin, trimethoprim-sulfamethoxazole, tetracycline, vancomycin, teicoplanin, rifampicin, fosfomycin, tigecycline and daptomycin), using an enhanced version of the library used by Köser et al. [46] in their earlier study.

Using the WGS data and phenotype data for 501 S. aureus isolates as a derivation set to optimize predictions for 12 agents (penicillin, methicillin, erythromycin, clindamycin, tetracycline, ciprofloxacin, vancomycin, trimethoprim, gentamicin, fusidic acid, rifampicin and mupirocin), Gordon et al. then conducted a blind validation of their refined method on a query set of 491 isolates and demonstrated sensitivity of 0.97 and specificity of 0.99 [49]. As with their previous study, the authors used de novo assemblies and BLASTn for their in silico prediction [47]. In their resistance database 18 acquired genes were included: blaZ, mecA, msr(A), erm(A), erm(B), erm(C), erm(T), tet(K), tet(L), tet(M), vanA, fusB, far, dfrA, dfrG, aacA-aplD, mupA and mupB, and variation in six core genes: gyrA (n = 6), gyrB (n = 13), grlB (n = 6), fusA (n = 59), rpoB (n = 28) and dfrB (n = 8).

In a significant departure from previous studies in S. aureus, Bradley et al. described a stand-alone tool, MYKROBE PREDICTOR (http://www.mykrobe.com/), for antimicrobial resistance prediction directly from FASTQ files and which does not rely on mapping or assembly-based approaches [51]. This tool uses a de Bruijn graph-based approach to compare sequence reads to a reference graph representation. The method has the advantage of being faster than the mapping and assembly-based methods, and it can also identify minority variants in sequencing data and therefore identify potential contamination issues. In their study, Bradley et al. used 495...
isolates as a training set, and then validated the tool with a collection of WGS data from a further 471 isolates. The tool uses the Gordon et al. database [49] with some additional refinements and makes predictions for the same 12 agents. Using the tool, Bradley et al. were able to demonstrate sensitivity of 99.1% and specificity of 99.6% for the genotypic predictions in comparison with the phenotypes.

**Problems found or anticipated—gaps in the knowledge base**

**Genetic instability.** The evolution of antimicrobial resistance in *S. aureus* occurs by point mutation in core genes, and also by horizontal acquisition of resistance genes via mobile genetic elements. In the studies conducted so far, it is apparent that in some cases the relative genetic instability of mobile genetic elements carrying resistance genes can be a cause of discrepancy in the genotypic and phenotypic comparisions. One of the agents most prone to this problem is erythromycin. In *S. aureus*, genes encoding erythromycin resistance are often found on plasmids, such as in the case of *erm(C)*. The instability of the *erm(C)*-carrying plasmid has been well documented, and it can be lost during passage of isolates in the laboratory. In the studies by Holden et al. [48] discrepancies in erythromycin resistance prediction were thought to be due to the loss of the *erm(C)* during propagation and transfer between testing and genomics laboratories. Similar observations about the loss of the SCCmec element from the chromosome have also been made, which can account for discrepancies in cefoxitin resistance, albeit at a far lower frequency. In this case, evidence of the deletion of the whole SCCmec element carrying the *mecA* gene encoding erythromycin resistance can be observed. Genetic stability of core components can also affect the observed resistance levels for some agents in *S. aureus*. Hetero-resistance has been observed whereby a sub-population of cells in a cultured population exhibits a higher MIC than their ‘siblings’. WGS of sub-populations has uncovered genetic variation associated with hetero-resistance to vancomycin, daptomycin and oxacillin [53,54].

**Gaps in the knowledge base.** Although the studies that have been published so far have generally demonstrated the effectiveness of *in silico* resistance prediction, there is evidence emerging that the performance for some agents will be less accurate than for others. Aanensen et al. [55] recently conducted a blinded study of 308 isolates and phenotypically tested a range of agents (16 agents were tested against all isolates: penicillin, cefoxitin, ciprofloxacin, moxifloxacin, amikacin, gentamicin, tobramycin, erythromycin, clindamycin, tetracycline, tigecycline, fusidic acid, linezolid, mupirocin, rifampicin, trimethoprim; and for MRSA isolates teicoplanin, vancomycin and daptomycin were also tested). Overall, the total performance of the *in silico* prediction was in line with previous studies, with 98.6% concordance, although for some agents, such as amikacin (92.5% concordance) and teicoplanin (97.5% concordance), the *in silico* prediction proved less effective.

For some agents there are clearly gaps in the knowledge base for the genetic basis of resistance that require further investigation. For example, in the case of glycopeptides, such as vancomycin, the multiplicity and diversity of mutational changes that are suspected to be linked with increased MICs in *S. aureus* is so great that they confound an accurate prediction of susceptibility or resistance [56–58].

**Laboratory variation.** Discrepancies between the genotypic and phenotype in some studies have been revealed to be laboratory artefacts and errors, where phenotypic re-testing led to concordance. Technical variation in some of the tests is also a possible contributing factor in mis-matches. For example, in the study by Aanensen et al. [55] it was noted that in the case of five isolates that had incorrect *in silico* predictions for mupirocin. The inhibition zone of these isolates (all at 29 mm) were so close to the ECOFF (WT ≥ 30 mm), that it cannot be ruled out that technical variations in the phenotypic AST might have influenced results or that it signals that a revision of the zone diameter ECOFF for mupirocin is needed.

**Summary**

The *in silico* prediction of AMR susceptibility for *S. aureus* is effective for most clinically relevant agents. There are, however, some agents for which it is more challenging to make predictions and further investigation is required to characterize the genetic and phenotypic basis of resistance.

**Streptococcus pneumoniae**

**Background**

*Streptococcus pneumoniae* is clinically a highly important community pathogen, in which genetic detection of acquired resistance is particularly challenging because most resistance results from development of mosaic genes or mutations in chromosomally encoded genes [59]. The most clinically important groups of antimicrobial agents with activity against pneumococci are the β-lactams, macrolides, tetracyclines and newer fluoroquinolones (e.g. moxifloxacin). WGS-based approaches have been used for characterizing resistance mechanisms for several of these antimicrobial groups. However, no specific, user-friendly database has been developed so far, and WGS has mostly been used to study new mechanisms, and not in the context of predicting phenotypic resistance from whole genome data.

**Published studies**

For β-lactams, resistance is mostly mediated through the development of mosaic genes encoding altered penicillin-binding proteins (PBP), as a result of intraspecies and interspecies DNA transfer by natural transformation [60]. Variants of PBP2x, PBP2b and PBP1a are considered most relevant to penicillin resistance in pneumococci. However, there have also been reports of non-PBP-mediated resistance, such as enrichment in branched-chain murpeptides and mutations in genes encoding other enzymes involved in the peptidoglycan synthesis [61]. Fani et al. used WGS to study mutants selected for penicillin resistance by stepwise penicillin increments until reaching a final MIC of 2 mg/L [61]. Sequencing was done by the 454 platform, generating a genome assembly of 28 × coverage, with 97% of the reads assembled into 78 large contigs. Comparative sequence analysis identified mutations that were confirmed by Sanger sequencing. PBP2x mutations were shown to be important, but the relationship between genotypic and phenotypic resistance was complex, with a mutated iron transport system found in several of the resistant mutants.

Later work by the same research group proposed phenotypic reconstruction by whole genome transformation of penicillin-susceptible *S. pneumoniae* of known genetic backgrounds with genomic DNA from resistant clinical isolates [62]. This procedure would then be followed by WGS of the antimicrobial-resistant transformants. Selection of transformants was done by gradually increasing increments of penicillin concentrations. The genome sequences of the fully resistant and intermediate-step transformants were compared with the reference genome of the wild-type *S. pneumoniae* strains used in the experiments. The study confirmed the importance of mosaic PBP2x, PBP2b and PBP1a, but also suggested a role for PBP2a in some isolates. In another study, analysis of cefotaxime-resistant mutants revealed mosaic PBP as well as mutations in other genes important for peptidoglycan synthesis [63]. Although these data suggest that predicting phenotypic β-lactam resistance based on WGS could be feasible in
S. pneumoniae, there are so far no studies with clinical isolates to confirm this.

Macrolides also have clinically important activity against S. pneumoniae. Resistance is often mediated through RNA methylase (erm) or macrolide efflux (mef) genes, both of which are coupled to mobile genetic elements. One study was conducted in the USA with 147 pneumococcal isolates collected over an 18-year period both before and after the introduction of conjugate vaccines [64]. Genomes were then compared and mapping of macrolide resistance genes and their genetic environment was carried out. Resistance genes were detected in all isolates, but the study was in no way investigator-blinded, as all included isolates were macrolide-resistant by phenotypic methods.

Lupien et al. investigated mutants selected for resistance to tetracyclines [65]. Resistance to tetracycline in bacteria occurs through enzymatic inactivation or, more often, by active efflux (via intrinsic or acquired pumps) or by ribosome protection. Resistance to tetracyclines in pneumococci is very common, and most often mediated by tet genes, which are found on mobile genetic elements. Lupien et al. used WGS to investigate not only genomic DNA, but also RNA sequencing libraries depleted of rRNA. RNA expression was compared in parent strains and mutants, identifying differentially expressed genes. Quantitative RT-PCR was used to confirm over-expression of some of the genes identified by comparison of sequenced mRNA in mutants and wild-type strains. Gene ontology classification of genes whose expression is significantly altered in S. pneumoniae therefore seems to be a feasible way of studying new chromosomal resistance mechanisms, although this approach has not been used on clinical isolates. Finally, new genomic DNA transformation combined with WGS has also been used to study isolates with resistance to ciprofloxacin [66]. In addition to identifying efflux (using quantitative RT-PCR) and quinolone-resistance-determining region mutations, the methodology could also point to the potential role of mutations in drug transporters and redox enzymes in ciprofloxacin resistance.

Summary

A number of mechanistic studies, including whole genome transformation, have been carried out with laboratory mutants. These studies have shed light on a number of putative new mechanisms, but there is at present a lack of studies of the utility of WGS for predicting phenotypic resistance to agents used in the treatment of S. pneumoniae.

Neisseria gonorrhoeae

Background

Gonorrhoea is a global public health concern with the World Health Organization estimating 106 million cases every year [67]. Neisseria gonorrhoeae has developed resistance to every class of antimicrobial used to treat infections caused by the organism [68]. The emergence of multidrug-resistant N. gonorrhoeae has led the US CDC to classify it at an urgent threat level requiring serious public health attention [69]. Empiric combination therapy with intramuscular ceftriaxone and oral azithromycin is recommended for both ceftriaxone and cefixime at ≤0.25 mg/L, although no breakpoint for resistance has been published [72]. EUCAST established a clinical resistance breakpoint for ceftriaxone and cefixime at >0.125 mg/L [73]. However, treatment failures with cefalosporin monotherapy have recently been observed in a number of countries in cases where the cefixime MICs for the infecting gonococci were as low as 0.032 mg/L [74–78]. Although there are no azithromycin breakpoints established by CLSI, susceptible and resistant breakpoints have been set by EUCAST at ≤0.25 mg/L and >0.5 mg/L, respectively [73]. In addition, the CDC has used a resistance breakpoint for azithromycin at ≥2 mg/L [79].

Numerous genetic mechanisms exist in N. gonorrhoeae for the development of elevated MICs to the extended-spectrum cephalosporins (ceftriaxone and cefixime). Alterations in penA, which encodes PB2, have been described either through amino acid alterations (A501, G542, P551) or through the acquisition of a penA mosaic allele, which contains segments of penA from non-gonococcal Neisseria species [80–82]. These alterations confer reduced susceptibility to extended-spectrum cephalosporins mediated by reduced binding to PB2 [83]. Upregulation of the MtrCDE efflux pump via a deletion in the promoter at –35 (A–del) or alterations of the MtrR repressor protein at positions G45D and A39T have also been associated with decreased susceptibility to the extended-spectrum cephalosporins [80,84]. A third mechanism for increased MICs to extended-spectrum cephalosporins involves alterations in the PorB1b porin at amino acid positions G120 and A121. These permeability changes may reduce entry of extended-spectrum cephalosporins into the cell leading to reduced susceptibility [80,84].

Published studies

Only a few published studies have used WGS to examine the phenotypic and genotypic antimicrobial resistance patterns observed in N. gonorrhoeae. The first such study examined the genomes of 236 isolates of N. gonorrhoeae collected in the USA from 2009 to 2010, and included 118 isolates with decreased susceptibility to cefixime (MIC ≥ 0.25 mg/L) [85]. The mosaic penA XXXIV allele was present in 115 of 118 isolates with decreased cefixime susceptibility and one other isolate carried a novel mosaic allele termed XLII [86]. No other alleles were clearly associated with reduced susceptibility to cefixime. A second study examined the genomes of 76 N. gonorrhoeae from numerous countries [87]. The mosaic penA XXXIV allele once again had the best positive predictive value, with this locus detected in six of seven cefixime-resistant isolates. Mutations in the mtrR/mtrCDE operon promoter region and penB gene mutations did not have such a strong predictive value, being found in only two of seven and three of seven cefixime-resistant isolates, respectively. A third study applied WGS to 169 Canadian isolates of N. gonorrhoeae with various antimicrobial resistance patterns [88]. There were 67 isolates with ceftriaxone MICs ranging from 0.125 to 2 mg/L. Of these, 40 (59.7%) harboured the penA mosaic, and all but one isolate had either porB mutations or the mtrR promoter mutations. Of the remaining 27 isolates without the penA mosaic, all had porB mutations and only a single isolate did not contain a mutation in the mtrR promoter. However, when isolates with low MICs of ceftriaxone (<0.032 mg/L; n = 65) were examined, one isolate (1.5%) was found to have the penA mosaic. 33 (50.7%) isolates contained mtrR promoter mutations, and five isolates also contained mutations in porB.

A second highly clinically significant phenotype that has been examined for N. gonorrhoeae is azithromycin resistance. The rates of azithromycin resistance in N. gonorrhoeae have been increasing in recent years in many countries, and the emergence of high-level resistance (MIC ≥ 256 mg/L) has been reported from the UK, Argentina, Canada and the USA [89–93]. The genetic mechanisms of resistance to azithromycin in N. gonorrhoeae include: accumulated changes in the four different alleles of the 23S rRNA genes; the presence of a 23S rRNA methylase encoded by erm(A),erm(B),erm(C) or erm(F); mutations in rplD and rplV; as well as the penB
and mtr operon genes described above for cephalosporin resistance [94–96].

Although reports describing WGS-based detection of these mechanisms are limited, there have been attempts to compare the azithromycin resistance phenotype with the WGS genotype and these are discussed below.

**Problems found or anticipated—gaps in the knowledge base**

Ezewudo et al. examined two isolates that were resistant to azithromycin and found that one isolate contained the 23S rRNA mutation whereas the other did not contain any of the known mutations examined [87]. In a second study, involving the WGS analysis of 213 Canadian azithromycin-resistant isolates, 23S rRNA mutations A2045G and/or C2597T, disruptions in the mtrR promoter, or the presence of erm(C) were strongly associated with phenotypic resistance [88]. Seventy N. gonorrhoeae contained only the mtrR deletion, and of these 21 were susceptible to azithromycin, suggesting that other potential but unknown mechanisms conferring resistance may exist.

Although there is strong association with penA mosaic alleles and reduced susceptibility to extended-spectrum cephalosporins, caution is required with this data set as it relates to predicting the extended-spectrum cephalosporin MIC phenotype. The study by Demczuk et al. suggested that reduced susceptibility to ceftriaxone remains complex, involving additional genetic markers [88]. In addition, there has been a report of the mosaic penA allele XXXVIII displaying a susceptible phenotype to extended-spectrum cephalosporins [97].

**Summary**

Inferring resistance to extended-spectrum cephalosporins and azithromycin in N. gonorrhoeae is possible with a high probability if certain genetic markers are detected in WGS data. However, the elevated MICs for some isolates result from combinations of multiple genetic changes, and further mechanisms of resistance have yet to be elucidated. Hence, predicting resistance to these antimicrobial agents can be problematic. Additional studies are required before the use of WGS can be advocated for use on a routine basis to predict resistance for these antimicrobial agents.

**Mycobacterium tuberculosis complex**

**Background**

Tuberculosis (TB) is caused by members of the *Mycobacterium tuberculosis* complex (MTBC) and, more rarely, by *Mycobacterium canettii* [98]. MTBC is monomorphic and strictly clonal and antimicrobial resistance is therefore only caused by chromosomal changes. These are single nucleotide polymorphisms (SNPs) in the vast majority of cases, but small in-frame insertions/deletions (indels) in essential resistance genes and large indels in non-essential genes are also possible [99,100]. By contrast, *M. canettii* is genetically diverse and displays evidence of recombination and lateral gene transfer [98]. The resistance mechanisms in this species are unclear [101–103].

Phenotypic susceptibility testing for TB is expensive, technically challenging and time-consuming owing to the slow growth rate of MTBC [104]. Current molecular AST assays only interrogate the most frequent mutations conferring resistance to a limited number of drugs. In theory, this limitation can be overcome by WGS. In practice, however, routine WGS of all isolates from TB cases is unlikely to be cost-effective if performed to predict antimicrobial susceptibility alone, despite the decreasing cost of WGS [105,106]. This is primarily because rates of drug resistance are low (typically <5%) in countries that can afford WGS [107]. Instead, the main driver for the introduction of WGS for all TB cases will be the desire to replace traditional typing techniques by the ultimate resolution provided by WGS to improve outbreak investigations [108–110]. Moreover, WGS from the initial liquid culture can replace current techniques for pathogen identification (WGS directly from a clinical sample is technically possible, but less reliable and prohibitively expensive for clinical practice at the moment [111,112]). In this scenario, the sequence data can also be used to rule in resistance at no additional cost [106].

**Published studies**

Several large-scale studies reporting WGS of MTBC have focused on elucidating the genetic basis of antimicrobial resistance, which complement more focused efforts [106,113–122]. Nevertheless, major gaps in this area remain, as discussed below. Several tools have been developed to automate WGS data analysis, although most do not meet clinical standards because they do not provide the necessary record keeping capabilities, have not been evaluated extensively, and often there are no plans to accredit them [51,106,123–127]. One exception is the analysis infrastructure that is currently being evaluated by Public Health England for clinical accreditation and routine use of WGS for TB in the UK [106].

**Problems found/anticipated—gaps in the knowledge base**

As discussed briefly in the section above on ‘Categories of systematic errors in WGS predictions of AMR’, three main challenges limit the utility of genotypic AST compared with phenotypic alternatives.

(a) **Systematic errors due to inadequate limit of detection of WGS.**

AST for TB is usually done on a significant fraction of the primary culture, as opposed to just one to three colonies from a primary agar plate, which is the approach taken for the vast majority of other clinically relevant bacterial pathogens. Resistance is deemed clinically significant if resistant organisms are present at or above a critical proportion, which is set at 10% for pyrazinamide and at 1% for the remaining antimicrobial agents, and reference standard phenotypic AST (i.e. the proportion method) is calibrated to detect resistance at this limit [128,129]. The limit of detection of traditional genotypic AST methods is poorer and depends on the assay and specific mutation, which can result in systematic false-negative results for strains with low-level hetero-resistance [130–132]. The magnitude of this source of error depends on several factors, including the frequency of mixed infections with unrelated strains that have different susceptibilities and the proportion of resistance that is transmitted versus resistance that evolves during treatment [133]. In practice, these factors vary between patient groups, geographic settings and antimicrobial agents. Moreover, the precise mechanism of resistance is relevant in this context. Low-level hetero-resistance SNPs can be identified by increasing the sequencing coverage, although this makes WGS prohibitively expensive in a clinical context at the moment [106,134]. This strategy is not an option for hetero-resistance indels, particularly large ones, because of the limited read lengths of the most commonly used platforms for clinical sequencing, coupled with the fact that most analysis algorithms are not optimized to identify indels [135].

(b) **Systematic errors due to poorly defined breakpoints for phenotypic AST used as the reference standard for the validation of WGS-based AST.**

The clinical breakpoints, known as critical concentrations (CCs) in the TB field, are currently defined by the CLSI and the WHO [136,137]. Clinical breakpoints should be defined by committees based on representative MIC distributions, pharmacokinetic–pharmacodynamic data and,
The complexity of this task depends on whether a resistance gene is essential or non-essential. In the former case, only a limited spectrum of resistance mutations is possible. Consequently, the correlation between genotype and phenotype should be relatively easy to resolve, provided that methodological problems such as poorly defined CCs are addressed (e.g. there is a near perfect correlation between genotype and phenotype for rifampcin resistance and rpoB mutations, although this can depend on the medium used for phenotypic AST) [145–147]. The situation with non-essential genes is more complicated. For genes that are non-essential in vitro as well as in vivo, it is impossible to study the genetic basis of resistance comprehensively given that there are too many possible resistance mutations. The best example of this type of resistance gene is pncA, which is responsible for the activation of the pro-drug pyrazinamide [148]. Any loss-of-function mutation in this gene can confer resistance and a wide variety of mutations are found clinically (e.g. more than 4000 single codon changes are possible in pncA, excluding start codon changes and nonsense mutations, not all of which will cause resistance). By combining large data sets it is possible to distinguish resistance mutations from neutral polymorphisms, but novel mutations will continue to be discovered, albeit at a lower rate over time [122,149]. Similar considerations apply to the catalase-peroxidase KatG, which is required for the activation of the pro-drug isoniazid [150]. The gene encoding KatG has been found to be non-essential in vitro, but clinically most resistance-associated mutations within it are not selected, suggesting that they may be counter-selected [151]. By contrast, the KatG S315T change confers only a low fitness cost [152]. Consequently, this mutation accounts for the majority of isoniazid resistance clinically [122]. The remaining isoniazid resistance is due to a large number of rare mutations that are impossible to study in their entirety.

**Summary**

Some of the aforementioned challenges to introducing and validating WGS for AST of TB can be overcome over time. For example, the ability of WGS to detect hetero-resistance will improve as the cost of sequencing decreases and read lengths improve. Similarly, the ongoing re-evaluation of CCs will probably resolve some of the current systematic differences between genotype and phenotype. Moreover, the pooling of large data sets will clarify the role of rare resistance mechanisms and the level of resistance conferred by different resistance mutations or mechanisms. For example, some low-level isoniazid-resistant strains due to inhA mutations remain treatable with higher doses of the agent and the same may apply for some strains with low-level resistance to new-generation fluoroquinolones (codon 90 mutations of gyrA) [136,153,154]. However, it is impossible to study the genetic basis of antimicrobial resistance to all clinically relevant agents comprehensively because of the large number of possible resistance mutations for some agents. This means that WGS can mainly be used to rule in resistance, as opposed to rule out resistance. Nevertheless, this constitutes a significant improvement in current clinical practice, as WGS directly from the first positive culture would allow for established resistance mutations to key agents to be identified rapidly, thereby allowing treatment regimens to be adjusted within days as opposed to weeks or even months for phenotypic AST [155]. Based on these results, reference laboratories could also immediately commence phenotypic AST for all remaining relevant agents, including second-line agents (which are usually only tested if resistance to first-line agents is found, which introduces long delays). Consequently, WGS is unlikely to completely replace phenotypic AST for TB in the near future, but will result in less phenotypic testing over time and in more rapid identification of resistant isolates in many cases. However, it is likely that different countries will adopt their own policies in terms of how much phenotypic confirmation of genotypic results is required, based on the resources available and the local rates of resistance.

**Clostridium difficile**

**Background**

*Clostridium difficile* is the leading cause of healthcare-associated diarrhoea, the severity of which may vary from mild and self-limiting symptoms to fulminant disease, including pseudomembranous colitis. Hospital outbreaks are occurring with an increasing frequency, and the most severe outbreaks have been caused by hypervirulent *C. difficile* strains 027/NAP1, although other ribotypes (such as 078/NAP7&8) also seem to have the ability both to cause outbreaks and severe disease in affected individuals.

Acquired phenotypic resistance to tetracyclines, clindamycin, fluoroquinolones and rifampicin and corresponding resistance genes have frequently been reported in *C. difficile* and moxifloxacin resistance is used as an epidemiological marker for hypervirulent strains, and for ribotype 027 (NAP1) in particular. However, resistance to the agents that are used as primary therapeutics for *C. difficile* infection (i.e. vancomycin, metronidazole and fidaxomicin) is less common.

Phenotypic AST of *C. difficile* suffers from some drawbacks. As anaerobic conditions are required, it is costly and time-consuming and therefore often not routinely performed in the clinical laboratory and the correlation between in vitro susceptibility and clinical outcome in the individual patient has not been thoroughly studied. In light of this, genotypic AST using WGS appears an attractive alternative. Although resistance to fidaxomicin has been associated with mutations in genes encoding RNA polymerase (rpoB and rpoC) or in the marK homologue CD22120, the mechanisms underlying resistance to vancomycin and metronidazole are less well defined.

Some of the main lineages of *C. difficile* contain a *vanG* locus, which is expressed but does not appear to play a role in resistance...
to vancomycin in *C. difficile* [156]. To date, no clinical isolates have been identified that are resistant to vancomycin. Two laboratory-derived vancomycin-resistant isolates have been described. One had a substitution mutation in the *rpoC* gene and the second had two mutations, one in *murG* (CD2725) and the second in a locus named CD3659 [157].

Nitroimidazole genes (*nimA–E*) associated with metronidazole resistance in other anaerobic species, including several species of the *Clostridium* genus, have not been described in *C. difficile*. The exact mechanism(s) behind reduced susceptibility to metronidazole in *C. difficile* still remains to be determined, although there have been several reports of strains exhibiting elevated MICs. Such reports of metronidazole resistance have all observed loss of the resistant phenotype after passaging or low temperature storage [158–160]. There has been one reported clinical isolate, 027/NAP1 from Canada, that initially had an unstable resistance phenotype, but after serial passage in the presence of metronidazole the phenotype became stable [161]. WGS comparisons between the resistant and reverted susceptible isolate revealed many SNPs between the two isolates [161]. Proteomic analysis of these isolates suggested that a multifactorial response may be associated with the high-level metronidazole resistance observed [162]. Hence, the genetic mechanism for metronidazole resistance in *C. difficile* remains elusive.

Gaps in the knowledge base

To date there have not been any publications of large-scale studies comparing phenotypic with WGS-based AST for *C. difficile*.

Summary

AST of *C. difficile* using WGS could be a useful tool, both for guiding the choice of treatment of the individual patient and for epidemiological purposes. However, the knowledge gaps regarding the mechanisms underlying resistance to several of the first-line treatment options pose a great challenge. Studies comparing WGS-based approaches with phenotypic testing are needed and future work on resistance mechanisms to frontline antimicrobials are required.

*Acinetobacter baumannii* and *Pseudomonas aeruginosa*

Background

Among non-fermentative Gram-negative bacteria *Pseudomonas aeruginosa* and *Acinetobacter baumannii* are important pathogens due to their ability to cause a variety of opportunistic infections, persist in the hospital environment and acquire antimicrobial resistance [163–165]. Genomic studies have shown that both *P. aeruginosa* and *A. baumannii* are associated with high genomic diversity and gene content due to frequent transfer/acquisition of mobile genetic elements, mobilization of insertion sequence elements, insertion sequence-mediated deletions and genome-wide homologous recombination [166–172]. In addition to being intrinsically resistant to many antimicrobial agents [11], the increase in multidrug-resistant and in particular carbapenem-resistant *P. aeruginosa* and *A. baumannii* has resulted in infections caused by extensively drug-resistant and even pan-drug resistant isolates with very limited or no validated therapeutic options [166,173–175]. High-risk clones of both species have been identified as responsible for the spread of important resistance genes, such as those encoding carbapenemases [175,176].

In both *P. aeruginosa* and *A. baumannii* acquired resistance genes are associated with various horizontally acquired resistance elements, although the majority of acquired resistance genes exist as gene cassettes in integron structures [175,177]. In *A. baumannii* these resistance elements are frequently clustered in antimicrobial resistance islands, which vary in structure, size and genomic location [167], and plasmid-borne resistance genes also contribute to the evolution of antimicrobial resistance in both species [175,178]. In addition, both species, and in particular *P. aeruginosa*, have an extraordinary capacity for modification of endogenous genes affecting functions such as membrane permeability, efflux, expression of intrinsic β-lactamases, antimicrobial targets and regulatory genes contributing to multidrug resistance [164,175,179].

Published studies

To date, few comprehensive studies have investigated the concordance between phenotypic AST and WGS-based resistance prediction for *P. aeruginosa* or *A. baumannii*. Kos et al. related phenotypic susceptibility data for meropenem, levofloxacin and amikacin to the genome sequences of approximately 390 clinical isolates of *P. aeruginosa* [180]. The sensitivity and specificity for genotypic inference of meropenem and levofloxacin resistance were 91% and 94%, respectively. In contrast, a genotypic marker for amikacin resistance was identified for only 60% of the amikacin non-susceptible isolates. In addition, 30 of 283 amikacin-susceptible isolates were found to harbour genes associated with amikacin resistance. This is in contrast to a study by Wright et al., where a strong association between amikacin resistance and the presence of *aphA6* and *armA* genes was observed in a collection of 75 clinical isolates of *A. baumannii* [181].

Although there is a lack of phenotypic–genotypic comparison studies with respect to prediction of clinical resistance, several genotypic studies have been performed for epidemiological purposes and to decipher mechanisms of resistance to various agents in selected resistant isolates [167,182,183]. These studies are important to identify both intrinsic and acquired genotypic resistance determinants associated with resistance to various agents. For instance, recent investigation of isogenic colistin-susceptible and colistin-resistant isolates of both *P. aeruginosa* and *A. baumannii* from single patients revealed novel determinants associated with colistin resistance [184,185]. Further, the use of WGS as a tool to predict antimicrobial resistance has recently been studied using 178 *A. baumannii* bacterial genomes to evaluate the antimicrobial resistance gene database ARG-ANNOT and it was shown that such an approach could be used as a routine test [13].

Problems found or anticipated—gaps in the knowledge base

Although prediction of antimicrobial resistance based on the presence of a relatively limited number of acquired resistance genes and chromosomal resistance-associated mutations might give high sensitivity and specificity, the major challenge with respect to both *P. aeruginosa* and *A. baumannii* lies in the identification or prediction of resistance due to chromosomal alterations resulting in modification of expression levels, particularly with respect to efflux pumps, outer membrane proteins and intrinsic β-lactamases. For instance, resistance to β-lactams in *A. baumannii* can occur due to the insertion of elements such as ISAba1 and ISAba125 upstream of the intrinsic β-lactamase genes *blaOXA-51* and *blaOXA-24*, increasing the expression of these genes and consequently resistance to cephalosporins and carbapenems, respectively [186,187]. Further, insertion sequence elements have been implicated in disruption of genes encoding outer membrane proteins that contribute to resistance, in particular to β-lactams [164]. Screening of genomes for insertion sequence elements in close association with resistance-associated genes, as well as for gene loss, will pose a significant challenge.

For *P. aeruginosa* the challenge is expected to be even greater due to the plethora of genes associated with intrinsic resistance and alterations in these genes or regulatory genes can confer resistance to several agents, even from different antimicrobial classes [179].
Further, alterations to one or, mostly, several of these mechanisms might be required to achieve clinical resistance (e.g. combination of decreased porin expression, increased efflux and/or increased β-lactamase expression)\[188\]. Altered expression of genes could be overcome by investigation of gene-expression analysis by RNA sequencing. However, specific studies on P. aeruginosa indicate that correlation between expression of genes on exposure to sub-MIC concentrations of antimicrobial agents and the genes implicated in intrinsic resistance is not always clearly observed\[189\].

Summary

In general these studies showed that prediction of resistance based on the detection of known acquired resistance genes and resistance-conferring mutations in antimicrobial targets can be used to investigate the phenotype-genotype relationship. However, additional comparative studies between phenotypic and genotypic methods using representative strain collections of P. aeruginosa and A. baumannii are required to evaluate the possibility of confidently predicting antimicrobial susceptibility/resistance by WGS. Further, for both species a greater understanding of the contribution to clinical resistance of alterations in intrinsic resistance genes is required. This will require not only WGS, but also knock-out and complementation studies of deleted/mutated determinants.

The epidemiological implications of using WGS

The epidemiology of AMR is determined by the spread of the host organisms harbouring resistance genes, and the spread of the resistance genes by different routes of horizontal gene transfer.

Classical methodologies used to study the epidemiology of AMR include strain genotyping with a variety of methods with a large variation in reproducibility and discriminatory power. This includes techniques such as multilocus sequence typing (MLST), pulsed field gel electrophoresis, variable number tandem repeat, multiple locus variable number tandem repeat analysis, amplified fragment length polymorphism and Enterobacterial repetitive intergenic consensus-based PCR [12]. Resistance genes can be identified by micro-array approaches, PCR and sequence analysis. A variety of molecular techniques are needed to identify and characterize epidemiologically relevant mobile genetic elements involved in the horizontal transfer of AMR genes, such as plasmids, conjugative transposons or genomic islands. Dedicated PCRs and sequencing are needed to identify the genetic environment of the AMR genes such as integrons and/or transposons and insertion sequences, and this is crucial to understand the epidemiological behaviour of specific AMR genes [190].

PCR-based replicon typing is most commonly applied for plasmid characterization in Enterobacteriaceae. Relaxase typing is more comprehensive and phylogenetically more informative, but is less discriminatory within the Enterobacteriaceae, where there are major concerns regarding resistance at present. Plasmid MLST and similar techniques, such as Double Locus Sequence Typing or restriction fragment length polymorphism, are used to sub-type plasmids. In addition, toxin/anti-toxin systems encoded on plasmids in Enterobacteriaceae may be key epidemiological determinants. As a result of its complexity, plasmid epidemiology is currently beyond the capabilities of most clinical microbiology laboratories and is labour-intensive even for the reference laboratory.

WGS opens a world of opportunities for enhanced (molecular) epidemiology of AMR because, in principle, all essential information needed to study the epidemiology of AMR will be available in the sequences obtained. WGS is particularly effective for identifying and characterizing clonal distribution of monomorphic species such as S. aureus [55] and M. tuberculosis [191] and high-risk clones such as CTX-M-15-producing E. coli O25:H4-ST131 [192] or K. pneumoniae carbapenemase-producing K. pneumoniae (e.g. ST11, 258 and 512) [193]. Importantly, WGS provides high-resolution typing information making most if not all of the traditional molecular typing approaches redundant. With its potential for an objective assessment of the gene content, such as presence of absence of resistance genes of particular public health importance, multicentre surveillance approaches would greatly benefit from the reporting of genomic resistance markers, obviating the need to rely on phenotypic AST profiles of doubtful inter-laboratory reproducibility.

However, WGS also has its weaknesses. For example, WGS is weak in managing direct repeats and insertions in plasmids and current bioinformatic cleaning often omits those from contigs. As a result, short-read WGS data can be misleading if studying plasmid-mediated outbreaks in which a broad host-range plasmid is moving freely between different species, in each of which it has a different phenotype.

To be able to analyse sequence output rapidly and identify all information needed for epidemiology as listed above, the AMR genes and plasmid types need to be determined in these sequences using genomic databases such as ResFinder or PlasmidFinder [194]. For pMLST-typing, identification of genomic islands, insertion sequences and transposons the genomic database PLACNET is available [195]. In silico arrays or PCRs are also commonly applied but short-read sequencing (e.g. as obtained with Illumina) is generally not sufficient to study the genetic environment of AMR genes.

Transfer (by transformation or conjugation) of a plasmid of interest into a ‘workhorse’ bacterium with a known genetic background, with or without subsequent plasmid enrichment during DNA extraction, will facilitate complete and correct plasmid sequence assembly. Long-read or single-cell sequencing (e.g. by PacBio or Oxford Nanopore) may be necessary [25,196], either alone or as a ‘scaffold’ for high-coverage short-read data (e.g. Illumina). However, both are beyond the capabilities of most clinical microbiology laboratories, and plasmid handling is relatively labour intensive. This may still leave specialized annotation problems, although direct annotation grammars can be helpful with these [198,199].

Since different users may have different demands for WGS data, a tiered approach can be applied.

Diagnostic information

(a) Rapid identification of a targeted set of AMR genes may provide important information at a clinical level. The output could vary from answering a specific dichotomous question (e.g. does a sample contain an ESBL-producer or an MRSA isolate) to a more complete resistance/susceptibility profile [20,200,201]. The output should be based on the bacterial species and the information required for the clinicians.

(b) The question of the positive and negative predictive values of WGS will be important, although their usefulness will depend on the targets to be identified, their diversity and their prevalence in the gene pools [202,203].

Epidemiological information

(a) Identification of genes and subtypes (e.g. blaCTX-M-15 versus blaCTX-M-1 or blaCTX-M-3, or mecA versus mecC) may be important for outbreak management, infection control or even phylogeographic analysis.

(b) In silico strain typing in cultured organisms is based on:

- MLST (seven or more gene targets) for evolutionary relatedness;
- Genomic islands such as SCCmec, SG11, SXT;
• SNPs, insertions/deletions (indels), and large structural DNA rearrangements (e.g. for tracking outbreaks/ mapping transmission chains).

(c) *In silico* plasmid typing by mapping to a reference database (in *silico* microarray)

• Sub-typing of plasmids by *in silico* PCR (pMLST, double locus sequence typing, replicon sequence typing (RST))

(d) Phylogenetic analysis of the total sequence output.

WGS approaches can be used to track markers from the chromosome (e.g. MLST), from plasmids (e.g. incompatibility markers, or post-segregational killing/ toxic/antitoxin markers) and individual genes—barcoding to tie them together in individual isolates by WGS using third-generation approaches such as PacBio or Nanopore.

The ideal method will provide sufficient depth and coverage to answer all of these questions, but will vary with the starting material: metagenomic approaches to DNA extracted directly from clinical samples will require a considerably higher number of sequencing reads than for analysis of a microorganism in pure culture.

At the time of writing the availability of reference databases for epidemiological questions remains limited both in the number of species and typing methods that are represented. Further development in this area will be crucial.

**Clinical and wider impacts**

The routine use of WGS in diagnostic and public health laboratories holds the promise of a revolution in the identification, typing, antimicrobial susceptibility testing and determination of pathogenicity of potential pathogens [104]. At present, at the initiation of antimicrobial chemotherapy or thereafter when definitive therapy is selected based on phenotypic AST, clinicians have no routine data provided on the likely pathogenic potential of any pathogens isolated. Future data from WGS linking pathogenicity determinants to adverse clinical outcomes for certain highly pathogenic strains may have significant impacts on chemotherapy—perhaps by identifying those at higher risk of infection-related complications, those who may require more aggressive or combination chemotherapy or prolonged intravenous courses of antimicrobial agents. Conversely, reassurance that some potential pathogens are of low pathogenicity may allow for shorter duration therapy, oral therapy, less intensive patient monitoring, fewer investigations and perhaps earlier discharge. Such approaches are, at present, almost entirely speculative but may have a greater clinical impact than the work done so far on the value of WGS in predicting phenotypic susceptibility or resistance when tested by conventional methodologies.

At present, proof-of-principle studies have been completed for WGS on common pathogens already isolated in pure cultures and hence most data related to WGS for predicting antimicrobial susceptibility assume an initial culture step. This is an obvious limitation in terms of speed of diagnosis compared with direct testing, but rather we might focus on a limited number of agents for each pathogen initially and let more detailed data become available later.

At present, the use of WGS outside reference or research laboratories to determine antimicrobial susceptibility has not been tested. Preliminary data are promising and feasibility studies need to be conducted in a more clinical environment. It is likely that WGS will first be used as a tool to predict antimicrobial susceptibility in public-health microbiology laboratories in the coming years with subsequent use closer to the patient to predict susceptibility in pathogens such as *M. tuberculosis* before its wider application in diagnostic laboratories.

**Conclusions and recommendations**

This EUCAST Subcommittee report on the role of WGS in AST of bacteria has reviewed the state-of-the−art as a first approach. It refers to almost 200 published works and describes where we are at the time of writing (late 2015 to early 2016). Despite the volume of published literature already available we conclude that, at present, there are insufficient data to present a definitive document on the topic. Instead, this report is intended to form a baseline discussion document, which can be revisited and updated at regular intervals (probably every 18–24 months). This will be important as sequencing technologies become more affordable and more widely applied. This first version will provide the baseline against which to compare and assess future progress in the area.

We are aware of many ongoing, as yet unpublished, studies of phenotypic–genotypic AST concordance and it is certain that the amount of available data will increase in the near future. However, the quality of those data needs to improve and to be assured via more rigorous and ‘standardized’ approaches to data analysis.
Bacterial AST is a fundamental activity that can be undertaken in any microbiology laboratory, but it is important to appreciate that the MIC or zone diameter measured reflects more than gene presence/absence; these values reflect multiple and complex interactions between different systems including cellular permeability, influx/efflux, target availability and binding, as well as enzymatic expression levels and activities. So there are many challenges in gathering and assessing evidence to consider whether AST can be replaced by a genotypic method such as WGS, which does not assess bacterial growth in the presence of antimicrobial agents.

At the present time, WGS-based analyses cannot yield an inferred MIC or zone diameter. Hence the potential utility of WGS-based approaches for AST must be considered at the level of detecting gene presence or absence. We will need more powerful bioinformatics tools in future if we seek to make inferences about antimicrobial susceptibility based on combinations of multiple different genes or contributory mutations. Furthermore, WGS does not directly provide information on levels of gene expression. Although other technologies can do so, e.g. RNA sequencing, it seems unlikely that these will find a place in the clinical laboratory before WGS.

It is our recommendation that the primary AST comparator for WGS-based prediction should be the ECOFF, wherever possible, to assess WGS-inferred ‘antibograms’ (based on gene positivity) against phenotypically defined categories of wild-type or non-wild-type. Adoption of ECOFFs as the primary comparator would make meta-analysis across different publications simpler, as comparison of data would not be subject to confounding factors such as differences in breakpoints from different organizations. Nevertheless, demonstrating concordance with interpretation based on clinical breakpoints will ultimately be necessary for the use of WGS-based testing to guide clinical decision making, but this will probably be more difficult to demonstrate for all organisms and antimicrobial agents. For this reason, assessing WGS-derived data against clinical breakpoints represents a tougher challenge, but should be encouraged as a secondary comparator and should ideally be done using the same data sets that are used for ECOFF-based assessments.

The challenges of harmonizing antimicrobial susceptibility breakpoints across multiple parallel and independent national and international systems have been ongoing for >50 years, and we still lack a globally harmonized system. When considering the introduction of WGS-based approaches, we need to balance the needs of clinical laboratories, where standardized and validated procedures are required to meet accreditation standards, with the need for intellectual and innovative academic challenge, which drives many of those who generate bioinformatics tools. We recommend that there should be international agreement on the most appropriate and effective principles to facilitate early standardization and harmonization of analytical approaches and interpretative criteria for WGS-based predictive AST. However we also recommend at the present time that we need to be pragmatic and must accept that bioinformatics algorithms will vary. It is unrealistic to suggest a single analytical approach. We recommend that different bioinformatics tools should perform to minimum standards and should be calibrated and equivalent in terms of the results generated.

To facilitate such comparisons, we recommend that performance of different bioinformatics tools should be calibrated against a single database of all known resistance genes/mutations. There have been efforts and investments in this direction, but multiple solutions exist and are used, thereby confounding comparisons. Establishing a single database will ensure that there is parity of analysis and will facilitate measurement of comparative accuracies across different systems. Such a global reference database would need to be updated regularly, and must have strictly curated minimum standards for the inclusion of new resistance genes and mutations. An important function of a centralized database would be to control resistance gene nomenclature (since poor annotation can confound current analyses, where multiple ‘hits’ from searches may reflect inconsistent annotation of the same gene). The inclusion criteria for any new determinant would probably need to be set higher than those accepted for publication because strong evidence of causal association would maximize the predictive values of inferring AST phenotype from genotype.

The organisms considered in this report can be divided into three main groups in terms of the available evidence for predicting AMR using WGS. First, at present most is known for *S. aureus* and *M. tuberculosis* and it is apparent that there is now momentum behind their deeper investigation. For a second group of organisms, including the *Enterobacteriaceae* (including *Salmonella*), initial studies have shown promise, but serve to highlight, through poor concordance, where gaps exist in the knowledge base about resistance mechanisms either in some genera or species or for some agents. For a third group of organisms, including *S. pneumoniae*, *N. gonorrhoeae*, *P. aeruginosa*, *A. baumannii* and *C. difficile* it is apparent that more studies are required before we can even define the extent of the gaps in the knowledge base accurately. More focused study and additional funding resources are needed as a priority to improve knowledge for the second and third of these groups.

Expansion of the knowledge base is a critical priority if WGS is to be considered seriously as a rival to phenotypic AST, better defining resistance determinants across all organisms. It seems likely that WGS may replace phenotyping testing ‘soon’ for surveillance purposes, where the low error rate has low impact. This would need to be phased to reflect the evidence base for the organism—agent being reported, and would require surveillance schemes to expand their inclusion criteria to accept WGS-inferred data. In reference laboratories, WGS-based AST may also be adopted ‘soon’, unless the reason for investigation relates to individual patient management, is for agents or species shown to have poor genotypic—phenotypic concordance, or is to assess the activity of novel agents.

Available published evidence does not currently support use of WGS-inferred susceptibility to guide clinical decision making. Such a paradigm shift would require large-scale education and behavioural change among microbiologists and prescribers. Gene (or mutation) absence cannot always reliably predict susceptibility, so robust evidence will be needed to show that the potential of genotypic tests for very major errors does not adversely impact on treatment outcomes. It seems likely that this may first be considered for *M. tuberculosis*, where the speed of WGS-generated results offers advantage over traditional AST methods. However, even if the evidence can be generated and expectations changed, for most bacteria and in most countries the current cost and speed of inferring antimicrobial susceptibility from WGS data remain prohibitive to wide adoption in routine clinical laboratories. Nevertheless, as advances in the knowledge of polymorphisms associated with antimicrobial resistance, technology, data sharing and training become more widely available in high-burden countries, sequencing technologies will be more attractive and cost effective as the cost of goods comes down.

Finally, there may even be scope for WGS-based approaches to be used to better understand and improve some areas of phenotypic AST. For some agents, there are technical challenges in measuring susceptibility in any way that meaningfully correlates with outcome. If WGS data could be correlated directly with outcome, then this revolutionary tool might aid development of improved criteria for interpreting phenotypic data.

**Acknowledgements**

During the 2015 ECCMID meeting in Copenhagen, Neil Woodford was approached (separately) by Derek Brown and Gunnar
Kahlmeter, who were both very keen for me to put together a group to consider how well WGS can predict antimicrobial susceptibility patterns and how this game-changing technology could impact on clinical microbiology, now and in the future. This coordinated and two-pronged approach clearly achieved its goal, and shortly afterwards this EUCAST Subcommittee came into being.

Over the past year the team has ‘met’ virtually and this report marks its first output. We quickly agreed that there are too few data to present a definitive document on the topic, but that it would be necessary to review the state-of-the–art as a first approach. This document is therefore presented as a baseline and as a discussion document, and should be considered as such. It marks where we are now, and we present it in the knowledge that it will require updating, probably regularly, as sequencing technologies become more affordable and more widely applied, and as the analysis of the WGS data becomes more rigorous and standardized, and the quantity and quality of evidence for phenotypic–genotypic concordance (or lack thereof) relating to antimicrobial susceptibility improves.

Neil Woodford would like to thank all members of the Subcommittee for their efforts over the last year. It has been a pleasure working with them all. Special thanks are owed to Matt Ellington and Oskar Eklund, who compiled the multiple contributions to the report.

Thanks also to those who contributed comments during the consultation phase, many of which served to improve our arguments and the report.

The Subcommittee thanks Thomas Schon for editing and assistance with the TB evidence report. He is a member of the EUCAST Subcommittee on Antimycobacterial Susceptibility Testing and is supported by research grants from the Swedish Heart and Lung Foundation and Marianne and Marcus Wallenberg Foundation.

Thanks also to Mrs Sushma Udani for providing administrative support for the Subcommittee.

Transparency declaration

CUK is a consultant for the Foundation for Innovative New Diagnostics and was technical advisor for the Tuberculosis Guideline Development Group of the World Health Organization. The Bill & Melinda Gates Foundation and Janssen Pharmaceutica covered his travel and accommodation to present at meetings. The European Society of Mycobacteriology awarded CUK the Gertrud Meissner Award, which is sponsored by Hain Lifescience. CUK collaborated with Illumina Inc. on a number of scientific projects.

MJE, KLH, MD and NW are part of Public Health England’s AMR HAI Reference Unit which has received financial support for conference attendance, lectures, research projects or contracted evaluations from numerous sources, including: Accelerate Diagnostics, Achaogen Inc, Allegra Therapeutics, Amplex, AstraZeneca UK Ltd, Basilea Pharmaceutica, Becton Dickinson Diagnostics, BioMérieux, Bio-Rad Laboratories, The BSAC, Cepheid, Check-Points B.V., Cubist Pharmaceuticals, Department of Health, Enigma Diagnostics, Food Standards Agency, GlaxoSmithKline Services Ltd, Henry Stewart Talks, IHA Ltd, Kalidex Pharmaceuticals, Melinta Therapeutics, Merck Sharpe & Dohme Corp, Meiji Seika Pharma Co, Mobidiag, Momentum Biosciences Ltd, Nordic Pharma Ltd, Norgine Pharmaceuticals, Rempex Pharmaceuticals Ltd, Roche, Rokitan Ltd, Smith & Nephew UK Ltd, Trius Therapeutics, VenatoRx Pharmaceuticals and Wockhardt Ltd.

CG received conference support and had research collaboration with AB Biodisk (later purchased by bioMérieux); research collaboration with bioMérieux, Checkpoints, Q-linea; received speaker’s honoraria from BioRad, Liofillchem, Pfizer, Cepheid, Cubist; and carried out consultancy work for the bioinformatics company 1928 Diagnostics.

JI has received travel support and honoraria from Astra Zeneca, MSD, and Pfizer in the last 5 years for advisory board attendance and lectures. All current funding is from the National Health and Medical Research Council of Australia (NHMRC) grant 1001021.

RC has participated in educational programmes from Cepheid, Roche, AstraZeneca, MSD and Novartis and has received financial research support from Amplex, AstraZeneca, Cepheid, Cubist Pharmaceuticals, Ferrer International Laboratories and MSD.

GK serves as consultant for Oxoid Ltd on technical matters related to antibiotic disc quality, media for AST performance and Quality Control.

DM leads the Dutch National Reference Laboratory for Antimicrobial Resistance in Animals at Central Veterinary Institute in Leystad. Financial support is provided by the Ministry of Economic Affairs, the EU and he coordinates several public private partnerships in which 50% of the funds come from the public domain and 50% from animal-producing organizations including Aviagen, Vion Food Group and van Drie Group. Travel and accommodation is solely paid by public funds.

AMaC, JMR, MM, UH, OS, MTGHE, OE, TN, HG, FMA and TP have stated explicitly that there are no conflicts of interest in connection with this article.

References


