Using Genomics to Track Global Antimicrobial Resistance

Hendriksen, Rene S.; Bortolaia, Valeria; Tate, Heather; Tyson, Gregory H; Aarestrup, Frank Møller; McDermott, Patrick F.

Published in:
Frontiers in Public Health

Link to article, DOI:
10.3389/fpubh.2019.00242

Publication date:
2019

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

Link back to DTU Orbit

Citation (APA):
Using Genomics to Track Global Antimicrobial Resistance

Rene S. Hendriksen 1*, Valeria Bortolaia 1, Heather Tate 2, Gregory H. Tyson 2, Frank M. Aarestrup 1 and Patrick F. McDermott 2

1 European Union Reference Laboratory for Antimicrobial Resistance, World Health Organisation, Collaborating Center for Antimicrobial Resistance and Genomics in Foodborne Pathogens, FAO Reference Laboratory for Antimicrobial Resistance, National Food Institute, Technical University of Denmark, Lyngby, Denmark, 2 Center for Veterinary Medicine, Office of Research, United States Food and Drug Administration, Laurel, MD, United States

The recent advancements in rapid and affordable DNA sequencing technologies have revolutionized diagnostic microbiology and microbial surveillance. The availability of bioinformatics tools and online accessible databases has been a prerequisite for this. We conducted a scientific literature review and here we present a description of examples of available tools and databases for antimicrobial resistance (AMR) detection and provide future perspectives and recommendations. At least 47 freely accessible bioinformatics resources for detection of AMR determinants in DNA or amino acid sequence data have been developed to date. These include, among others but not limited to, ARG-ANNOT, CARD, SRST2, MEGARes, Genefinder, ARIBA, KmerResistance, AMRFinder, and ResFinder. Bioinformatics resources differ for several parameters including type of accepted input data, presence/absence of software for search within a database of AMR determinants that can be specific to a tool or cloned from other resources, and for the search approach employed, which can be based on mapping or on alignment. As a consequence, each tool has strengths and limitations in sensitivity and specificity of detection of AMR determinants and in application, which for some of the tools have been highlighted in benchmarking exercises and scientific articles. The identified tools are either available at public genome data centers, from GitHub or can be run locally. NCBI and European Nucleotide Archive (ENA) provide possibilities for online submission of both sequencing and accompanying phenotypic antimicrobial susceptibility data, allowing for other researchers to further analyze data, and develop and test new tools. The advancement in whole genome sequencing and the application of online tools for real-time detection of AMR determinants are essential to identify control and prevention strategies to combat the increasing threat of AMR. Accessible tools and DNA sequence data are expanding, which will allow establishing global pathogen surveillance and AMR tracking based on genomics. There is however, a need for standardization of pipelines and databases as well as phenotypic predictions based on the data.

Keywords: global, antimicrobial resistance, surveillance, genomic, bioinformatics tools, microbiology
INTRODUCTION

The science of infectious disease, along with other medical and biological specialties, is undergoing rapid change brought on by the advent of affordable whole genomic sequencing (WGS) technologies (1–3). These technologies are rapidly gaining acceptance as routine methods, and in the process, are transforming laboratory procedures.

The amount of bacterial genomic data being generated is immense. As of this writing, for example, over 190,000 *Salmonella* genomes alone are in the public domain with hundreds being added weekly. A complete genomic DNA sequence represents the highest practicable level of structural detail on the individuating traits of an organism or population. As such, it can be used to provide more reliable microbial identification, definitive phylogenetic relationships, and a comprehensive catalog of traits relevant for epidemiological investigations. This is having a major impact on outbreak investigations and the diagnosis and treatment of infectious diseases, as well as the practice of microbiology and epidemiology (4). Furthermore, DNA sequences are a universal dataset from which, theoretically, any biological feature can be inferred. In clinical applications, this includes the ability to detect antimicrobial resistance (AMR), and to track the evolution and spread of AMR bacteria in a hospital or the community.

AMR is a global health problem that contributes to tens of thousands of deaths per year [Chaired by Jim O’Neill, (5)]. Historically, AMR has been detected as a measurement of the growth inhibitory effects of a chemotherapeutic agent on a bacterial population cultured under specific laboratory conditions. Despite some ancillary enhancements, clinical laboratories to this day rely mainly on diffusion and dilution methods to guide clinical therapy and to monitor AMR over time. Accumulating data show that AMR can be accurately predicted from the genomic sequence for many bacteria. The sequence-based approach to AMR detection requires robust bioinformatics tools to analyze and visualize the genomic structure of the microbial “resistome,” defined by AMR genes and their precursors (6). This review summarizes the state of the science in using single isolate WGS to track global AMR.

THE ADVANTAGES OF WHOLE GENOME SEQUENCING

A major advancement enabling resistome surveillance is the demonstrated power to predict AMR from genomic data alone. Several studies including those focused on foodborne pathogens and *Enterobacteriaceae* have shown a high concordance (>96%) between the presence of known AMR genes or mutations and Minimum Inhibitory Concentration (MIC) of several antimicrobials at or above the epidemiological cut-off value or clinical breakpoint for resistance. High sensitivity of >87%, defined by the ability to correctly identify AMR determinants associated with an antimicrobial resistance phenotype (true positive rate) and high specificity of >98%, defined by the ability to correctly identify the absence of AMR determinants in an antimicrobial susceptible phenotype (true negative rate), have been observed depending on the bacterial species analyzed (Table 1) (7–18). Furthermore, a growing body of data shows that it is possible to predict AMR, and perhaps the MIC of an antimicrobial, applying machine or deep learning to genome sequence data (19–21). The comparison between phenotype and genotype as well as the application of machine or deep learning are however still in their infancy and additional data on bacterial species beyond the foodborne pathogen domain are needed.

The most obvious advantage of WGS for microbial typing and AMR surveillance is the unprecedented level of detail in one assay that can be used to describe current trends and distinguish emerging tendencies (22). AMR bacteria can be typed and traced by specific allele profiles, rather than just according to phenotypic patterns by drug class. This is exemplified by a study of emerging aminoglycoside-resistant *Campylobacter* in the USA, where WGS revealed that the rising trend was driven by nine different resistance alleles, six of which had never been detected in *Campylobacter* previously and would not have been found easily using PCR (10). Similarly, in one of the first large-scale applications of WGS to investigate a drug-resistant foodborne outbreak in the US in 2011, inconsistent resistance patterns among indistinguishable PFGE types of *Salmonella* serovar Heidelberg were revealed by sequence analysis to be a polymicrobial contamination event, involving various combinations of plasmids and strain types (23).

DNA sequence-based surveillance makes it possible also to define multidrug-resistance (MDR) with much greater precision compared to phenotypic tests (22). It has long been a common practice to define MDR as resistance to compounds from three or more drug classes (24), a definition with limited practical value. Bioinformatic analysis can reveal the co-carriage of specific genes underlying different MDR patterns, allelic trends over time, their genetic context including the potential for horizontal transfer, and their distribution by source. In addition, the presence of co-resistances not assayed on standard drug panels is revealed, such as disinfectant and heavy metal resistance. This level of “deep surveillance” can uncover other potential drivers of AMR persistence and evolution, and the opportunity for a more refined microbial risk analysis based on the association of resistance traits with specific sources.

ONLINE RESOURCES FOR IN SILICO ANTIMICROBIAL RESISTANCE DETECTION

The high level of agreement between phenotype and genotype coincides with the development of new and updated versions of bioinformatics tools to predict AMR, and the maturation of well-curated AMR gene databases. In principle, *in silico* AMR detection is performed by using a search algorithm to query input DNA or amino acid sequence data for the presence of a pre-determined set of AMR determinants contained in AMR reference databases (Figure 1). This can be performed using proprietary systems offered by commercial companies or open-access systems requiring different levels of user expertise.
<table>
<thead>
<tr>
<th>Year</th>
<th>Pathogen</th>
<th>No. of pathogens</th>
<th>AST method</th>
<th>No. of antimicrobials</th>
<th>Bioinformatic tool</th>
<th>Sequencing data</th>
<th>Concordance</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Comment</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td><em>S. Typhimurium</em></td>
<td>49</td>
<td>MIC</td>
<td>17</td>
<td>ResFinder</td>
<td>Assembled, Velvet</td>
<td>99.74%</td>
<td>96%</td>
<td>97%</td>
<td>Disagreement: 7 isolates including 6 E. coli resistant to Spec</td>
<td>(7)</td>
</tr>
<tr>
<td>2013</td>
<td><em>E. coli</em></td>
<td>48</td>
<td>DD</td>
<td>7</td>
<td>BLASTn, selected panel</td>
<td>Assembled, Velvet</td>
<td>96%</td>
<td>97%</td>
<td>97%</td>
<td>VM rate: 1.2%/M rate: 2.1%</td>
<td>(5)</td>
</tr>
<tr>
<td>2013</td>
<td><em>E. faecalis</em></td>
<td>50</td>
<td>MD</td>
<td>14</td>
<td>ResFinder</td>
<td>Assembled, Velvet</td>
<td>99.3%</td>
<td>99%</td>
<td>99%</td>
<td>VM rate: 0.5%/M rate: 0.7%</td>
<td>(5)</td>
</tr>
<tr>
<td>2013</td>
<td><em>K. pneumonia (ESBL)</em></td>
<td>69</td>
<td>DD</td>
<td>12</td>
<td>BLASTn, selected panel</td>
<td>Assembled, Velvet</td>
<td>97%</td>
<td>99%</td>
<td>96%</td>
<td>Lower concordance to VM (0.5)/M (0.2)</td>
<td>(5)</td>
</tr>
<tr>
<td>2014</td>
<td><em>S. aureus</em></td>
<td>501</td>
<td>DD/MIC (Vitek)</td>
<td>12</td>
<td>BLASTn, selected panel</td>
<td>Assembled, Velvet</td>
<td>97%</td>
<td>97%</td>
<td>97%</td>
<td>Lower concordance to VM (0.5)/M (0.2)</td>
<td>(5)</td>
</tr>
<tr>
<td>2016</td>
<td><em>C. jejuni</em></td>
<td>32</td>
<td>MIC</td>
<td>9</td>
<td>BLASTx</td>
<td>Assembled, CLC-bio</td>
<td>99.2%</td>
<td>99.2%</td>
<td>99.3%</td>
<td>Lower concordance to</td>
<td>(10)</td>
</tr>
<tr>
<td>2016</td>
<td><em>C. coli</em></td>
<td>82</td>
<td>MIC</td>
<td>14</td>
<td>ResFinder/ ARG-ANNOT/ CARD/BLAST</td>
<td>Assembled, CLC-bio</td>
<td>97.6%</td>
<td>98.0%</td>
<td>Lower concordance to aminoglycosides/β-lactams</td>
<td>(11)</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td><em>E. coli</em></td>
<td>31</td>
<td>MIC</td>
<td>4</td>
<td>Custom DB based on ARDB/CARD/β-lactamase alleles</td>
<td>Assembled, SPAdes</td>
<td>98.4%</td>
<td>98.4%</td>
<td>98.4%</td>
<td>Disagreement: 2/2 C. jejuni to FQ/ERY</td>
<td>(13)</td>
</tr>
<tr>
<td>2017</td>
<td><em>K. pneumonia</em></td>
<td>24</td>
<td>MIC</td>
<td>5</td>
<td>ResFinder/NCBI Pathogen DB/BLAST</td>
<td>Assembled, CLC-bio</td>
<td>96.5%</td>
<td>Neg. predictive value: 97%</td>
<td>(14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td><em>P. aeruginosa</em></td>
<td>22</td>
<td>MIC</td>
<td>4</td>
<td>GeneFinder/ Mykrobe/ Typewriter</td>
<td>FASTQ/assembled, BLAST</td>
<td>96.5%</td>
<td>Neg. predictive value: 91%</td>
<td>(14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td><em>S. enterica</em></td>
<td>50</td>
<td>MIC</td>
<td>4</td>
<td>ResFinder/NCBI Pathogen DB/BLAST</td>
<td>Assembled, CLC-bio</td>
<td>96.3%</td>
<td>Neg. predictive value: 91%</td>
<td>(15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td><em>E. faecalis</em></td>
<td>97</td>
<td>MIC</td>
<td>11</td>
<td>GeneFinder/ Mykrobe/ Typewriter</td>
<td>FASTQ/assembled, BLAST</td>
<td>96.3%</td>
<td>Neg. predictive value: 91%</td>
<td>(15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td><em>E. faecium</em></td>
<td>100</td>
<td>MIC</td>
<td>11</td>
<td>GeneFinder/ Mykrobe/ Typewriter</td>
<td>FASTQ/assembled, BLAST</td>
<td>96.3%</td>
<td>Neg. predictive value: 91%</td>
<td>(15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td><em>S. aureus</em></td>
<td>501</td>
<td>DD/MIC</td>
<td>12</td>
<td>GeneFinder/ Mykrobe/ Typewriter</td>
<td>FASTQ/assembled, BLAST</td>
<td>96.3%</td>
<td>Neg. predictive value: 91%</td>
<td>(15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td><em>M. tuberculosis</em></td>
<td>397</td>
<td>MIC</td>
<td>4</td>
<td>Cortex</td>
<td>Assembled</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>Phenotype issues to metronidazole</td>
<td>(16)</td>
</tr>
<tr>
<td>2019</td>
<td><em>H. pylori</em></td>
<td>140</td>
<td>MIC</td>
<td>2</td>
<td>4</td>
<td>ARIBA</td>
<td>FASTQ</td>
<td>99%</td>
<td>99%</td>
<td>99%</td>
<td>Phenotype issues to metronidazole</td>
</tr>
</tbody>
</table>


References:
(7) Hendriksen et al. Bioinformatics Resources for Prediction of AMR.
Open-access systems are available at public genome data centers such as the Center for Genomic Epidemiology (CGE) http://www.genomicepidemiology.org/ online or downloadable for local install from github (https://github.com/), bitbucket (https://bitbucket.org/account/user/genomicepidemiology/projects/DB) and similar.

The various bioinformatics software can process sequence data either as reads or as assemblies (25). Generally, available resources do not include quality control of input sequence data thus it is the users’ responsibility to ensure the quality of submitted sequences or assemblies. When using assembly-based methods, differences among assemblers may compromise comparability of the outcome (15, 26). Following assembly, the most common approaches to compare the input data with the AMR reference databases rely on BLAST and Hidden Markov Model searches, among others. BLAST-based tools can give different outputs based on default settings for gene length and percentage of similarity. This can negatively affect specificity if the settings are too low or too high. Moreover, assembly-based methods are computationally demanding. Despite these caveats, assembly-based methods may have an added value in an AMR surveillance context as they allow analysis of the genetic context of the AMR genes such as their presence on mobilizable potential. Read-based methods may use different tools to align reads to AMR databases, including Bowtie2, BWA, and KMA (25). Recently, the KMA (k-mer alignment) has been develop to map raw reads directly against redundant AMR databases (27). The KMA tool was developed specifically for rapid and accurate bacterial genome analyses in contrast to other mapping methods such as BWA that were developed for large reference genome, such as the human genome and subsequently applied empirically to microbiology (27). KMA uses k-mer seeding to speed-up mapping and the Needleman-Wunsch algorithm to accurately align extensions from k-mer seeds. Multi-mapping reads are resolved using a novel sorting scheme (ConClave scheme) to ensure an accurate selection of templates (27). Read-based methods allow identification of AMR genes present in low abundance which might be overlooked where assemblies are incomplete (25).

Independent of the bioinformatics approach chosen, the performance of in silico AMR prediction is critically dependent on the availability of accurate AMR databases. AMR reference databases can be subdivided into solutions specialized for detection of resistance to specific antimicrobials and/or in specific bacterial species or in solutions allowing detection of virtually any possible AMR determinant in any DNA/amino acid sequence. Besides their focus area, AMR reference databases have important differences which users need to acknowledge for choosing the optimal fit-for-purpose database. First, AMR reference databases differ for criteria of inclusion of entries. For example, entries in CARD must have been published in scientific literature. In ResFinder, publication is not a strict requirement. Genes must have a GenBank number and expert review of the GenBank entries. Also, the types of entries differ across databases, with most databases including AMR genes and only a few databases including mutations of chromosomal genes mediating AMR. Finally, the available AMR databases differ regarding the format of the entries (fasta, json, etc.), the possibility of download, and the availability and frequency of curation (Table 2).

At present, at least 47 online available resources for in silico AMR prediction are published in the scientific literature (13, 26, 28–63) (Table 2). They range from basic AMR reference databases that can be embedded in the user's own bioinformatics pipeline, to systems having a well-curated database with integrated search tools. These bioinformatics resources have interfaces of different complexity that require different skills in bioinformatics and microbiology for performing the sequence analyses and interpreting the results (Table 2). As the features of these systems differ widely, the outputs obtained by different tools may not be fully comparable. Moreover, employing the same tool for different input formats of the same data (e.g., raw reads vs. assembled sequences, trimmed vs. non-trimmed reads; assemblies obtained by different software, etc.) can produce different results (64). A reliable genomic approach to assaying AMR gene content requires accurate curated reference databases.
### TABLE 2 | Open-access resources for in silico antimicrobial resistance detection in bacteria.

<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Software</th>
<th>Database</th>
<th>Input sequence</th>
<th>Link</th>
<th>Year of development</th>
<th>Curation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABRES Finder</td>
<td>General AMR</td>
<td>Profile HMM</td>
<td>No</td>
<td>Own</td>
<td><a href="http://scbt.sastra.edu/ABRES/index.php">http://scbt.sastra.edu/ABRES/index.php</a></td>
<td>2017</td>
<td>Not specified</td>
<td>Unpublished</td>
</tr>
<tr>
<td>ARDB</td>
<td>General AMR</td>
<td>BLAST</td>
<td>Yes</td>
<td>Own</td>
<td><a href="https://ardb.cbcb.umd.edu/">https://ardb.cbcb.umd.edu/</a></td>
<td>2009</td>
<td>2009 (28)</td>
<td></td>
</tr>
<tr>
<td>ARG-ANNOT</td>
<td>General AMR</td>
<td>–</td>
<td>–</td>
<td>Own</td>
<td>Discontinued</td>
<td>2014</td>
<td>2018 (29)</td>
<td></td>
</tr>
<tr>
<td>CARD</td>
<td>General AMR</td>
<td>BLAST, RGI</td>
<td>Yes</td>
<td>Own</td>
<td><a href="https://card.mcmaster.ca/home">https://card.mcmaster.ca/home</a></td>
<td>2013</td>
<td>2019 (31)</td>
<td></td>
</tr>
</tbody>
</table>

*(Continued)*
TABLE 2 | Continued

<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Software</th>
<th>Database</th>
<th>Input sequence</th>
<th>Link</th>
<th>Year of development (last update)</th>
<th>Curation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCBI AMR Finder</td>
<td>General AMR</td>
<td>BLAST, HMMER</td>
<td>Yes</td>
<td>Own</td>
<td>Nucleotide, amino acid FASTA, GFF</td>
<td>2017 2019</td>
<td>(33)</td>
<td></td>
</tr>
<tr>
<td>Noradab</td>
<td>General AMR</td>
<td>BLAST</td>
<td>No</td>
<td>Derived from ARDB and CARD³</td>
<td>Nucleotide, amino acid FASTA</td>
<td>2018 Not specified</td>
<td>(34)</td>
<td></td>
</tr>
<tr>
<td>Patric</td>
<td>General AMR</td>
<td>BLAST</td>
<td>Yes</td>
<td>Own</td>
<td>Nucleotide, amino acid FASTA, FASTQ</td>
<td>2004 2019</td>
<td>(35)</td>
<td></td>
</tr>
<tr>
<td>ResFinder</td>
<td>General AMR</td>
<td>BLAST, KMA</td>
<td>Yes</td>
<td>Own</td>
<td>Nucleotide, amino acid FASTA, FASTQ</td>
<td>2012 2019</td>
<td>(36)</td>
<td></td>
</tr>
<tr>
<td>SRST2</td>
<td>General AMR</td>
<td>BOWTIE2</td>
<td>Yes</td>
<td>Derived from ARG-ANNOT³</td>
<td>Nucleotide, amino acid FASTA, FASTQ and any other format readable by BOWTIE2</td>
<td>2014 2019</td>
<td>(37)</td>
<td></td>
</tr>
<tr>
<td>SSTAR</td>
<td>General AMR</td>
<td>BLAST</td>
<td>Yes</td>
<td>Derived from ARG-ANNOT and Resfinder³</td>
<td>Nucleotide, amino acid FASTA</td>
<td>2015 2018</td>
<td>(38)</td>
<td></td>
</tr>
<tr>
<td>INTEGRALL</td>
<td>AMR genes and associated integrons</td>
<td>BLAST</td>
<td>No</td>
<td>Own</td>
<td>Nucleotide, amino acid FASTA</td>
<td>2008 2019</td>
<td>(39)</td>
<td></td>
</tr>
<tr>
<td>MvirDB</td>
<td>AMR genes, protein toxins and virulence factors for bio-defense applications</td>
<td>BLAST</td>
<td>No</td>
<td>Derived from Tox-Prot, SCOPION, the PRINTS virulence factors, VFDB, TVFac, Islander, ARGO and a subset of VIDA³</td>
<td>Nucleotide, amino acid FASTA</td>
<td>Discontinued (<a href="http://mvirdb.llnl.gov/">http://mvirdb.llnl.gov/</a>)</td>
<td>2007 Not specified</td>
<td>(40)</td>
</tr>
<tr>
<td>BacMet</td>
<td>Biocide and metal resistance</td>
<td>BLAST</td>
<td>No</td>
<td>Own</td>
<td>Nucleotide, amino acid FASTA</td>
<td>2013 2018</td>
<td>(41)</td>
<td></td>
</tr>
<tr>
<td>ResCap</td>
<td>Antibiotic, heavy metal and biocide resistance</td>
<td>BLAST, Bowtie2</td>
<td>Yes</td>
<td>Derived from ARG-ANNOT, CARD, RED-DB, ResFinder, Bacmet³</td>
<td>Nucleotide, amino acid FASTA, FASTQ</td>
<td>2017 2017</td>
<td>(42)</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
### TABLE 2 | Continued

<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Software</th>
<th>Database</th>
<th>Input sequence</th>
<th>Link</th>
<th>Year of development</th>
<th>Curation (last update)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetracycline MLS nomenclature</td>
<td>β-lactamases Database</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td><a href="https://faculty.washington.edu/marilynn/Not">https://faculty.washington.edu/marilynn/Not</a> specified</td>
<td>2019</td>
<td>Unspecified</td>
<td>Unpublished</td>
</tr>
<tr>
<td>BLAD</td>
<td>β-lactamases</td>
<td>–</td>
<td>–</td>
<td>Own</td>
<td><a href="http://www.blad.co.in/">http://www.blad.co.in/</a></td>
<td>2012</td>
<td>Not specified</td>
<td>Unpublished</td>
</tr>
<tr>
<td>BLDB</td>
<td>β-lactamases</td>
<td>BLAST</td>
<td>No</td>
<td>Own</td>
<td><a href="http://blbd.eu/">http://blbd.eu/</a></td>
<td>2017</td>
<td>2019</td>
<td>(44)</td>
</tr>
<tr>
<td>CBMAR</td>
<td>β-lactamases</td>
<td>BLAST</td>
<td>No</td>
<td>Own</td>
<td><a href="http://proteininformatics.org/mkumar/lactamasesdb/">http://proteininformatics.org/mkumar/lactamasesdb/</a></td>
<td>2014</td>
<td>2014</td>
<td>(45)</td>
</tr>
<tr>
<td>LacED</td>
<td>β-lactamases</td>
<td>BLAST</td>
<td>No</td>
<td>Own</td>
<td><a href="http://www.laced.uni-stuttgart.de/Not">http://www.laced.uni-stuttgart.de/Not</a> specified</td>
<td>2009</td>
<td>Not specified</td>
<td>(46)</td>
</tr>
<tr>
<td>AMR-time</td>
<td>AMR genes in metagenomic data</td>
<td>DIAMOND</td>
<td>Yes</td>
<td>CARD</td>
<td><a href="https://github.com/beiko-lab/AMR-time">https://github.com/beiko-lab/AMR-time</a></td>
<td>2017</td>
<td>2019</td>
<td>(47)</td>
</tr>
<tr>
<td>DeepARG</td>
<td>AMR genes in metagenomic data</td>
<td>BLAST, DIAMOND</td>
<td>Yes</td>
<td>Derived from RDB, CARD, UNIPROTb</td>
<td><a href="https://bench.cs.vt.edu/deeparg">https://bench.cs.vt.edu/deeparg</a></td>
<td>2017</td>
<td>2019</td>
<td>(48)</td>
</tr>
<tr>
<td>GROOT</td>
<td>AMR genes in metagenomic data</td>
<td>LSH Forest indexing</td>
<td>Yes</td>
<td>Derived from ARDB, CARD, Resfinder</td>
<td><a href="https://github.com/will-rowe/groot">https://github.com/will-rowe/groot</a></td>
<td>2018</td>
<td>2019</td>
<td>(49)</td>
</tr>
<tr>
<td>SARG (ARGs-OAP, ARQpore)</td>
<td>AMR genes in metagenomic data</td>
<td>BLAST, HMMER, UBLAST</td>
<td>Yes</td>
<td>Derived from ARQ-ANNOT, CARD, Resfinder</td>
<td><a href="https://smile.hku.hk/SARGs">https://smile.hku.hk/SARGs</a></td>
<td>2016</td>
<td>2019</td>
<td>(50)</td>
</tr>
<tr>
<td>SEAR</td>
<td>AMR genes in metagenomic data</td>
<td>BLAST, BWA-MEM</td>
<td>Yes</td>
<td>ARG-ANNOT</td>
<td>Discontinued (<a href="https://github.com/will-rowe/SEAR">https://github.com/will-rowe/SEAR</a>)</td>
<td>2015</td>
<td>2018</td>
<td>(51)</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Software</th>
<th>Database</th>
<th>Input sequence</th>
<th>Link</th>
<th>Year of development (last update)</th>
<th>Curation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ShortBRED</td>
<td>AMR genes in metagenomic data</td>
<td>BLAST, USEARCH</td>
<td>Derived from ARDB and CARD&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Yes</td>
<td><a href="http://huttenhower.sph.harvard.edu/shortbred">http://huttenhower.sph.harvard.edu/shortbred</a></td>
<td>2015 2019</td>
<td></td>
<td>(52)</td>
</tr>
<tr>
<td>Mustard</td>
<td>AMR determinants in the human gut microbiota</td>
<td>BLAST</td>
<td>Derived from Resfinder, ARG-ANNOT, the Lahey Clinic (<a href="http://www.lahey.org/studies/">http://www.lahey.org/studies/</a>), RED-DB (<a href="http://www.fibim.unisi.it/REDDB/">http://www.fibim.unisi.it/REDDB/</a>), Marilyn Roberts’ website for macrolides and tetracycline resistance (<a href="http://faculty.washington.edu/marilyn/">http://faculty.washington.edu/marilyn/</a>) and different functional metagenomics studies&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Yes</td>
<td><a href="http://mgps.eu/Mustard/">http://mgps.eu/Mustard/</a></td>
<td>2017 2017</td>
<td></td>
<td>(53)</td>
</tr>
<tr>
<td>FARMEDB</td>
<td>AMR genes discovered by functional metagenomics</td>
<td>BLAST</td>
<td>Own</td>
<td>Yes</td>
<td><a href="http://staff.washington.edu/wallace/farme/index.html">http://staff.washington.edu/wallace/farme/index.html</a></td>
<td>2016 Not specified&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>Unpublished</td>
</tr>
<tr>
<td>ResFinderFG</td>
<td>AMR genes discovered by functional metagenomics</td>
<td>BLAST</td>
<td>Own</td>
<td>No</td>
<td><a href="https://cge.cbs.dtu.dk/services/ResFinderFG-1.0/">https://cge.cbs.dtu.dk/services/ResFinderFG-1.0/</a></td>
<td>2016 Not specified&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>Unpublished</td>
</tr>
<tr>
<td>Galileo AMR (MARA, RAC)</td>
<td>AMR genes in Gram-negative bacteria (ATTACCA)</td>
<td>BLAST</td>
<td>Own</td>
<td>Yes</td>
<td><a href="https://galileoamr.arcbio.com/mara/">https://galileoamr.arcbio.com/mara/</a></td>
<td>2017 Not specified&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
<td>(55)</td>
</tr>
<tr>
<td>LREfinder</td>
<td>Linezolid resistance in enterococci</td>
<td>KMA</td>
<td>Own</td>
<td>Yes</td>
<td><a href="https://cge.cbs.dtu.dk/services/LRE-finder/">https://cge.cbs.dtu.dk/services/LRE-finder/</a></td>
<td>2019 2019</td>
<td></td>
<td>(56)</td>
</tr>
</tbody>
</table>

<sup>1</sup> Additional references related to ShortBRED are provided. For Mustard, the references include specific studies and resources. For FARMEDB, the development year is specified as 2016, with the last update not specified. For ResFams, the development year is 2014, and the last update is 2018. For Galileo AMR, the development year is 2017, and the last update is not specified. For LREfinder, the development year is 2019, and the last update is 2019. Further details and additional references are provided in the specified links and resources.
<table>
<thead>
<tr>
<th>Name</th>
<th>Target</th>
<th>Software</th>
<th>Database</th>
<th>Input sequence</th>
<th>Link</th>
<th>Year of development (last update)</th>
<th>Curation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUBII-TB-DB</td>
<td>AMR mutations in Mycobacterium tuberculosis</td>
<td>BLAST</td>
<td>Own</td>
<td>No</td>
<td>Downloadable&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2013</td>
<td>Not specified</td>
<td>(57)</td>
</tr>
<tr>
<td>Mykrobe</td>
<td>AMR in Mycobacterium</td>
<td>Own</td>
<td>Yes</td>
<td>Nucleotide</td>
<td>FASTQ</td>
<td>2015</td>
<td>2019</td>
<td>(58)</td>
</tr>
<tr>
<td>TBDreamdb</td>
<td>AMR in Mycobacterium – tuberculosis</td>
<td>Own</td>
<td>Yes</td>
<td>–</td>
<td>–</td>
<td>2009</td>
<td>2014</td>
<td>(59)</td>
</tr>
<tr>
<td>PointFinder</td>
<td>Selected mutations in chromosomal genes of Escherichia coli, Salmonella sp., Campylobacter sp., Staphylococcus aureus, Enterococcus sp., Mycobacterium tuberculosis, Neisseria gonorrhoeae</td>
<td>BLAST, KMA</td>
<td>Own</td>
<td>Yes</td>
<td>Downloadable&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2017</td>
<td>2019</td>
<td>(13)</td>
</tr>
<tr>
<td>SCCmec Finder</td>
<td>SCCmec elements in Staphylococcus aureus</td>
<td>BLAST, KMA</td>
<td>Own</td>
<td>Yes</td>
<td>Nucleotide, FASTQ</td>
<td>2016</td>
<td>2018</td>
<td>(60)</td>
</tr>
<tr>
<td>U-CARE</td>
<td>AMR in Escherichia coli</td>
<td>BLAST</td>
<td>Own</td>
<td>Yes</td>
<td>Amino acid, FASTA</td>
<td>2013</td>
<td>Not specified</td>
<td>(61)</td>
</tr>
<tr>
<td>ARG-miner</td>
<td>Robust and comprehensive curation of AMR gene databases</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>–</td>
<td>2018</td>
<td>2019 (crowdcuration)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Yes, standalone version is available (usually in Bitbucket or in GitHub) either with or without a corresponding web version; no, only web version is available.

<sup>b</sup>Curation to avoid redundancies and remove selected sequences (see respective references for details).

<sup>c</sup>Active, based on authors’ knowledge; discontinued databases may still be available for download via WayBack Machine.
that should be synchronized and harmonized in a way to ensure comparable outputs worldwide. Once that is achieved, the bioinformatics method of monitoring will undeniably lead to a paradigm shift in the way that we conduct AMR surveillance and compare results internationally. Importantly, the currently available tools may detect new gene variants, but they are not presently equipped to detect new AMR genes. Identifying novel resistance elements from genomic data is being pursued using iterative kmer-based analytics and other machine learning schemes but these strategies still require well-characterized reference genomes with phenotypic data for training (11, 19–21).

**BENCHMARKING OF BIOINFORMATICS TOOLS TO DETECT ANTIMICROBIAL RESISTANCE DETERMINANTS**

Benchmarking exercises are important to assess the performance, and reliability of the available bioinformatics tools which have different complexity in design and function.

Designing and executing a benchmarking trial offers several challenges. At a recent meeting (October 2017) organized by the European Commission Joint Research Center, the challenges of designing a benchmarking strategy for assessing bioinformatics tools to detect AMR determinants was discussed (65). Here, several challenges were identified, and considerations discussed which included: (1) the origin of the dataset tested; (2) sustainable reference datasets; (3) quality of the test genomes; (4) what determinants to include in a dataset; (5) the, expected result; and (6) performance thresholds. The sequence dataset could either be real or artificially composed. In both cases, this will have implications for accurate benchmarking. A real dataset needs to be properly characterized and the true reference result defined. Furthermore, a real dataset may be biased in content for certain resistance determinants, such as mutations in the ampC promoter of E. coli, and thereby affect some bioinformatics tools more than others (26). In contrast, a simulated dataset needs to be accurate and correct but also contain a variety of different determinants or mechanisms. Ideally, a combination could be applied designing a desired benchmarking dataset to represent real-life scenarios aligned with the test objective (e.g., only focused on extended spectrum β-lactamases). The scope of bacterial species represented can also influence the results (65).

The quality and type of sequence data are also important factors. This also needs to mimic a real-life scenario where genomes will differ in error rates, read lengths, and read quality and may be raw reads or assemblies. The robustness of bioinformatics tools will differ in performance when dealing with low quality genomes and assemblies compared to optimal conditions (26, 65).

Prior to executing a benchmarking exercise, the reference AMR classes need to be determined as to whether all known or acquired determinants will be included, or only specific mechanisms such as certain enzymes, efflux pumps, mutations/single nucleotide polymorphisms (SNPs), upregulated or downregulated genes or porins. Ideally, the bioinformatics tools should enable the detection of all known determinants if used for surveillance or guiding clinical treatment unless the scope is different and agreed upon (65).

Since the main objective of a benchmarking exercise is to assess the ability of the bioinformatics tool to provide reliable analysis of AMR gene content, it is vital that the concordance is high between the reference result and the expected outcome (65). The sensitivity is especially important as the misidentification of a resistant strain is more consequential than the finding of silent resistance genes in phenotypically susceptible isolates. As previously mentioned, discrepancies observed between phenotypic reference result and the expected genomic outcome is often due to incorrect phenotypic antimicrobial susceptibility test data.

Assessing the performance of bioinformatics tools is often based on a comparison between the genotypic and phenotypic results and a calculation of the specificity, sensitivity, positive predictive (PPV) and negative predictive values (NPV), accuracy [Simple Matching Coefficient (SMC)] and performance [Matthew's Correlation Coefficient (MCC)] followed by a comparison of these parameter’s between the different bioinformatics tools (26, 66).

Surprisingly, only a few studies have benchmarked bioinformatics tools against each other to detect AMR determinants. 24 used two previously published pair-end Miseq datasets (7, 8) of 196 genomes of four species and 143 genomes from two species (five species in total), respectively. Phenotypic susceptibility test data was used as the reference result in predicting AMR determinants when benchmarking the KmerResistance vers 1.0 (target only enzymes) (70% identity and 10% depth corr (co-occurrence of K-mers), ResFinder vers. 2.0 (target only enzymes) [98% identity and 60 coverage (assembly/BLAST)], and SRST2 (90% identity 90% coverage) (clustering/Bowtie2). To further challenge the sensitivity, the datasets were down-sampled to 1% of the reads and re-analyzed. Overall, the three bioinformatics tools performed equally well with almost the same accuracy, SMC and performance, MCC testing the two datasets; SMC and MCC were app. 96% and 0.90 for the Stoe tess et al. collection, respectively whereas the SMC and MCC ranged from 98 to 100% and 0.91 to 0.99 for the Zankari et al. collection, respectively with the lowest performance by SRST2 and the highest by KmerResistance (26). The KmerResistance tool performed significant better than the two others when data were contaminated or down-sampled to contain a few reads—all bioinformatics tools performed best using raw reads input data (26).

Another study (ENGAGE) (66) evaluated the Public Health England’s GeneFinder tool, which targets enzymes and some chromosomal point mutations for fluoroquinolone resistance using two HiSeq datasets, 125 Salmonella genomes and 164 E. coli genomes of which a large proportion harbored upregulated ampC-mediated resistance to extended spectrum cephalosporins. ResFinder provided the highest accuracy, SMC and performance, MCC predicting resistance in the E. coli genomes and GeneFinder for Salmonella genomes. The correlation to phenotypic susceptibility testing was for Salmonella spp. Ninety percent for all bioinformatics tools but higher for GeneFinder specifically for fluoroquinolones. The
accuracy, SMC revealed to be lower in *E. coli* than testing *Salmonella* for all bioinformatics tools due to the bias of the *E. coli* dataset containing a high number of upregulated ampC genotypes not predicted by any of the bioinformatics tools (66). Hunt et al. similarly benchmarked the same bioinformatics tools as in Clausen et al. including also the ARIBA tool (30). The ARIBA tool contain in addition to enzymes also chromosomal point mutations thus, outperforming both KmerResistance (26) and SRST2 (37).

Following the benchmarking described above, both the ResFinder and the KmerResistance bioinformatics tools have been updated. Thus, the ResFinder tool now includes a number of chromosomal point mutations such as those to detect resistance to colistin, fluoroquinolones, etc. Overall, the benchmarking exercises revealed that all bioinformatics tools evaluated performed almost similarly good but were affected by the type and quality of input data.

In an assessment of the accuracy of NCBI’s AMRFinder, a 2018 study by Feldgarden et al compared it with a 2017 version of ResFinder (33). AMRFinder was evaluated first using a set of 6,242 genomes with 87,679 AST data points for 14 antimicrobial drugs. Overall, 98.4% were consistent with predictions. When compared with ResFinder, most gene calls were identical. While there were 1,229 gene symbol differences, 81% were attributed to differences in database composition. AMRFinder and ResFinder use HMM- and BLAST-based approaches, respectively, and are the commonly used resources for genome-based AMR tracking. Synchronized harmonization of the databases, as is done globally with genomic sequence databases, is needed to minimize inconsistent outputs due to algorithmic differences.

ENSURING HIGH QUALITY GENOMIC DATA BY PROFICIENCY TESTING

Standardization of WGS procedures from DNA preparation to the final genome is paramount to ensure reliable prediction of AMR determinants for surveillance and clinical purposes. To ensure the production of reliable high quality genomic data, laboratories routinely performing WGS should participate in laboratory proficiency testing (PT) or external quality assurance systems (EQAS) (67, 68). For decades, global and regional EQAS in phenotypic AST of foodborne pathogens has been conducted to ensure the quality of performed dilution and diffusion AST (69–71). There is an urgent need to also establish a mechanism to provide a global proficiency testing in the area of WGS to establish standardization in the field (68). This goal is part of the charter of the Global Microbial Identifier (GMI), launched in 2011, to help establish a “global system of DNA genome databases for microbial and infectious disease identification and diagnostics” (https://www.globalmicrobialidentifier.org/).

In 2014, GMI launched its first pilot PT in WGS lead by the DTU and US FDA to trial test the WGS platforms, procedures, test material and the functionality of the assessment pipeline (72). In 2015, a full roll-out of the pilot was delivered by GMI to a global audience. The GMI continued to provide proficiency testing in 2016 and 2017. Cultures and pure DNA for library construction were provided to participating laboratories for DNA purification, library preparation, and WGS followed by *in silico* prediction of wgMLST and AMR determinants. The genomes and analysis were submitted to DTU for quality control assessment using closed genomes of the test strains as a reference. The quality control assessment was facilitated by an in-house developed PT QC pipeline measuring a large number of parameters. These included the numbers of reads after trimming, unmapped reads, map to the total reference DNA, reference chromosome, reference plasmids; proportion of reads that map to reference chromosome; coverage of the reference chromosome and reference plasmids; depth of coverage of total DNA, reference chromosome, and reference plasmids; Phred quality score (Q score), total size and proportion of assembly map to the reference DNA, number of contigs including above a length above 200 bp, N50, and NG50. Underperformance was observed and reported in each trial mainly caused by laboratory contamination or poor performance.

DATA SHARING—PUBLIC/PRIVATE

An important element of genomics as a tool for AMR surveillance and diagnostics is that, once data quality standards are met, the data set is platform-independent, discrete and portable. The analytical outputs and data sharing then become the most important considerations (Figure 2). A plethora of international and governmental position papers have stressed the need for global cooperation and data sharing to combat infectious diseases and worsening antimicrobial resistance (73–82). Countries have different levels of legal restriction on the sharing of medical information and biological material with potential commercial value or compliance to the EU General Data Protection Regulation. While the legal issue may be more intractable, the public health advantages to global data sharing are obvious. In the US, where fewer restrictions are in place, WGS data from national surveillance systems are continuously placed in the public domain both for public health purposes, and for exploitation by innovators to develop and update new technologies. This permits global access to information on common microbiological threats, something that will become more important as travel and trade increase and as new threats arise.

ONLINE REPOSITORIES TO HOST AND LINK GENOME AND ANTIMICROBIAL SUSCEPTIBILITY DATA

Concurrently with the vast amount of genomic data being produced, traditional antimicrobial susceptibility testing is being conducted in parallel on a large scale. Up until recently, it was only possible to submit and store DNA sequence data in the International Nucleotide Sequence Database Collaboration (INSDC), whereas all AST data was stored separately in closed local or national repositories. Furthermore, not all genomic data is submitted to the online open genomic repositories of INSDC and shared globally due to difficulties to submit, a lack of appreciation for its value, access to local or national repositories,
fear of being data being published by others, or privacy of the data (83). Nonetheless, today the NCBI and the European Bioinformatics Institute (EMBL-EBI) can accommodate AST data along with the WGS information, to facilitate a global monitoring of AMR in bacteria to strengthen global public health (84, 85).

**EUROPEAN NUCLEOTIDE ARCHIVE REPOSITORY**

At European Nucleotide Archive (ENA), a mechanism to host and link submitted genomic and AST data has been developed by the EU COMPARE partners and EMBL-EBI (85). Briefly, the EMBL-EBI system allows submitted genomes and associated metadata in the ENA to be stored as open access or privately in a secured login protected repository with named data hubs (86). The system is designed to accommodate submission of susceptibility data from both dilution or diffusion methods. Novel software has been developed to validate conformity of the AST data to ensure harmonization of the data (85). The submitted genomic and AST data could be analyzed by using existing bioinformatics infrastructure and implemented cloud-based bioinformatics workflows in specific an extended version of the Bacterial Analysis Pipeline consisting of ContigAnalyzer-1.0, KmerFinder-2.1, MLST-1.6, ResFinder-2.1, VirulenceFinder-1.2, PlasmidFinder-1.2, pMLST-1.4 (87) with the inclusion of also the cgMLSTFinder 1.0. The submitted data could be queried and downloaded in multiple ways including via the Pathogen Data Portal for surveillance, identification, and investigation https://www.ebi.ac.uk/ena/pathogens/home. Subsequently, the data could be visualized by using a developed Notebook tool integrated the Pathogen Data Portal to query and display all typing data including distribution of the phenotypic AST data enable a potential real time monitoring of AMR (85). The advantage of the data hub model and similar embassy cloud system is the possibility for privacy to control own data having restricted access to only owners or collaborators while analyzing or publishing the data or await less political sensitivity due to GDPR which all a major barriers in data sharing (88–90).

**NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION REPOSITORY**

The National Center for Biotechnology Information (NCBI) is the US member of the INSDC and part of the United States National Institutes of Health, and houses hundreds of thousands of bacterial genomes from around the world. Sequences are submitted from global research studies, but the majority are from national public health surveillance programs with systematic sampling schema. With the expansion of WGS capacity, the number of genome submission is expected to rise soon to over 100,000 annually from US sources alone.

To help make these large datasets accessible, the NCBI Pathogens page (https://www.ncbi.nlm.nih.gov/pathogens/) was developed. This resource is designed for exploring the genomic features of various bacterial pathogens. These include major foodborne and zoonotic pathogens, such as *Salmonella enterica*, *Escherichia coli*, and *Campylobacter* spp. Included in these datasets is a variety of metadata, including strain ID, source, date collected, geographical location, antimicrobial resistance, and more. This page was established in collaboration with GenomeTrakr, an international consortium of laboratories organized by the U.S. Food and Drug Administration (FDA) that
collect and sequence bacterial strains from a variety of food and environmental sources (91).

A major feature of the Pathogens page is the phylogenetic trees, as genomes are arranged into clusters based on relatedness according to SNPs. These allow users to explore and interpret the relatedness of bacterial strains. These have provided a robust database of bacterial species that can be used for genomic comparisons with isolates collected from human patients. This information can be used to help identify foodborne disease outbreaks and support regulatory actions by the FDA.

Another major aspect of the Pathogens page is the AMR reference gene database mentioned above (33). The tool, AMRFinder is automatically run on all genomes submitted to NCBI, resulting in AMR genotype outputs that identify resistance genes from each sequence (33). This, combined with the phylogenetic tree outputs, allows for identification and potential prioritization of investigations into resistant outbreaks of pathogenic organisms.

The NCBI Pathogens web portal also contains phenotypic information, when submitters of these data choose to include it. Over 7,000 isolates now have phenotypic MIC data associated with them, allowing users to interrogate the data for various resistance phenotypes, including those conferred by mutations not tracked presently by the genotypic outputs of AMRFinder (33).

To help make the resistance information accessible, the US Food and Drug Administration developed a tool called ResistomeTracker (https://www.fda.gov/animal-veterinary/national-antimicrobial-resistance-monitoring-system/global-salmonella-resistome-data). This suite of data dashboards is focused exclusively on analysis and visualization of AMR genes extracted from the complete genomes at the NCBI. ResistomeTracker was developed for the U.S. National Antimicrobial Resistance Monitoring System (NARMS) to better understand the epidemiological aspects of resistance by making the large amounts of resistome data accessible to a broad user audience. This includes the identification of new resistance determinants, differences in the prevalence of resistance genes among various food commodities, and geographical spread over time. Additionally, continuous updates to ResistomeTracker enable users to detect early resistance threats. ResistomeTracker allows for user-directed queries of the data that are informative for individual interests. Because it is linked directly to the NCBI pathogen database, it allows the user to begin a query with a specific resistance allele, and end with a phylogenetic analysis of related strains. It currently is focused on foodborne bacteria, but can be modified to exploit and genome for resistance gene content.

**USING WGS IN AMR SURVEILLANCE**

In the United States, national laboratory capacity for AMR monitoring and WGS is growing. It consists of federally coordinated networks operated by State public health laboratories and Universities. The Centers for Disease Control and Prevention (CDC) coordinates the Antibiotic Resistance Laboratory Network (ARLN) to rapidly detect emerging resistance threats in healthcare, food and the community. Among many activities, this comprehensive network performs WGS for numerous pathogens, including all isolates of *Mycobacterium tuberculosis*. WGS is used also as a routine method to characterize *Neisseria gonorrhoeae*, and other major pathogens, including those involved in outbreaks.

The National Antimicrobial Resistance Monitoring System (NARMS) is a long-standing program focused on bacteria transmitted commonly through food (92). NARMS is a partnership of the CDC, the FDA and United States Department of Agriculture Food Safety and Inspection Service (FSIS); it is focused on tracking resistance in enteric bacteria from humans, retail meats and food animals, respectively. NARMS began systematic WGS of *Salmonella* in 2013 and has incorporated WGS data for *Salmonella* and *Campylobacter* in its reports since 2014. Online tools enable users to examine resistance trends at the genetic level using various query filters. These tools provide graphical visualizations of the genotypes behind changing resistance patterns over time by source and serotype.

As national resistance surveillance matures to better fit the One Health model, animal pathogens and environmental testing are beginning. In the US, the Department of Agriculture National Animal Health Laboratory Network (NAHLN) and the FDA Veterinary Laboratory Investigation and Response Network (Vet-LIRN) are starting to gather resistance information and WGS data on pathogens from food animals and companion animals, respectively. The US Environmental Protection Agency (EPA) conducts periodic water surveys that includes detection of resistance genes. While in the early stages, national public health surveillance programs using DNA sequence information will continue to expand and permit new associations to be inferred from resistomic analyses of the data.

In Europe, its mandatory by law, Directive 2003/99/EC (https://eur-lex.europa.eu/eli/dir/2003/99/oj) for Member States (MSs) to monitor AMR phenotypically by MIC determination in *Salmonella*, *Campylobacter*, and *E. coli* obtained from healthy food-producing animals and from food. The monitoring also include a specific monitoring of extended-spectrum beta-lactamase (ESBL)-, AmpC- and carbapenemase-producing *Salmonella* and indicator commensal *E. coli* stipulated in the Commission Implementing Decision 2013/652/EU of 12 November 2013 (http://data.europa.eu/eli/dec_impl/2013/652/oj). The data collection on human diseases including AMR from MSs is optimal and based on either MIC or disk diffusion and conducted in accordance with Decision 1082/2013/EU (http://data.europa.eu/eli/dec/2013/1082/oj).

A number of MSs providing data for the specific monitoring of AmpC- and carbapenemase-producing *Salmonella* and indicator commensal *E. coli* from healthy food-producing animals and from food, has expressed an interest to replace the mandatory phenotypic MIC determination with WGS due to this already been implemented locally in the specific MSs. Thus, in the preparatory work of updating the Commission Implementing Decision 2013/652/EU coming into force in 2021, the preliminary draft of the technical specifications on
harmonized monitoring of resistance in zoonotic and indicator bacteria from food-producing animals and food from EFSA suggested to allow replacing MIC determination with WGS combined with using the CGE ResFinder tool till 2025 (36). From 2025, the using of WGS combined with using the CGE ResFinder tool will be mandatory for the specific monitoring of AmpC- and carbapenemase-producing *Salmonella* and indicator commensal *E. coli* from healthy food-producing animals and from food and considered to be expended replacing all phenotypic MIC determinations as well as species identification. The resulting AMR determinant profile will be submitted to EFSA and used to predict the phenotype which will be reported in the European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food. It will be optional for the individual MSs to also submit the DNA sequences and metadata data to ENA. It’s believed that all MSs by 2015 have acquired WGS and conducting bioinformatics analysis of DNA sequences of single isolates for monitoring purposes.

**AMR SURVEILLANCE USING METAGENOMICS**

Current AMR surveillance often focuses on few pathogens mainly based on passive reporting of phenotypic laboratory results for a few selected specific pathogens as in the Danish monitoring system, DANMAP https://www.danmap.org/, leading to a narrow pathogen spectrum that does not capture all relevant AMR genes. The majority of AMR genes may be present in the commensal bacterial flora of healthy humans and animals or the environment.

Metagenomics techniques, using short-read next-generation sequencing data, benefit from the ability to quantify thousands of especially transmissible resistance genes in a single sample without any prior selection of which genes to look for. Moreover, it can provide additional information about the presence of bacterial species, pathogens and virulence genes and the data can be re-analyzed, if novel genes of interest are identified.

It was recently shown that metagenomics is superior to conventional methods for AMR surveillance in pig herds (93), useful for comparing AMR across livestock in Europe (94), as well as investigations related to epidemiological data (95). The utility for surveillance of global AMR gene dissemination through international flights (96) and using urban sewage to determine the local and global resistome has also been proven (97, 98).

Metagenomics will sequence all DNA present in the sample including food and host DNA, which may result in low sensitivity. Quantitative PCR procedures, including large scale capture PCR methodologies have been developed, likely providing higher sensitivity (42). However, these methodologies have not been compared with respect to sensitivity and specificity.

In the future the application of metagenomics directly on samples from healthy and clinical ill individuals and animals as well as potential reservoir might results in the ultimate One Health surveillance of AMR allowing determination of all resistance genes and their context in all reservoirs. However, as for single isolates different pipelines and databases are also used for such metagenomics studies and there is a need for global standardization.

**PERSPECTIVES**

An important advantage of using WGS technologies in detecting and tracking AMR is the opportunity to expand it to align with a One Health surveillance framework and allowing for exact comparisons across reservoirs. This cannot be done using WGS only on the phenotypic antimicrobial class level, but at the exact genetic mechanism level. This One Health goal has so far been impeded by the high cost of testing animal and environmental samples using classical methods based on metabolic and biochemical characterization. As the NGS technology becomes more affordable, it will become more common to use metagenomics to explore the potential role of different environments in the ecology of resistance. Thus, One Health monitoring is now poised to evolve into nucleotide surveillance of complex microbial ecosystems. And to the extent that the data can be generated and reported without delay, it appears that something analogous to a “weather map” of infectious diseases and resistance is possible. This was not practicable in the past, where *ad hoc* gene detection was the norm and PFGE was the typing tool of choice.

**CONCLUSION**

The advancement in whole genome sequencing and the application of online tools for real-time detection of AMR determinants is essential for control and prevention strategies to combat the increasing threat of AMR. We identified a number of accessible tools available in the prediction of AMR determinants to support expanding to establish global pathogen surveillance and AMR tracking based on genomics. In addition, we identified a number of preceding requirements for a successful transition such as curated AMR databases ensuring a high concordance between pheno- and genotypes, benchmarking designs, PT schemes, sharing options etc. There is however, a vital need for standardization of pipelines and databases as well as phenotypic predictions based on the genomic data.

**AUTHOR CONTRIBUTIONS**

RH and PM conceived, outlined and critically revised the manuscript. All authors wrote, read and accepted the manuscript.

**FUNDING**

This study has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No. 643476 (COMPARE) and from the Novo Nordisk Foundation (NNF16OC0021856: Global Surveillance of Antimicrobial Resistance).
REFERENCES


77. Sane JEM. Overcoming Barriers to Data Sharing in Public Health A Global Perspective. (2015). London, UK: Chatham House, the Royal Institute of International Affairs.


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hendriksen, Bortolaia, Tate, Tyson, Aarestrup and McDermott. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.