



Using Genomics to Track Global Antimicrobial Resistance

Rene S. Hendriksen^{1*}, Valeria Bortolaia¹, Heather Tate², Gregory H. Tyson², Frank M. Aarestrup¹ and Patrick F. McDermott²

¹ European Union Reference Laboratory for Antimicrobial Resistance, World Health Organisation, Collaborating Center for Antimicrobial Resistance and Genomics in Food borne 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

OPEN ACCESS

Edited by:

Marc Jean Struelens,
European Centre for Disease
Prevention and Control
(ECDC), Sweden

Reviewed by:

Sergey Eremin,
World Health Organization
(Switzerland), Switzerland
Ana Afonso,
University of São Paulo, Brazil

*Correspondence:

Rene S. Hendriksen
rshe@food.dtu.dk

Specialty section:

This article was submitted to
Infectious Diseases - Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Public Health

Received: 18 June 2019

Accepted: 13 August 2019

Published: 04 September 2019

Citation:

Hendriksen RS, Bortolaia V, Tate H,
Tyson GH, Aarestrup FM and
McDermott PF (2019) Using
Genomics to Track Global
Antimicrobial Resistance.
Front. Public Health 7:242.
doi: 10.3389/fpubh.2019.00242

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 1 | Concordance between phenotypic susceptibility testing and WGS based predicted antimicrobial resistance.

	Pathogen	No. of pathogens	AST method	No. of antimicrobials	Bioinformatic tool	Sequencing data	Concordance	Sensitivity	Specificity	Comment	References
2013	<i>S. Typhimurium</i>	49	MIC	17	ResFinder	Assembled, Velvet	99.74%			Disagreement: 7 isolates including 6 <i>E. coli</i> resistant to Spec	(7)
	<i>E. coli</i>	48									
	<i>E. faecalis</i>	50									
	<i>E. faecium</i>	50									
2013	<i>E. coli</i> (ESBL)	74	DD	7	BLASTn, selected panel	Assembled, Velvet		96%	97%	VM rate: 1.2%/M rate: 2.1%	(8)
	<i>K. pneumonia</i> (ESBL)	69									
2014	<i>S. aureus</i>	501	DD/MIC (Vitek)	12	BLASTn, selected panel	Assembled, Velvet		97%	99%	VM rate: 0.5%/M rate: 0.7%	(9)
2016	<i>C. jejuni</i>	32	MIC	9	BLASTx	Assembled, CLC-bio	99.2%			Lower concordance to Gen, Azi, Clin, Tel	(10)
	<i>C. coli</i>	82									
2016	<i>S. enterica</i>	104	MIC	14	ResFinder/ ARG-ANNOT/ CARD/BLAST	Assembled, CLC-bio	99.0%	99.2%	99.3%	Lower concordance to aminoglycosides/β-lactams	(11)
		536									
2017	<i>E. coli</i>	31	MIC	4	Custom DB based on ARDB/CARD/β-lactamase alleles			87%	98%	Neg. predictive value: 97% Pos. Predictive value: 91%	(12)
	<i>K. pneumonia</i>	24									
	<i>P. aeruginosa</i>	22									
	<i>E. cloacae</i>	13									
2017	<i>S. enterica</i>	50	MIC	4	ResFinder/ PointFinder	Assembled, SPAdes	98.4%			Disagreement: 2/2 <i>C.jejuni</i> to FQ/ERY 5 <i>E. coli</i> to COL (pmrB)	(13)
	<i>E. coli</i>	50									
	<i>C. jejuni</i>	50									
2018	<i>E. faecalis</i>	97	MIC	11	ResFinder/NCBI Pathogen DB/BLAST	Assembled, CLC-bio	96.5%				(14)
	<i>E. faecium</i>	100									
2018	<i>S. aureus</i>	501	DD/MIC	12	GeneFinder/ Mykrobe/ Typewriter	FASTQ/assembled, BLAST	98.3%			Disagreements: 0.7% predicted resistant 0.6% predicted susceptible	(15)
		491									
		397									
2018	<i>M. tuberculosis</i>	10,209	MGIT 960	4	Cortex	Assembled	89.5%			97.1%/99.0% predicted R/S 97.5%/98.8% predicted R/S 94.6%/93.6% predicted R/S 91.3%/96.8% predicted R/S	(16)
				4							
				4							
				4							
2019	<i>H. pylori</i>	140	MIC (E-test)	5	ARIBA	FASTQ	99%			Phenotype issues to metronidazole	(17)

1) ESBL: Extended Spectrum Beta-Lactamase, 2) MIC: Minimum Inhibitory Concentration, 3) DD: Disk diffusion, 4) VM: Very Major, 5) M: Major, 6) R/S: Resistant/Susceptible, 7) SPEC: Spectinomycin, 8) GEN: Gentamicin, 9) AZI: Azithromycin, 10) CLIN: Clindamycin, 11) TEL: Telithromycin, 12) FQ: Fluoroquinolone, 13) ERY: Erythromycin, 14) COL: colistin.

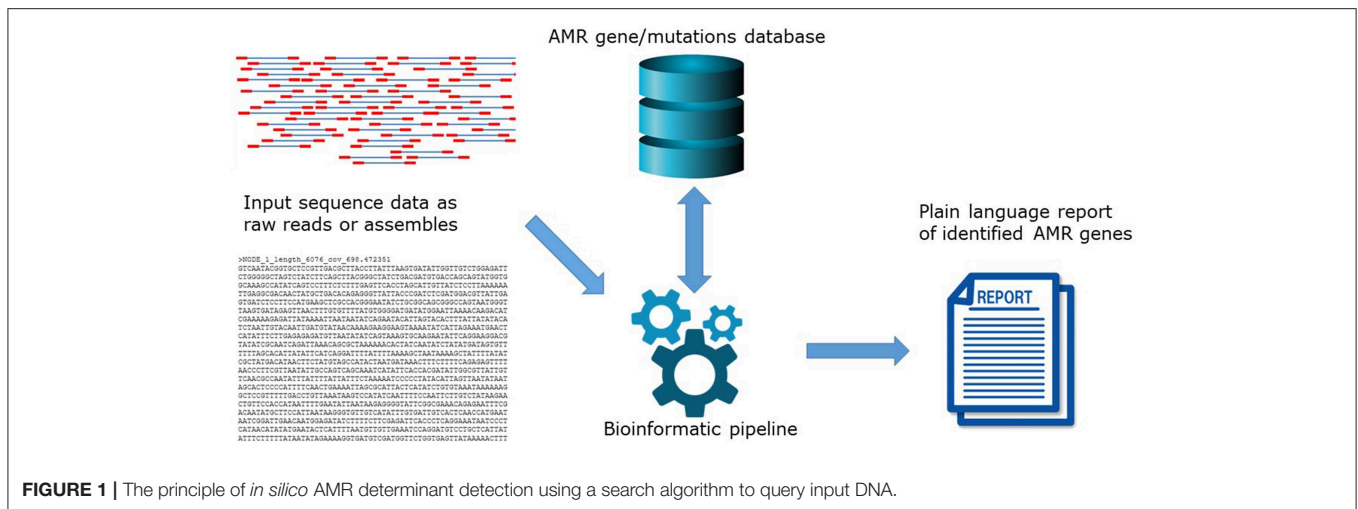


FIGURE 1 | The principle of *in silico* AMR determinant detection using a search algorithm to query input DNA.

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 developed 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.

Name	Target	Software		Database		Input sequence		Link	Year of development	Curation (last update)	References
		Type	Downloadable ^a	Source	Downloadable	Type	Format				
ABRES Finder	General AMR	Profile HMM	No	Own	No	Amino acid	FASTA	http://scbt.sastra.edu/ABRES/index.php	2017	Not specified	Unpublished
ABRICATE	General AMR	BLAST	Yes	ResFinder, CARD, ARG-ANNOT, NCBI AMRFinder, EcOH, PlasmidFinder, Ecoli_VF and VFDB	Yes	Nucleotide	FASTA	https://github.com/tseemann/abricate	2016	2019	Unpublished
ARDB	General AMR	BLAST	Yes	Own	Yes	Nucleotide	FASTA	https://ardb.cbcb.umd.edu/	2009	2009	(28)
ARG-ANNOT	General AMR	–	–	Own	Yes	–	–	Discontinued	2014	2018	(29)
ARIBA	General AMR (single isolate sequences)	Minimap, Bowtie2	Yes	Derived from ARG-ANNOT, CARD, PlasmidFinder, ResFinder, VFDB ^b ; customizable	No	Nucleotide	FASTQ	https://github.com/sanger-pathogens/ariba	2017	2019	(30)
CARD	General AMR	BLAST, RGI	Yes	Own	Yes	Nucleotide, amino acid	FASTA	https://card.mcmaster.ca/home	2013	2019	(31)
IRIDA plugin AMR detection	General AMR	RGI, staramr	Yes	CARD, PointFinder, PlasmidFinder and ResFinder	Yes	Nucleotide	FASTQ	https://github.com/phac-nml/irida-plugin-amr-detection	2019	2019	Unpublished
Kmer resistance	General AMR	KMA	Yes	ResFinder	Yes	Nucleotide	FASTA, FASTQ	https://cge.cbs.dtu.dk/services/KmerResistance-2.2/	2016	2019	(26)
MEGARes (AMRplusplus)	General AMR	BWA	Yes	Derived from ARG-ANNOT, CARD, NCBI Lahey Clinic beta-lactamase archive, ResFinder ^b	Yes	Nucleotide	FASTQ	https://megares.meglab.org/	2016	2016	(32)

(Continued)

TABLE 2 | Continued

Name	Target	Software		Database		Input sequence		Link	Year of development	Curation (last update)	References
		Type	Downloadable ^a	Source	Downloadable	Type	Format				
NCBI AMRFinder	General AMR	BLAST, HMMER	Yes	Own	Yes	Nucleotide, amino acid	FASTA, GFF	https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/	2017	2019	(33)
Noradab	General AMR	BLAST	No	Derived from ARDB and CARD ^b	Yes	Nucleotide, amino acid	FASTA	http://noradab.bi.up.ac.za/	2018	Not specified	(34)
Patric	General AMR	BLAST	Yes	Own	Yes	Nucleotide, amino acid	FASTA	https://www.patricbr.org/	2004	2019	(35)
ResFinder	General AMR	BLAST, KMA	Yes	Own	Yes	Nucleotide	FASTA, FASTQ	https://cge.cbs.dtu.dk/services/ResFinder/	2012	2019	(36)
SRST2	General AMR	BOWTIE2	Yes	Derived from ARG-ANNOT ^b	Yes	Nucleotide	FASTA, FASTQ and any other format readable by BOWTIE2	https://github.com/katholt/srst2	2014	2019	(37)
SSTAR	General AMR	BLAST	Yes	Derived from ARG-ANNOT and Resfinder ^b	Yes	Nucleotide	FASTA	https://github.com/tomdeman-bio/Sequence-Search-Tool-for-Antimicrobial-Resistance-SSTAR	2015	2018	(38)
INTEGRALL	AMR genes and associated integrons	BLAST	No	Own	Yes	Nucleotide	FASTA	http://integrall.bio.ua.pt/?	2008	2019	(39)
MvirDB	AMR genes, protein toxins and virulence factors for bio-defense applications	BLAST	No	Derived from Tox-Prot, SCORPION, the PRINTS virulence factors, VFDB, TVFac, Islander, ARGO and a subset of VIDA ^b	Yes	Nucleotide, amino acid	FASTA	Discontinued (http://mvirdb.llnl.gov/)	2007	Not specified	(40)
BacMet	Biocide and metal resistance	BLAST	No	Own	Yes	Nucleotide, amino acid	FASTA	http://bacmet.biomedicine.gu.se/	2013	2018	(41)
ResCap	Antibiotic, heavy metal and biocide resistance	BLAST, Bowtie2	Yes	Derived from ARG-ANNOT, CARD, RED-DB, ResFinder, Bacmet ^b	Yes	Nucleotide	FASTA, FASTQ	https://github.com/vallanza/ResCap	2017	2017	(42)

(Continued)

TABLE 2 | Continued

Name	Target	Software		Database		Input sequence		Link	Year of development	Curation (last update)	References
		Type	Downloadable ^a	Source	Downloadable	Type	Format				
ARGO	Beta-lactam and vancomycin resistance	–	–	Own	–	–	–	Discontinued (http://bioinformatics.org/argo/beta/antibioticresistance.php)	2005	–	(43)
RED-DB	Beta-lactam, glycopeptide, aminoglycoside, tetracycline, sulphonamide, macrolide, lincosamide, streptogramin b, oxazolidinone and quinolone resistance	BLAST	No	Own	Yes	Nucleotide, amino acid	FASTA	http://www.fibim.unisi.it/REDDB/	2007-2013	Not specified	Unpublished
Tetracycline MLS nomenclature	Macrolide, lincosamide, – streptogramin and tetracycline resistance	–	–	Own	Yes	–	–	https://faculty.washington.edu/marylryn/	Not specified	2019	Unpublished
β -lactamases Database	β -lactamases	–	–	Own	Yes	–	–	http://ifr48.timone.univ-mrs.fr/beta-lactamase/public/	Not specified	Not specified	Unpublished
BLAD	β -lactamases	–	–	Own	No	Nucleotide, amino acid	FASTA	http://www.blad.co.in/	2012	Not specified	Unpublished
BLDB	β -lactamases	BLAST	No	Own	Yes	Nucleotide, amino acid	FASTA	http://bladb.eu/	2017	2019	(44)
CBMAR	β -lactamases	BLAST	No	Own	Yes	Nucleotide, amino acid	FASTA	http://proteininformatics.org/mkumar/lactamasedb/	2014	2014	(45)
LacED	β -lactamases	BLAST	No	Own	Yes	Amino acid	FASTA	http://www.laced.uni-stuttgart.de/	2009	Not specified ^c	(46)
AMRtime	AMR genes in metagenomic data	DIAMOND	Yes	CARD	Yes	Nucleotide	FASTQ	https://github.com/beiko-lab/AMRtime	2017	2019	(47)
DeepARG	AMR genes in metagenomic data	BLAST, DIAMOND	Yes	Derived from RDB, CARD, UNIPROT ^b	Yes	Nucleotide, amino acid	FASTA, FASTQ	https://bench.cs.vt.edu/deeparg	2017	2019	(48)
GROOT	AMR genes in metagenomic data	LSH Forest indexing	Yes	Derived from ARG-ANNOT, CARD, Resfinder	Yes	Nucleotide	FASTQ	https://github.com/will-rowe/groot	2018	2019	(49)
SARG (ARGs-OAP; ARGpore)	AMR genes in metagenomic data	BLAST, HMMER, UBLAST	Yes	Derived from ARDB and CARD ^b	Yes	Nucleotide	any format is supported	https://smile.hku.hk/SARGs	2016	2019	(50)
SEAR	AMR genes in metagenomic data	BLAST, BWA-MEM	Yes	ARG-ANNOT	Yes	Nucleotide	FASTQ	Discontinued (https://github.com/will-rowe/SEAR)	2015	2018	(51)

(Continued)

TABLE 2 | Continued

Name	Target	Software	Database		Input sequence		Link	Year of development	Curation (last update)	References	
			Type	Downloadable ^a	Source	Downloadable					Type
ShortBRED	AMR genes in metagenomic data	BLAST, USEARCH	Yes	Derived from ARDB and CARD ^b	Yes	Amino acid	FASTA	http://hutterhower.sph.harvard.edu/shortbred	2015	2019	(52)
Mustard	AMR determinants in the human gut microbiota	BLAST	No	Derived from Resfinder, ARG-ANNOT, the Lahey Clinic (http://www.lahey.org/studies/), RED-DB (http://www.fibim.unisi.it/REDDB/), Marilyn Roberts' website for macrolides and tetracycline resistance (http://faculty.washington.edu/marilynr/) and different functional metagenomics studies ^b	Yes	Nucleotide, amino acid	FASTA	http://mgps.eu/Mustard/	2017	2017	(53)
FARMEDB	AMR genes discovered by functional metagenomics	BLAST	No	Own	Yes	Nucleotide, amino acid	FASTA	http://staff.washington.edu/jwallace/farme/index.html	2016	Not specified ^c	Unpublished
ResFams	AMR genes discovered – by functional metagenomics	–	–	Derived from CARD, LacED, Lahey beta-lactamases (now at NCBI) ^b	Yes	–	–	http://www.dantaslab.org/resfams	2014	2018	(54)
ResFinderFG	AMR genes discovered by functional metagenomics	BLAST	Yes	Own	No	Nucleotide	FASTA, FASTQ	https://cge.cbs.dtu.dk/services/ResFinderFG-1.0/	2016	Not specified	Unpublished
Galileo AMR (MARA, RAC)	AMR genes in Gram-negative bacteria	BLAST (ATTACCA)	Yes	Own	Yes	Nucleotide	FASTA	https://galileoamr.arcbio.com/mara/	2017	Not specified ³	(55)
LREfinder	Linezolid resistance in enterococci	KMA	Yes	Own	Yes	Nucleotide	FASTA, FASTQ	https://cge.cbs.dtu.dk/services/LRE-finder/	2019	2019	(56)

(Continued)

TABLE 2 | Continued

Name	Target	Software		Database		Input sequence		Link	Year of development	Curation (last update)	References
		Type	Downloadable ^a	Source	Downloadable	Type	Format				
MUBII-TB-DB	AMR mutations in Mycobacterium tuberculosis	BLAST	No	Own	No	Nucleotide	FASTA	https://umr5558-bibiserv.univ-lyon1.fr/mubii/mubii-select.cgi	2013	Not specified	(57)
Mykrobe	AMR in Mycobacterium tuberculosis and Staphylococcus aureus	Own (based on de Bruijn graph)	Yes	Own	Yes	Nucleotide	FASTQ	http://www.mykrobe.com/	2015	2019	(58)
TBDReaM	AMR in Mycobacterium tuberculosis	–	–	Own	Yes	–	–	https://tbdreamdb.ki.se/Info/	2009	2014	(59)
PointFinder	Selected mutations in chromosomal genes of Escherichia coli, Salmonella sp., Campylobacter sp., Staphylococcus aureus, Enterococcus sp., Mycobacterium tuberculosis, Neisseria gonorrhoeae	BLAST, KMA	Yes	Own	Yes	Nucleotide	FASTA, FASTQ	https://cge.cbs.dtu.dk/services/ResFinder/	2017	2019	(13)
SCCmec Finder	SCCmec elements in Staphylococcus aureus	BLAST, KMA	Yes	Own	Yes	Nucleotide	FASTA, FASTQ	https://cge.cbs.dtu.dk/services/SCCmecFinder/	2016	2018	(60)
U-CARE	AMR in Escherichia coli	BLAST	No	Own	Yes	Amino acid	FASTA	http://www.e-bioinformatics.net/ucare/	2013	Not specified	(61)
ARGDIT	Toolkit for validation and integration of AMR gene database	–	Yes	–	–	Nucleotide, amino acid	FASTA	https://github.com/phglab/ARGDIT	2018	2019	(62)
ARG-miner	Robust and comprehensive curation of AMR gene databases	–	–	Derived from ARDB, ARG-ANNOT, CARD, DeepARG-DB, MEGARes, NDARO, ResFinder, SARG, UniProt ^b	Yes	–	–	https://bench.cs.vt.edu/argminer/#/home	2018	2019 (crowd-curation)	(48)

^aYes, standalone version is available (usually in Bitbucket or in GitHub) either with or without a corresponding web version; no, only web version is available.

^bCuration to avoid redundancies and remove selected sequences (see respective references for details).

^cActive, 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 Stoesser 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

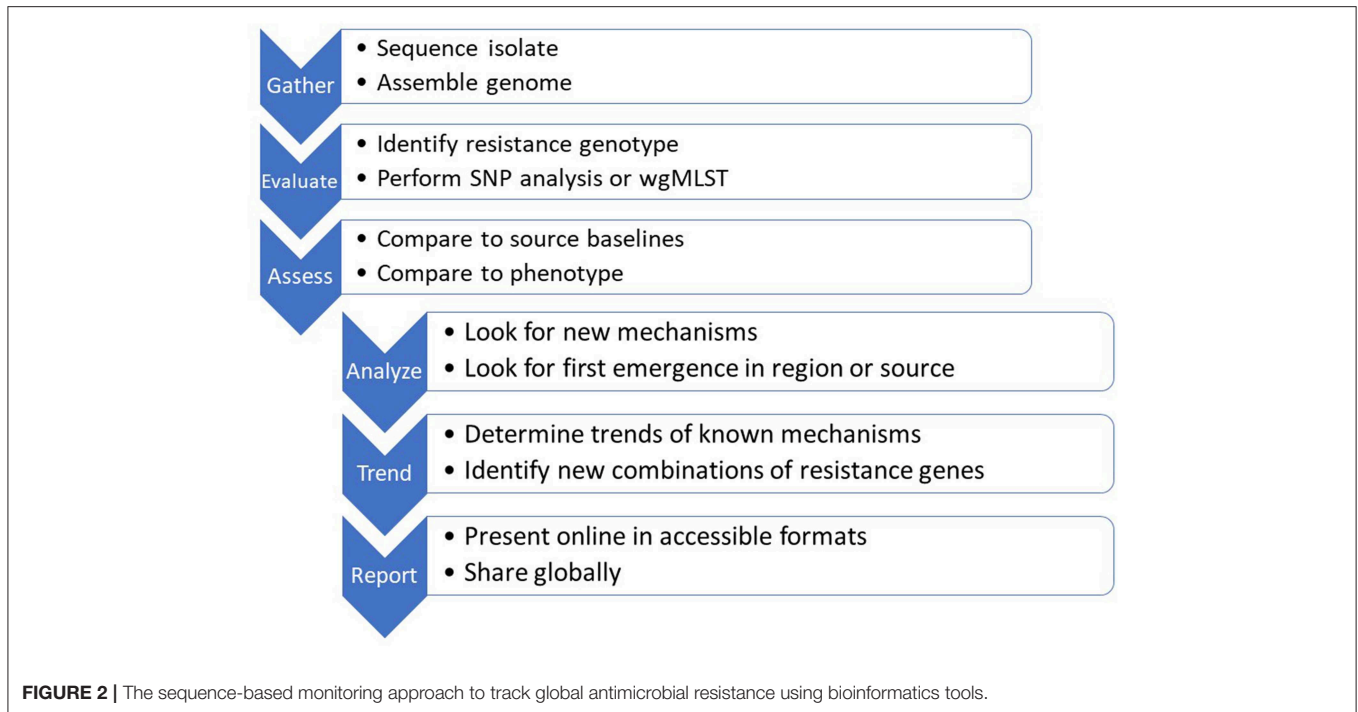
library construction were provided to participating laboratories for DNA purification, library preparation, and WGS followed by *in silico* prediction of wgMLST and AMR determinants. The genomes and analysis were submitted to DTU for quality control assessment using closed genomes of the test strains 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

- Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet.* (2012) 13:601–12. doi: 10.1038/nrg3226
- Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet.* (2014) 15:49–55. doi: 10.1038/nrg3624
- Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* (2012) 8:e1002824. doi: 10.1371/journal.ppat.1002824
- Allard MW, Stevens EL, Brown EW. All for one and one for all: the true potential of whole-genome sequencing. *Lancet Infect Dis.* (2019) 19:683–4. doi: 10.1016/S1473-3099(19)30172-0
- O'Neill J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations. The Review on Antimicrobial Resistance* London, UK: Wellcome Trust and the UK Department of Health (2016).
- Cartwright EJ, Patel MK, Mbopi-Keou FX, Ayers T, Haenke B, Wagenaar BH, et al. Recurrent epidemic cholera with high mortality in Cameroon: persistent challenges 40 years into the seventh pandemic. *Epidemiol Infect.* (2013) 141:2083–93. doi: 10.1017/S0950268812002932
- Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agerso Y, Lund O, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother.* (2013) 68:771–7. doi: 10.1093/jac/dks496
- Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo EC, et al. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother.* (2013) 68:2234–44. doi: 10.1093/jac/dkt180
- Gordon NC, Price JR, Cole K, Everitt R, Morgan M, Finney J, et al. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol.* (2014) 52:1182–91. doi: 10.1128/JCM.03117-13
- Zhao S, Tyson GH, Chen Y, Li C, Mukherjee S, Young S, et al. Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in *Campylobacter* spp. *Appl Environ Microbiol.* (2015) 82:459–66. doi: 10.1128/AEM.02873-15
- McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, et al. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal salmonella. *Antimicrob Agents Chemother.* (2016) 60:5515–20. doi: 10.1128/AAC.01030-16
- Shelburne SA, Kim J, Munita JM, Sahasrabhojane P, Shields RK, Press EG, et al. Whole-genome sequencing accurately identifies resistance to extended-spectrum beta-lactams for major gram-negative bacterial pathogens. *Clin Infect Dis.* (2017) 65:738–45. doi: 10.1093/cid/cix417
- Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother.* (2017) 72:2764–8. doi: 10.1093/jac/dkx217
- Tyson GH, Sabo JL, Rice-Trujillo C, Hernandez J, McDermott PF. Whole-genome sequencing based characterization of antimicrobial resistance in *Enterococcus*. *Pathog Dis.* (2018) 76:4931055. doi: 10.1093/femspd/fty018
- Mason A, Foster D, Bradley P, Golubchik T, Doumith M, Gordon NC, et al. Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from *Staphylococcus aureus* whole-genome sequences. *J Clin Microbiol.* (2018) 56:e01815-17. doi: 10.1128/JCM.01815-17
- CRYPtic Consortium and the 100,000 Genomes Project, Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med.* (2018) 379:1403–15. doi: 10.1056/NEJMoa1800474
- Lauener FN, Imkamp F, Lehours P, Buissonniere A, Benejat L, Zbinden R, et al. Genetic determinants and prediction of antibiotic resistance phenotypes in *Helicobacter pylori*. *J Clin Med.* (2019) 8:E53. doi: 10.3390/jcm8010053
- Su M, Satola SW, Read TD. Genome-based prediction of bacterial antibiotic resistance. *J Clin Microbiol.* (2019) 57:e01405-18. doi: 10.1128/JCM.01405-18
- Eyre DW, De SD, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict antibiotic MICs for *Neisseria gonorrhoeae*. *J Antimicrob Chemother.* (2017) 72:1937–47. doi: 10.1093/jac/dkx067
- Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, et al. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal salmonella. *J Clin Microbiol.* (2019) 57:e01260-18. doi: 10.1128/JCM.01260-18
- Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, et al. Developing an *in silico* minimum inhibitory concentration panel test for *Klebsiella pneumoniae*. *Sci Rep.* (2018) 8:421–18972. doi: 10.1038/s41598-017-18972-w
- Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clin Microbiol Infect.* (2017) 23:2–22. doi: 10.1016/j.cmi.2016.11.012
- Hoffmann M, Zhao S, Pettengill J, Luo Y, Monday SR, Abbott J, et al. Comparative genomic analysis and virulence differences in closely related salmonella enterica serotype heidelberg isolates from humans, retail meats, and animals. *Genome Biol Evol.* (2014) 6:1046–68. doi: 10.1093/gbe/evu079
- Magiorakos AP, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect.* (2012) 18:268–81. doi: 10.1111/j.1469-0691.2011.03570.x
- Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. *Nat Rev Genet.* (2019) 20:356–70. doi: 10.1038/s41576-019-0108-4
- Clausen PT, Zankari E, Aarestrup FM, Lund O. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. *J Antimicrob Chemother.* (2016) 71:2484–8. doi: 10.1093/jac/dkw184
- Clausen PT, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics.* (2018) 19:307–2336. doi: 10.1186/s12859-018-2336-6
- Liu B, Pop M. ARDB—antibiotic resistance genes database. *Nucleic Acids Res.* (2009) 37:D4437. doi: 10.1093/nar/gkn656
- Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother.* (2014) 58:212–20. doi: 10.1128/AAC.01310-13
- Hunt M, Mather AE, Sanchez-Buso L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom.* (2017) 3:e000131. doi: 10.1099/mgen.0.000131
- Jia B, Raphenya AR, Alcock B, Wagelchner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* (2017) 45:D566–D573. doi: 10.1093/nar/gkw1004
- Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, et al. MEGARes: an antimicrobial resistance database for high throughput sequencing. *Nucleic Acids Res.* (2017) 45:D574–80. doi: 10.1093/nar/gkw1009
- Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Using the NCBI AMRFinder tool to determine antimicrobial resistance genotype-phenotype correlations within a collection of NARMS isolates. *BioRxiv [Preprint].* (2019). doi: 10.1101/550707
- Van Goethem MW, Pierneef R, Bezuidt OKI, Van De PY, Cowan DA, Makhalanya TP. A reservoir of 'historical' antibiotic resistance genes in remote pristine Antarctic soils. *Microbiome.* (2018) 6:40–0424. doi: 10.1186/s40168-018-0424-5
- Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* (2017) 45:D535–D542. doi: 10.1093/nar/gkw1017
- Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* (2012) 67:2640–4. doi: 10.1093/jac/dks261
- Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* (2014) 6:90–0090. doi: 10.1186/s13073-014-0090-6

38. de Man TJ, Limbago BM. SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. *mSphere*. (2016) 1:e00050-15. doi: 10.1128/mSphere.00050-15
39. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*. (2009) 25:1096-8. doi: 10.1093/bioinformatics/btp105
40. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. *Nucleic Acids Res*. (2007) 35:D391-4. doi: 10.1093/nar/gkl791
41. Pal C, Gtsson-Palme J, Rensing C, Kristiansson E, Larsson DG. BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res*. (2014) 42:D737-43. doi: 10.1093/nar/gkt1252
42. Lanza VF, Baquero F, Martinez JL, Ramos-Ruiz R, Gonzalez-Zorn B, Andremont A, et al. In-depth resistome analysis by targeted metagenomics. *Microbiome*. (2018) 6:11-0387. doi: 10.1186/s40168-017-0387-y
43. Scaria J, Chandramouli U, Verma SK. Antibiotic Resistance Genes Online (ARGO): a Database on vancomycin and beta-lactam resistance genes. *Bioinformation*. (2005) 1:5-7. doi: 10.6026/97320630001005
44. Naas T, Oueslati S, Bonnin RA, Dabos ML, Zavala A, Dortet L, et al. Beta-lactamase database (BLDB) - structure and function. *J Enzyme Inhib Med Chem*. (2017) 32:917-9. doi: 10.1080/14756366.2017.1344235
45. Srivastava A, Singhal N, Goel M, Virdi JS, Kumar M. CBMAR: a comprehensive beta-lactamase molecular annotation resource. *Database*. (2014) 2014:bau111. doi: 10.1093/database/bau111
46. Thai QK, Bos F, Pleiss J. The lactamase engineering database: a critical survey of TEM sequences in public databases. *BMC Genomics*. (2009) 10:390. doi: 10.1186/1471-2164-10-390
47. Maguire F, Alcock B, Brinkman FS, McArthur AG, Beiko RG. AMRtime: Rapid Accurate Identification of Antimicrobial Resistance Determinants from Metagenomic Data. In: *Third American Society for Microbiology Meeting on Rapid Applied Microbial Next-Generation Sequencing and Bioinformatics Pipelines*. Washington, DC (2018).
48. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome*. (2018) 6:23-0401. doi: 10.1186/s40168-018-0401-z
49. Rowe WPM, Winn MD. Indexed variation graphs for efficient and accurate resistome profiling. *Bioinformatics*. (2018) 34:3601-8. doi: 10.1093/bioinformatics/bty387
50. Yin X, Jiang XT, Chai B, Li L, Yang Y, Cole JR, et al. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics*. (2018) 34:2263-70. doi: 10.1093/bioinformatics/bty053
51. Rowe W, Baker KS, Verner-Jeffreys D, Baker-Austin C, Ryan JJ, Maskell D, et al. Search engine for antimicrobial resistance: a cloud compatible pipeline and web interface for rapidly detecting antimicrobial resistance genes directly from sequence data. *PLoS ONE*. (2015) 10:e0133492. doi: 10.1371/journal.pone.0133492
52. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput Biol*. (2015) 11:e1004557. doi: 10.1371/journal.pcbi.1004557
53. Ruppe E, Ghazlane A, Tap J, Pons N, Alvarez AS, Maziers N, et al. Prediction of the intestinal resistome by a three-dimensional structure-based method. *Nat Microbiol*. (2019) 4:112-23. doi: 10.1038/s41564-018-0292-6
54. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J*. (2015) 9:207-16. doi: 10.1038/ismej.2014.106
55. Partridge SR, Tsafnat G. Automated annotation of mobile antibiotic resistance in Gram-negative bacteria: the Multiple Antibiotic Resistance Annotator (MARA) and database. *J Antimicrob Chemother*. (2018) 73:883-90. doi: 10.1093/jac/dkx513
56. Hasman H, Clausen TTL, Kaya H, Hansen F, Knudsen JD, Wang M, et al. LRE-Finder, a Web tool for detection of the 23S rRNA mutations and the *optrA*, *cfp*, *cfp(B)* and *poxTA* genes encoding linezolid resistance in enterococci from whole-genome sequences. *J Antimicrob Chemother*. (2019) 74:1473-6. doi: 10.1093/jac/dkz092
57. Flandrois JP, Lina G, Dumitrescu O. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics*. (2014) 15:107. doi: 10.1186/1471-2105-15-107
58. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat Commun*. (2015) 6:10063. doi: 10.1038/ncomms10063
59. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. *PLoS Med*. (2009) 6:e2. doi: 10.1371/journal.pmed.1000002
60. Kaya H, Hasman H, Larsen J, Stegger M, Johannesen TB, Allesoe RL, et al. SCCmecFinder, a web-based tool for typing of staphylococcal cassette chromosome *mec* in *Staphylococcus aureus* using whole-genome sequence data. *mSphere*. (2018) 3:e00612-17. doi: 10.1128/mSphere.00612-17
61. Saha SB, Uttam V, Verma V. u-CARE: user-friendly comprehensive antibiotic resistance repository of *Escherichia coli*. *J Clin Pathol*. (2015) 68:648-51. doi: 10.1136/jclinpath-2015-202927
62. Chiu JKH, Ong RT. ARGDIT: a validation and integration toolkit for antimicrobial resistance gene databases. *Bioinformatics*. (2019) 35:2466-74. doi: 10.1093/bioinformatics/bty987
63. Arango-Argoty GA, Guron GKP, Garner E, Riquelme MV, Heath LS, Pruden A, et al. ARGminer: a web platform for crowdsourcing-based curation of antibiotic resistance genes. *bioRxiv [Preprint]*. (2019). doi: 10.1101/274282
64. Xavier BB, Das AJ, Cochrane G, De GS, Kumar-Singh S, Aarestrup FM, et al. Consolidating and exploring antibiotic resistance gene data resources. *J Clin Microbiol*. (2016) 54:851-9. doi: 10.1128/JCM.02717-15
65. Angers-Loustau A, Petrillo M, Bengtsson-Palme J, Berendonk T, Blais B, Chan KG, et al. The challenges of designing a benchmark strategy for bioinformatics pipelines in the identification of antimicrobial resistance determinants using next generation sequencing technologies. *F1000Res*. (2018) 7:459. doi: 10.12688/f1000research.14509.1
66. Technical University of Denmark - National Food Institute. *Final report of ENGAGE - Establishing Next Generation sequencing Ability for Genomic analysis in Europe*. Istituto Zooprofilattico Sperimentale del Lazio e della Toscana; Federal Institute for Risk Assessment; National Institute of Public Health - National Institute of Hygiene; National Veterinary Research Institute; Public Health England; Animal and Plant Health Agency, and Istituto Zooprofilattico Sperimentale delle Venezie. EN-1431 (2018). 252 p.
67. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol*. (2012) 30:1033-6. doi: 10.1038/nbt.2403
68. Rossen JWA, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clin Microbiol Infect*. (2018) 24:355-60. doi: 10.1016/j.cmi.2017.11.001
69. Hendriksen RS, Seyfarth AM, Jensen AB, Whichard J, Karlsmose S, Joyce K, et al. Results of use of WHO Global Salm-Surv external quality assurance system for antimicrobial susceptibility testing of *Salmonella* isolates from 2000 to 2007. *J Clin Microbiol*. (2009) 47:79-85. doi: 10.1128/JCM.00894-08
70. Lo Fo Wong DM, Hendriksen RS, Mevius DJ, Veldman KT, Aarestrup FM. External quality assurance system for antibiotic resistance in bacteria of animal origin in Europe (ARBAO-II) 2003. *Vet Microbiol*. (2006) 115:128-39. doi: 10.1016/j.vetmic.2005.12.016
71. Pedersen SK, Wagenaar JA, Vigre H, Roer L, Mikoleit M, idara-Kane A, et al. Proficiency of WHO global foodborne infections network external quality assurance system participants in identification and susceptibility testing of thermotolerant *Campylobacter* spp. from 2003 to 2012. *J Clin Microbiol*. (2018) 56:e01066-18. doi: 10.1128/JCM.01066-18
72. Deng X, den Bakker HC, Hendriksen RS. Genomic epidemiology: whole-genome-sequencing-powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annu Rev Food Sci Technol*. (2016) 7:353-74. doi: 10.1146/annurev-food-041715-033259
73. Aarestrup FM, Koopmans MG. Sharing data for global infectious disease surveillance and outbreak detection. *Trends Microbiol*. (2016) 24:241-5. doi: 10.1016/j.tim.2016.01.009

74. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*. (2017) 544:309–15. doi: 10.1038/nature22040
75. Kaye J, Heeney C., Hawkins N, de Vries J, Boddington, P. Data sharing in genomics—re-shaping scientific practice. *Nat Rev Genet*. (2009) 10:331–5. doi: 10.1038/nrg2573
76. McArthur AG, Tsang KK. Antimicrobial resistance surveillance in the genomic age. *Ann N Y Acad Sci*. (2017) 1388:78–91. doi: 10.1111/nyas.13289
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.
78. Wielinga PR, Hendriksen RS, Aarestrup FM, Lund O, Smits SL, Koopmans MPG, et al. Global microbial identifier. In: Deng X, den Bakker HC, Hendriksen RS, editors. *Applied Genomics of Foodborne Pathogens*. Springer International Publishing Switzerland; Food Microbiology and Food Safety (2017). p. 13–32. doi: 10.1007/978-3-319-43751-4_2
79. World Health Organization. *Whole Genome Sequencing for Foodborne Disease Surveillance: Landscape Paper*. Geneva (2018). Available online at: <https://apps.who.int/iris/handle/10665/272430>.
80. Yozwiak NL, Schaffner SF, Sabeti PC. Data sharing: make outbreak research open access. *Nature*. (2015) 518:477–9. doi: 10.1038/518477a
81. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics*. (2011) 38:95–109. doi: 10.1016/j.jgg.2011.02.003
82. Ribeiro DS, van de Burgwal LHM, Regeer BJ. Overcoming challenges for designing and implementing the One Health approach: a systematic review of the literature. *One Health*. (2019) 7:100085. doi: 10.1016/j.onehlt.2019.100085
83. Ribeiro DS, Koopmans MP, Haringhuizen GB. Threats to timely sharing of pathogen sequence data. *Science*. (2018) 362:404–6. doi: 10.1126/science.aau5229
84. Otto M. Next-generation sequencing to monitor the spread of antimicrobial resistance. *Genome Med*. (2017) 9:68–0461. doi: 10.1186/s13073-017-0461-x
85. Matamoros S, Hendriksen RS, Pataki B, Pakseresht N, Rossello M, Silvester N, et al. Accelerating surveillance and research of antimicrobial resistance - an online repository for sharing of antimicrobial susceptibility data associated with whole genome sequences. *BioRxiv [Preprint]*. (2019). doi: 10.1101/532267
86. Amid C, Pakseresht N, Silvester N, Jayathilaka S, Lund O, Dynovski LD, et al. The COMPARE data hubs. *bioRxiv [Preprint]*. (2019). doi: 10.1101/555938
87. Thomsen MC, Ahrenfeldt J, Cisneros JL, Jurtz V, Larsen MV, Hasman H, et al. A bacterial analysis platform: an integrated system for analysing bacterial whole genome sequencing data for clinical diagnostics and surveillance. *PLoS ONE*. (2016) 11:e0157718. doi: 10.1371/journal.pone.0157718
88. Ribeiro CDS, van Roode MY, Haringhuizen GB, Koopmans MP, Claassen E, van de Burgwal LHM. How ownership rights over microorganisms affect infectious disease control and innovation: a root-cause analysis of barriers to data sharing as experienced by key stakeholders. *PLoS ONE*. (2018) 13:e0195885. doi: 10.1371/journal.pone.0195885
89. Contreras JL. NIH's genomic data sharing policy: timing and tradeoffs. *Trends Genet*. (2015) 31:55–7. doi: 10.1016/j.tig.2014.12.006
90. Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. *Expert Rev Mol Diagn*. (2014) 14:1053–65. doi: 10.1586/14737159.2014.961917
91. Timme RE, Sanchez Leon M, Allard MW. Utilizing the public GenomeTrakr database for foodborne pathogen traceback. *Methods Mol Biol*. (2019) 1918:201–12. doi: 10.1007/978-1-4939-9000-9_17
92. Karp BE, Tate H, Plumlee JR, Dessai U, Whichard JM, Thacker EL, et al. National antimicrobial resistance monitoring system: two decades of advancing public health through integrated surveillance of antimicrobial resistance. *Foodborne Pathog Dis*. (2017) 14:545–57. doi: 10.1089/fpd.2017.2283
93. Munk P, Andersen VD, de Knecht L, Jensen MS, Knudsen BE, Lukjancenko O, et al. A sampling and metagenomic sequencing-based methodology for monitoring antimicrobial resistance in swine herds. *J Antimicrob Chemother*. (2017) 72:385–92. doi: 10.1093/jac/dkw415
94. Munk P, Knudsen BE, Lukjancenko O, Duarte ASR, Van GL, Luiken REC, et al. Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nat Microbiol*. (2018) 3:898–908. doi: 10.1038/s41564-018-0192-9
95. Van GL, Luiken REC, Sarrazin S, Munk P, Knudsen BE, Hansen RB, et al. The antimicrobial resistome in relation to antimicrobial use and biosecurity in pig farming, a metagenome-wide association study in nine European countries. *J Antimicrob Chemother*. (2019) 74:865–76. doi: 10.1093/jac/dky518
96. Nordahl PT, Rasmussen S, Hasman H, Caroe C, Baelum J, Schultz AC, et al. Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. *Sci Rep*. (2015) 5:11444. doi: 10.1038/srep11444
97. Hendriksen RS, Munk P, Njage P, van Bunnik B., McNally, L., Lukjancenko, O., et al. (2019) Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun*. 10, 1124-08853. doi: 10.1038/s41467-019-08853-3
98. Pehrsson EC, Tsukayama P, Patel S, Mejia-Bautista M, Sosa-Soto G, Navarrete KM, et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature*. (2016) 533:212–6. doi: 10.1038/nature17672

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.