FIND
Diagnosis for all

Sequencing for antimicrobial resistance surveillance

Sequencing for antimicrobial resistance surveillance

February 2022

This report was developed by PHG Foundation for FIND. We would like to thank those who contributed to the development and review of this report:

Writers:

Chantal Babb de Villiers (PHG Foundation)

Laura Blackburn (PHG Foundation)

Joanna Janus (PHG Foundation)

Heather Turner (PHG Foundation)

Reviewers:

Sophie Crettaz (FIND)

Ana Belen Ibarz (FIND)

Cecilia Ferreyra (FIND)

Antonio Pedrotta (FIND)

Anita Suresh (FIND)

Swapna Uplekar (FIND)

URLs correct as of 11th February 2022

This report is the result of PHG Foundation's independent research and analysis and is not linked to a third party in any way. PHG Foundation has provided occasional analytical services to Oxford Nanopore Technologies (ONT) as part of a consultancy agreement.

Contents

1 Introduction

Antimicrobial resistance (AMR) is recognised as a major threat to human health globally. The World Health Organization (WHO) has declared that AMR is “one of the top 10 global public health threats facing humanity”. The emergence and spread of drug-resistant pathogens is growing with few new and innovative drugs being developed and entering the market. For example, in 2019 the WHO identified 32 antibiotics that were in clinical development to treat infections caused by bacteria on the WHO list of priority pathogens. Of these, only six were classified as innovative, meaning that they have novel mechanisms of action [1]. The Review on Antimicrobial Resistance, commissioned by the UK government in 2014, estimated that by 2050 10 million lives could be lost globally per year to drug resistant infections, at a cumulative cost of 100 trillion USD to the world economy [2].

In 2015, the WHO launched the *Global Action Plan to tackle AMR (GAP-AMR)*, which aims to ensure the ongoing successful treatment and prevention of infectious diseases with safe and effective medicines that are accessible to all [3]. The five objectives outlined in the plan include “strengthening the knowledge and evidence base through surveillance and research” – surveillance is recognised as being essential to inform infection prevention, control measures and policy. As part of GAP-AMR, the WHO also launched the *Global Antimicrobial Resistance and Use Surveillance System (GLASS)*, an initiative to standardise AMR surveillance worldwide. The *GLASS whole-genome sequencing for surveillance of antimicrobial resistance (2020)* report outlines how next generation sequencing (NGS), particularly whole genome sequencing (WGS), can support AMR surveillance efforts [4]. The report aims to assist decision-making around surveillance and surveillance technologies in countries looking to expand their efforts, while also outlining the advantages and disadvantages of WGS approaches.

1.1 Antimicrobial resistance

Antimicrobials are drugs used to treat infections caused by a range of organisms – bacteria, viruses, fungi and parasites – in humans, animals and plants, or other agents that kill or reduce the growth of microorganisms. AMR can occur through intrinsic or acquired mechanisms. Intrinsic resistance is due to naturally occurring mechanisms, whereas acquired resistance arises when these organisms evolve resistance in response to repeated antimicrobial use as a medicine or as a prophylactic, e.g. antibiotics used as growth promoters in agricultural animals. Acquired resistance can occur via a variety of genetic mechanisms which result in a partial or total loss of antimicrobial effectiveness. For example, in bacteria intrinsic resistance is due to naturally occurring genetic variation that results in a resistance mechanism, for example a cell membrane efflux pump. Acquired resistance occurs in response to selection pressure or due to gene/plasmid transfer from another bacteria e.g. horizontal gene transfer, gene duplication, new promoters being introduced, point mutations, or an insertion/deletion.

When AMR is discussed, the term can be conflated with antibiotic resistance, partly due to the scale of the challenge of drug resistant bacterial infections. However, when discussing AMR it should be clear that the principles established are applicable to the surveillance of all pathogenic organisms, not just bacteria. There is discussion in the literature about whether the distinction between antibiotic resistance and AMR should be kept when exploring these issues, to maintain focus on the importance of antibiotic resistance as a global threat [5]. This report will use both terms as appropriate depending on the context and the topic or case study under discussion, given that many of the issues discussed around the use of WGS for AMR surveillance are relevant to a range of pathogens, not just antibiotics and bacteria.

1.1.1 Overview of current AMR detection methods

A number of molecular methods are being employed to identify AMR with a large focus on antibiotic resistance of bacteria. Conventional methods use phenotypic approaches, known as antimicrobial susceptibility testing (AST), which depend upon culturing and isolating the pathogen under investigation and using different assays to identify and quantify the level of resistance to different antimicrobials. These approaches have been used extensively with the advantage that they are comparatively simple, quantify the minimum inhibitory concentration (MIC) and typically enable pathogen identification [6]. However, these methods are limited by the need to test individual purified strains with the ease of culturing differing between organisms and there is some disagreement between the various standards for AST. The main limitations of the currently available tools are [6]:

- The need for sample pre-treatment steps. For example, these can increase turnaround times or limit the number of samples that can be analysed at once – the use of MALDI-TOF mass spectrometry is not possible for mixed samples, which need to be processed and purified beforehand to ensure accurate characterisation by the device.
- Low sensitivity. Some technologies are not able to detect low levels of pathogenic organisms in a sample.
- Microorganism or resistance identification is sometimes not possible, for example when the pathogen or the mechanism causing resistance are novel.
- Lack of integration, automation, and portability of tools.

Molecular methods are able to overcome some of these challenges. They enable rapid identification of antimicrobial resistance genes and have the advantage of being readily adaptable to newly identified resistance factors [6]. Molecular-based methods fall into four broad categories [7]:

- Amplification tests
- Hybridisation tests
- Immunoassays
- Sequencing tests

Each of the methods outlined above require specific methods, equipment, trained personnel, and potentially costly reagents. They also provide different types of information about a pathogen that can be combined to provide more comprehensive information about a pathogen's biology and epidemiology.

Sequencing methods may be used for targeted detection of AMR genes or mutations – where only a selected subsection of the genome is sequenced – while whole genome sequencing (WGS) can be used to determine the nearly complete DNA sequence of a pathogen inclusive of all regions of the genome beyond/outside of the AMR genes [1].

1.2 AMR surveillance

The majority of AMR surveillance currently undertaken makes use of commonly available microbiology methods and phenotypic drug susceptibility testing, and some molecular methods. Sequencing technologies are only currently used in a limited capacity.

Some AMR surveillance efforts use the One Health approach, which involves collaboration across sectors and disciplines, from the local to the global level, with the aim of achieving optimal health outcomes while recognising the connections between humans, animals, plants and the environment [8]. Common One Health issues focus on shared health threats, for example zoonotic diseases, food safety, and AMR. Tackling AMR using a One Health approach requires inputs from stakeholders in human, animal, plant and environmental health and research. To provide fully comprehensive AMR surveillance using sequencing

data will require information from each of these sectors, and mechanisms to share and act on the data.

The settings in which AMR surveillance can take place include:

- Healthcare sector – In healthcare settings as a general surveillance/monitoring tool or for more targeted outbreak management, in hospitals and community healthcare settings. Samples can be taken from patients, their visitors, healthcare workers and/or the healthcare environment.
- Food sector - In the food sector to monitor and maintain food safety, from the food source, through the food chain to the consumer.
- Veterinary sector – to monitor the health and treatment of animals.
- Environmental sector – Wider environmental/regional surveillance, where the One Health approach can be deployed to investigate the prevalence of AMR in the environment e.g. waste water, agriculture e.g. both animals and their immediate environment or plants.

1.3 Next generation sequencing and surveillance

During the ongoing (as of February 2022) COVID-19 pandemic the value of using next generation sequencing (NGS) technologies to detect SARS-CoV-2 variants and monitor their spread during the pandemic has been demonstrated in both higher- (HICs) and lower- and middle-income countries (LMICs). These technologies are becoming more widely used to support infectious disease management and there is an opportunity to explore how these technologies could be leveraged to mitigate other ongoing and future threats such as AMR, particularly in LMICs.

Among NGS approaches, WGS provides the entire nucleotide sequence of an organism's genome which is the highest resolution information available. In many cases it provides greater sensitivity and specificity in terms of pathogen identification, relatedness to other pathogens, virulence information, and drug resistance or susceptibility, all via one assay. It is being proposed as a method to replace most, if not all, the currently used phenotypic and molecular methods for a range of pathogens.

Despite these benefits, WGS is not a 'one size fits all' technology and its utility varies between pathogens and surveillance use. This is due to: the depth and breadth of knowledge around pathogen genomics and genotype/phenotype interactions; access to sequencing infrastructure, expertise, analysis tools and data storage; availability of cheap and quick phenotypic or molecular tests. Different types of sequencing approaches (targeted sequencing, metagenomics or whole genome sequencing) will be suitable for different surveillance situations, and while there are challenges, there is also much potential to be realised. With the costs of sequencing continuing to fall, and the increasing availability of bioinformatics tools to analyse data, the technology could expand the reach of AMR surveillance since it can be applied to any pathogen of interest.

The use of sequencing in an AMR surveillance context can bring value in the following areas:

1. Informed clinical decision making
2. Support for quicker and more accurate identification of outbreaks
3. Information beyond AMR predictions to aid understanding of outbreaks and guide interventions
4. Stored WGS sequence data allows retroactive analysis when new information appears
5. Provides information about the genetic mechanisms underlying resistance
6. Allows identification of new and emerging resistance mutations
7. Supports understanding of how AMR is spreading between pathogens, animals, humans and the environment.

8. Allows linking of data from different fields: clinics, environment, food and animals (One Health), in turn providing a more joined-up approach to the emergence of potential threats
9. Data accumulation allows better understanding of the evolution of AMR over time

FIND (Foundation for Innovative New Diagnostics) has recognised the importance of sequencing technologies in the management of infectious diseases and recognises the need to support implementation of these technologies to improve AMR surveillance. However, further understanding is required on the variety of NGS approaches, including WGS, that can be used for AMR surveillance. Certain technical features and product characteristics will make some technologies more or less amenable for implementation in lower resource health settings. This report outlines the technical features and characteristics of NGS technologies that could be applied to support AMR surveillance, and explores the opportunities and challenges to be considered around implementing these technologies in lower resource health settings.

The analysis focuses on the priority pathogens outlined in the GLASS WGS-AMR surveillance report [4]. These pathogens were prioritised by the WHO due to the number of infections and associated morbidity/mortality, combined with resistance to more than one of the antibiotic classes used to treat them [9]:

- *E. coli*
- *K. pneumoniae*
- *Acinetobacter spp.*
- *S. aureus*
- *S. pneumoniae*
- *Salmonella spp.*
- *Shigella spp.*
- *N. gonorrhoeae*
- *P. aeruginosa* (not currently in GLASS but due to be added in the future)

While all the GLASS priority pathogens currently listed are bacteria, it should be noted that more pathogens are expected to be added in the future, most notably Candida spp. (fungus). Therefore, the term antimicrobial resistance is appropriate within the context of the GLASS pathogens.

1.4 Methods

This report summarises current knowledge on technology developments, protocols and global best practices. It is based on desk-based research and analysis informed by official publications, grey literature, peer-reviewed and pre-print literature.

Where appropriate, in-depth interviews (via telephone or video conference) have been conducted with experts and other relevant stakeholders to better understand the enablers and barriers to implementation and adoption of the techniques in appropriate settings. These experts are acknowledged in Appendix 8.6.

2 AMR surveillance

Infectious disease surveillance involves a range of stakeholders including the health system, public health authorities (including laboratories) and epidemiologists. These contribute to the different components of surveillance, which are collection, analysis, dissemination and response.

Collection and analysis of data can be done at a local, regional, national or international level. The public and private sector can contribute to these efforts; data can be collected through health records, health system reporting on particular pathogens or diseases, registries, or health system surveys. Data collection will differ depending on the event(s) undergoing surveillance, which can range from consistent background surveillance to more targeted surveillance of a particular population or geographical area in response to a disease outbreak. This process provides information to support the development of public health measures. These measures are then disseminated by the relevant authorities and the public health response is carried out via implementation of these measures.

Disease surveillance relies on ongoing and systematic data collection and analysis, supported by interpretation, and regular feedback of data on outcomes from public health response measures. AMR surveillance is now a core component of many countries' infectious disease surveillance programmes, however surveillance systems vary, depending on need, resources and the impact of the health threat on their population.

The World Health Organization describes AMR surveillance as a core pillar for the global action plan on AMR and the European strategic action plan on antibiotic resistance, stating that: "*Consistent and high-quality data on the incidence, prevalence, range across pathogens and geographical patterns related to AMR are needed to guide the treatment of patients; to inform local, national and regional actions; and to monitor the effectiveness of interventions" [10].*

AMR surveillance is complex and includes many components: host (human or animal), other contributory factors (food or environment), the pathogen being monitored, characterisation of resistance, type of monitoring (passive, enhanced), AST methodology and criteria for determining presence of AMR. Global surveillance programmes that monitor resistance in specific bacterial pathogens, such as *Mycobacterium tuberculosis*, have been in place for many years [11].

However, there are a number of challenges that have hampered global level surveillance efforts, including a lack of common standards for methods, data-sharing and coordination at local, national, regional and global levels. There are numerous AMR surveillance activities worldwide and initiatives have been in place for some time to develop standards, but these are restricted to certain types of AMR surveillance [12], and the overall standardisation of global surveillance processes are in the early stages [13]. Apart from the WHO GLASS programme no global forum currently exists for rapid sharing of standardised information on AMR. These gaps are hampering efforts to produce meaningful data at a global level to enable comprehensive monitoring and analysis of the occurrence and trends of resistance worldwide [14].

In this chapter we examine current global AMR surveillance initiatives, outline considerations and challenges for AMR surveillance programmes, before exploring how sequencing is contributing to these efforts.

2.1 National Action Plans for AMR

In 2015, the WHO adopted the Global Action Plan on Antimicrobial Resistance (GAP-AMR), which was drafted with the Food and Agriculture Organization of the United Nations (FAO) and the World Organization for Animal Health (OIE) (known as 'the Tripartite').

The five objectives outlined in the plan are:

- improve awareness and understanding of AMR through effective communication, education and training
- strengthen the knowledge and evidence base through surveillance and research
- reduce the incidence of infection through effective sanitation, hygiene and infection prevention measures
- optimise the use of antimicrobial medicines in human and animal health and
- develop the economic case for sustainable investment that takes account of the needs of all countries and to increase investment in new medicines, diagnostic tools, vaccines and other interventions.

The Plan included the goal that all Member States would have, by 2017, a national action plan (NAP) aligned with the GAP objectives. Therefore, for the majority of countries, AMR surveillance is coordinated within the context of their NAP. By 2020, more than 120 member states had developed NAPs, and a monitoring regime based on self-reporting has been established [15]. There is also a library on the WHO website of all approved NAPs [16].

To monitor national progress, the Tripartite relies on self-reporting by member states on AMR-related policies made available in the Global Database for Antimicrobial Resistance Country Self-Assessment. The annual Tripartite AMR country self-assessment survey (TrACSS) is a component of a broader approach for monitoring and evaluating GAP-AMR. The latest survey had a total of 136 (70.1%) out of 194 WHO Member States who responded to the 2019–2020 TrACSS [17]. Answers to the latest survey are publicly available on a website [18]. There is no specific mention of laboratory techniques used to determine AMR, and no mention of sequencing in the summary report or survey.

A cross-country analysis of 59 NAPs provided evidence of the variation in the content of the NAPs within and across regions and income groups [15]. It found that the NAPs included in their analysis were mostly aligned with the GAP's five overarching objectives and only moderately aligned with the recommended corresponding actions. Whilst the development of NAPs has been an important step towards AMR control there is still much that needs to be done. The analysis across these NAPs indicate that strengthening the regional governance regime as a mediating level between global governance (the Tripartite and GAP) and local delivery (national actors and NAPs), while preferable to uncoordinated national initiatives, will not solve the issue of limited global concerted action [15]. The authors of the analysis believe that the ideal role of WHO regional offices will be to support the development of the global governance regime, and to increase coordination across regions for implementing the GAP to avoid further fragmentation of effort [15].

2.2 Global AMR surveillance

Many international, national and local approaches are being employed for the management and surveillance of AMR, however to date many of these efforts are inconsistent and as a result global surveillance of AMR is fragmented [18-22]. Trends in AMR have been described in national and regional surveillance reports as well as numerous single-centre and population-based surveys conducted throughout the world. However, the dynamic nature of these trends indicate that considerable monitoring and more comprehensive, well organised surveillance programmes are needed. Consistent surveillance is required through time and over a broad geographic area to more comprehensively understand the emergence of specific strains or species, discover and monitor changes in the antimicrobial susceptibility profile of organisms, and develop an understanding of regional, national, and global trends and distribution of AMR. The majority of AMR surveillance is data is derived from clinical sources.

Since the WHO's Global Antimicrobial Resistance and Use Surveillance System (GLASS) report in 2018, participation in GLASS has grown exponentially. In 2019 the GLASS reporting system had aggregated surveillance data from more than 64,000 surveillance sites

with more than 2 million patients enrolled from 66 countries across the world. The 2021 GLASS report included 106,602 laboratory confirmed infections reports by 24,803 surveillance sites in 70 countries [23]. The drop in number of sites is due to 40,000 outpatient sites from the Americas no longer being listed, possibly due to the impact of the COVID-19 pandemic.

Key publications that cover information on AMR surveillance systems and networks for specific global regions include:

Mapping of different AMR surveillance systems in **Europe:**

- 24 national and 14 regional AMR surveillance systems in 19 European countries identified [20].
- The authors noted that although the European Antimicrobial Resistance Surveillance Network (EARS-Net) provides annual reports on monitored resistant bacteria, national surveillance efforts are still fragmented and heterogeneous, and have substantial structural problems and issues with laboratory data. Most incidence and prevalence data cannot be linked with relevant epidemiological, clinical, or outcome data.

Supranational networks in low- and middle-income countries [19]:

- Identified that there had been 45 surveillance systems implemented since 2000, of which 22 were still active at the time the review (2018).
- Surveillance networks have a positive impact by connecting laboratories in different countries. The Antibiotic Resistance in the Mediterranean Region (ARMed) network, reported improvement in participating laboratories' capacity to perform bacterial identification and AST, as a result of the EQA programme provided to the network [24]. The HIV, mycobacteria, influenza and gonorrhoea reference laboratory networks have been created thanks to global surveillance programmes.

Worldwide surveillance systems [18]. The most recently published review on global surveillance (2020) identified 71 surveillance systems. A summary of the results is presented in Table 1. One limitation is that the review only used papers from academic database (PubMed) and did not include grey literature such as surveillance reports which could support identification of additional surveillance systems. This review found:

- A total of 71 AMR surveillance systems from 35 countries were described, of which 64 (90.1%) were national surveillance systems and 7 (9.9%) were multinational (Supplementary Table S1 in the review). Two regions accounted for ~72% of systems: European region (37; 52.1%) and Region of the Americas (14; 19.7%). Other regions were Western Pacific region (12; 16.9%), African region (3; 4.2%), South East Asia region (3; 4.2%) and Eastern Mediterranean region (2; 2.8%).
- Of the 71 surveillance systems, 53 (74.6%) were exclusively using isolates from humans, 12 (16.9%) targeted isolates from both humans and animals and six (8.5%) focused on the surveillance of AMR in animals (Table 1). The latter six surveillance systems monitored bacteria of zoonotic origin, including *Campylobacter spp*., *Salmonella spp*. and commensal bacteria (*E. coli*) according to EU legislation on monitoring and reporting of AMR in zoonotic and commensal bacteria (2013/652/EU).
- 26 of 71 (36.6%) surveillance systems were considered up to date, whereas 45 of 71 (63.4%) were not. For 19 of 71 (26.8%) monitoring systems no report was found, and 26 of 71 (36.6%) were monitoring systems with at least one report found; 25 of 26 (96.2%) had published at least one report in the past five years. Of the 26 up-to-date systems, 3 (11.5%) of them are real-time monitoring systems and have an alarm detection system for critical phenotypes, namely Marseille Antibiotic Resistance Surveillance System (MARSS), EPIdemiological Surveillance and Alert Based on MICrobiological Data (EPIMIC) and Swedish Surveillance of Antimicrobial

Resistance (SVEBAR). Nine of 71 (12.7%) systems have an interactive database in which one could collect information on the percentage of resistance for specific antibiotics and/or phenotypes for a given period. Thirty of 52 (57.7%) reports are written in English, 14 of 52 (26.9%) in a local native language and 8 of 52 (15.4%) both in English and local native language. Information on frequency of the reports and if they reported to global databases was not provided.

- For the GLASS pathogens: 24 countries and four regional systems were covered by 49 surveillance systems (Table 1, Appendix 8.2). The regional systems include EASR-NET (EEA/EU, ECDC), EURO-GASP (EU), ARMed (Mediterranean region) and GLASS. Two of the surveillance systems – ANRESIS (Switzerland) and ARSP (Phillippines) – cover all nine GLASS pathogens. Two surveillance systems – NARST (Thailand) and ISKRA (Croatia) – cover seven pathogens. 11 of the 49 systems cover six GLASS pathogens.
- For the GLASS pathogens, the most commonly represented in the 49 surveillance programmes were *Staphylococcus spp.* (35/49 systems), followed by *E. coli* (31/49), *Streptococcus* (28/49), *Klebsiella spp*. (24/49), *Pseudomonas spp*. (20/49), *Acinetobacter spp.* (19/49), *Salmonella spp.* (17/49), *Neisseria spp*. (16/49) and finally *Shigella spp*. (7/49). For *Neisseria,* of the 16 surveillance systems that monitor this pathogen seven focus on this pathogen only and none of the other GLASS pathogens.

Table 1. Surveillance systems and their location that were identified as part of a single global literature review using PubMed and did not include grey literature such as surveillance reports. Only systems that included at least one of the nine GLASS pathogens of interest are included here (the eight main pathogens plus *Pseudomonas aeruginosa*, which is planned to be included). Adapted from [18].

			PATHOGEN									
Surveillance system	**Country**	**Surveillance of humans and / or animals**	***Acinetobacter spp.***	***Escherichia coli***	***Klebsiella spp***	***Neisseria spp.***	***Salmonella spp.***	***Shigella spp.***	***Staphylococcus spp***	***Streptococcus spp***	***Pseudomonas spp.***	**Number of GLASS pathogens /species covered**
ARSP	Philippine	Human	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
ANRESIS	Switzerland	Human and animal	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
GLASS	WHO - global	Human	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	8
ISKRA	Croatia	Human	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	7
NARST	Thailand	Human	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	7
CHINET	China	Human	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	6
EARS-NET	EU/EEA (ECDC)	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
ONERBA	France	Human and animal	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
MARSS	France - local	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
EPIMIC	France - local	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
SARI	Germany	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
ARS	Germany	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
ISIS-AR	Netherlands	Human and animal	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
GERMS-SA	South Africa	Human	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	6
KONIS	South Korea	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
BSAC (HISC/HPA)	United Kingdom (regional)	Human	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	6
CARAIERT	Australia	Human	No	No	No	Yes	Yes	Yes	Yes	Yes	No	5
CARSS	Canada	Human and animal	No	Yes	No	Yes	Yes	No	Yes	Yes	No	5
KISS	Germany	Human	Yes	Yes	Yes	No	No	No	Yes	No	Yes	5
GSSAR	Greece	Human	Yes	Yes	Yes	No	No	No	Yes	No	Yes	5
LABBASE2 (PHE)	United Kingdom (regional)	Human and animal	No	Yes	Yes	No	No	No	Yes	Yes	Yes	5
DANMAP	Denmark	Human and animal	No	Yes	No	No	Yes	No	Yes	Yes	No	4
FINRES-VET	Finland	Human and animal	No	Yes	No	No	Yes	No	Yes	Yes	No	4
JANIS	Japan	Human	No	Yes	Yes	No	No	No	Yes	No	Yes	4
JVARM	Japan	Animal	No	No	Yes	No	Yes	No	Yes	Yes	No	4
NORM	Norway	Human	No	Yes	Yes	No	Yes	No	Yes	No	No	4
KARMS	South Korea	Human	Yes	Yes	Yes	No	No	No	No	No	Yes	4
AURA	Australia	Human	Yes	No	No	Yes	No	No	Yes	No	Yes	3
BMR-RAISIN	France	Human	No	Yes	Yes	No	No	No	Yes	No	No	3
ITAVARM	Italy	Animal	No	Yes	No	No	Yes	No	Yes	No	No	3
AR-ISS	Italy	Human	No	No	Yes	No	No	No	Yes	Yes	No	3

ARMED	Mediterranean region (transnational)	Human	No	Yes	No	No	No	No	Yes	Yes	No	3
ARMOR	USA	Human	No	No	No	No	No	No	Yes	Yes	Yes	3
NARMS	USA	Human and animal	No	Yes	No	No	Yes	Yes	No	No	No	3
WHONET	Argentina	Human	No	No	No	No	No	No	Yes	Yes	No	2
CIPARS	Canada	Human and animal	No	Yes	No	No	Yes	No	No	No	No	2
MIB	Italy	Human	No	No	No	Yes	No	No	No	Yes	No	2
MARAN	Netherlands	Human	No	Yes	No	No	Yes	No	No	No	No	2
NTSS	USA	Human	No	No	No	Yes	No	No	No	Yes	No	2
VICNISS	Australia - regional	Human	No	No	No	No	No	No	Yes	No	No	1
BULSTAR	Bulgaria	Human	No	No	No	Yes	No	No	No	No	No	1
CNISP	Canada	Human	No	No	No	No	No	No	Yes	No	No	1
EURO-GASP	EU	Human	No	No	No	Yes	No	No	No	No	No	1
FIRE	Finland	Human	No	No	No	Yes	No	No	No	No	No	1
ARMIN	Germany	Human	No	No	No	Yes	No	No	No	No	No	1
SNARS	Slovakia	Human	No	No	No	Yes	No	No	No	No	No	1
SVEBAR	Sweden	Human and animal	No	No	No	Yes	No	No	No	No	No	1
CA-MRSA	Switzerland	Human	No	No	No	No	No	No	Yes	No	No	1
GISP	USA	Human	No	No	No	Yes	No	No	No	No	No	1

Key. Yes: species or genera monitored (genus listed where several species of the same genus are included); No and Red: not monitored by the systems.

Abbreviations for the surveillance systems listed in this table are in Appendix 8.2.

To understand the AMR surveillance situtation in **Africa** a useful resource is a recent situational assessment for the Africa CDC on disease surveillance, emergency preparedness, and outbreak response in Eastern and Southern Africa [21]. Five AMR surveillance networks and an additional four research, capacity building and other support structures for AMR surveillance were identifed. Although only one is focused on the GLASS pathogens it does demonstrate that resources and skills relevant to AMR surveillance are present on the African continent. The surveillance systems are focused on:

- Malaria (HANMAT [25])
- Pneumococcal Disease (netSPEAR)
- Typhoid
- Influenza (GISRS [26])

Lastly, Pfizer's Antimicrobial Testing Leadership and Surveillance (ATLAS) programme monitors the resistance of pathogens across more than 73 countries and shares data on AMR Register, an open-access data platform created by the Open Data Institute and Wellcome Trust. The report also indicates there is laboratory testing for AMR in the eight GLASS priority pathogens as well as *Pseudomonas aeruginosa*. In particular Kenya and Uganda have laboratory testing for these nine pathogens (Table 2).

Table 2. Laboratory testing abilities for AMR in 10 South and East African countries for 14 priority pathogens. The number of GLASS pathogens covered by each laboratory is indicated (eight main pathogens plus *Pseudomonas aeruginosa*, which is planned to be included; *indicates GLASS pathogens). Adapted from [21].

	PATHOGEN														
Country	*Acitenobacter spp.* *	*E. coli* *	*K. pneumoniae* *	*N. gonorrhoeae* *	*Salmonella spp.* *	*Shigella spp.* *	*S. aureus* *	*S. pneumoniae* *	*P. aeruginosa* *	*Vibrio cholerae*	*H. influenzae*	*Enterococci*	*Typhi*	*M. tuberculosis*	**Number of GLASS pathogens /species covered**
Kenya	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
Uganda	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
South Africa	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	8
Zambia	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	Yes	8
Ethiopia	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	Yes	No	No	7
Malawi	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	No	No	No	No	6
Zimbabwe	No	Yes	No	No	Yes	No	No	No	No	No	No	No	No	No	2
Mozambique	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	0
Somalia	No	No	No	No	No	No	No	No	No	No	No	No	No	No	0
South Sudan	No	No	No	No	No	No	No	No	No	No	No	No	No	No	0

For the **South and South East Asian** region a recent 2020 review looked at AMR Surveillance in eight South Asian and Southeast Asian countries [27]. Across the eight countries the number of surveillance sites were:

- Bangladesh (16 surveillance sites)
- Cambodia (8)
- India (55 hospitals)
- Laos (Not established)
- Nepal (42)
- Pakistan (9)
- Thailand (74)
- Vietnam (16 hospitals).

It is unclear within the 2020 review which pathogens are covered by surveillance within each country. The exception is Nepal, which has surveillance and monitoring of ten pathogens*: Vibrio cholerae, Shigella spp., S. pneumoniae, H. influenzae, Neisseria gonorrhoeae, Salmonella spp. and E. coli, Acinetobacter spp., Klebsiella spp.,* and *S. aureus*. All follow CLSI AST guidelines, although it is unclear whether this applies in Laos. Three countries (Bangladesh, Cambodia, Laos) do not report to GLASS, Vietnam is not enrolled, and the remainder only report from some sites [27].

2.2.1 AMR surveillance supported by industry

The AMR Benchmark, a research programme run by the Access to Medicine Foundation [28], compares the surveillance programmes of large research-based pharmaceutical companies and reports on whether they are active in the surveillance of bacterial or fungal pathogens and/or infections anywhere in the world and whether the results are shared publicly. Companies can fund external programmes run by established institutions, and they can run their own surveillance programmes. Data sharing is varied, with some requiring a data access agreement, others release aggregated results via publication, and one (Pfizer) releases raw data via an open access platform, ATLAS [29]. Registration is required to access the database. Surveillance programmes in which pharmaceutical companies are active and involve the pathogens covered in our report are listed in Table 3 [28]. Improved public-private collaborations and open data sharing are essential for enabling greater utilisation of private sector data [30]. Industry networks produce high-quality data, but they may not be representative and these networks do not usually support laboratory capacity building in LMICs or influence policy and guidelines [19].

Table 3. Surveillance programmes in which pharmaceutical companies are active, that involve the nine GLASS pathogens (eight main pathogens plus *Pseudomonas aeruginosa*, which is planned to be included). Adapted from [28].

			Pathogen						
AMR Surveillance programme that have companies involved	**Companies**	**Countries covered**	***Acinetobacter spp.***	***Neisseria spp.***	***Salmonella spp.***	***Staphylococcus aureus***	***Streptococcus pneumonia***	***Pseudomonas aeruginosa spp.***	**Number of GLASS pathogens/ species covered**
CANWARD	Abbott; MSD; Pfizer	1 (Canada)	Yes	Yes	Yes	Yes	Yes	Yes	6
Antimicrobial Testing Leadership and Surveillance (ATLAS)	Pfizer	81	Yes	Yes	No	Yes	Yes	Yes	5
SENTRY	Cipla; Pfizer: Shionogi	57	Yes	No	No	Yes	Yes	Yes	4
Study for monitoring antimicrobial resistance trends (SMART)	MSD	63	Yes	No	No	Yes	Yes	Yes	4
Study of bacterial resistance, Kinki region of Japan	Shionogi	1 (Japan)	Yes	No	No	Yes	Yes	Yes	4
BSAC Bacterial resistance surveillance programme	MSD; Pfizer	2	No	No	No	Yes	Yes	Yes	3
Data Development	Viatris	1	No	Yes	No	Yes	No	Yes	3
Abbott Restora Infectious Disease Scientific Excellence (ARISE)	Abbott	1	No	No	No	Yes	No	Yes	2
SIDERO-WT Programme	Shionogi	1 (Japan)	Yes	No	No	No	No	Yes	2
National reference centre for the diagnostics of central nervous system bacterial infections (KOROUN)	Novartis	1	No	No	No	No	Yes	No	1
Observations regionaux du pneumocoque (ORP)	Sanofi	1	No	No	No	No	Yes	No	1
Program to access ceftolozane/tazobactam susceptibility (PACTS)	MSD	29	No	No	No	No	No	Yes	1
Shionogi Japanese surveillance studies Programme	Shionogi	1 (Japan)	No	No	No	Yes	No	No	1
Surveillance of Tedizolid Activity and resistance (STAR)	MSD	14	No	No	No	Yes	No	No	1
Survey of antibiotic resistance (SOAR)	GSK	43	No	No	No	No	Yes	No	1

2.3 Considerations for surveillance

The data collected by the various AMR surveillance systems have the advantage of providing information on the actual burden of resistance at local, national and international levels. However, the reviews outlined above consistently found that, overall, surveillance systems are highly heterogenous and fragmented. Some systems will use incidence of AMR for reporting whereas others use prevalence of AMR. The frequency of reporting can vary from real time, monthly, quarterly, yearly or mixed, for example in Europe it was found that 3% of surveillance systems released data in real-time [20].

AMR surveillance is complex with different surveillance systems having different objectives and data collection tailored to each objective. There are many components involved from the application of the surveillance systems, AST, standards and criteria for AMR definitions. Below we will consider some broader aspects of that would enable successful AMR surveillance and that would impact on the implementation of surveillance using sequencing.

2.3.1 Current design and availability of surveillance settings

Laboratory-based surveillance systems have limitations [20]. The microbiological results reported usually have no associated relevant epidemiological, clinical, or outcome data. Thus, these systems provide no information on the identification of at-risk patient populations, types of infections, sources (i.e., community-onset, health-care-associated, or hospital-acquired infections), treatment failure, or real burden of disease associated with health-care-associated infections and antimicrobial resistance [20].

Animals and Food. AMR surveillance in animals can refer to a wide range of settings e.g. in the wild, veterinary settings in infected animals, and in animals intended for food consumption (monitoring of healthy animals to meat products). The latter constitutes a large part of food surveillance, which also includes surveillance of non-meat products and surveillance and investigation of foodborne infections. Systems for routine surveillance in each of these settings are typically still in their infancy, with the exception of humans diagnosed with foodborne infections, which are present in many countries [19, 20]. Some countries also have well established systems for AMR surveillance in animals intended for food consumption, especially of zoonotic pathogens. In Europe, it is mandatory for member states to monitor AMR based on certain indicator bacteria in healthy animals and food as well as zoonotic infections in humans [31]. However there is still progress to be made; for example, as results from food product surveillance are considered commercially sensitive they are rarely released publicly by regulators [20], meaning a complete picture of AMR within regions cannot be developed.

In areas of veterinary medicine, wildlife, and plant-based food, surveillance systems are much more limited, although efforts are underway to address this. For veterinary surveillance, it was recently proposed that the EARS-VET (European Antimicrobial Resistance Surveillance) network be established as a veterinary equivalent to EARS-NET(for AMR surveillance of clinical infections in animals [31]. Progress towards standardising AST breakpoints (the concentration of a drug that defines whether a pathogen is susceptible or resistant to a drug) in veterinary microbiology is behind that in human microbiology [19]. The need for surveillance in crops has been recognised by institutions such as the FAO [16], whilst the need for wildlife surveillance has been highlighted through research studies and One Health strategies [32] although currently no systems appear to routinely monitor AMR in wildlife.

One Health. The One Health approach recommends monitoring AMR in humans, animals and the environment with multi-sectoral teams and international collaboration between different AMR surveillance systems (or networks) [18]. The ongoing impact on the food chain and food monitoring confirms the need for greater integration between human and animal surveillance systems [33]. Very few current surveillance systems take into account the

monitoring of AMR in the environment, including the detection of AMR in animals. There is therefore a need for better alignment of human, veterinary and food surveillance systems to implement a One Health approach to monitor AMR. Some strategies have been piloted in LMICs under the guidance of the WHO Advisory Group on Integrated Surveillance of Antimicrobial Resistance (AGISAR) [19]. There is also one supranational European network, the European Food- and Waterborne Diseases and Zoonoses Network (FWD-Net) [34]. It does surveillance of food- and waterborne diseases and zoonoses collecting data on antimicrobial susceptibility in humans, animals and food [19].

2.3.2 Surveillance operations

Local relevance. Inadequate and delayed reporting of surveillance data can lead to suboptimal antimicrobial prescribing that has an impact on clinical outcomes, increase risk of ongoing transmission and drive further AMR development [20]. Many surveillance systems operate as an upwards reporting chain, which provides limited generation of locally usable results, being even more limited for clinical use and patient-centred surveillance. One way to overcome this is to promote national and institutional ownership of surveillance activity and frame it as being an important routine activity [19]. The benefits of collaboration between policymakers, academics and service providers has been previously highlighted, a sentiment echoed by the experience of malaria networks, where multi-stakeholder collaboration re-energised surveillance and also played a role in advocacy for policy change, acting as a bridge between research groups and national control programmes [19].

Representativeness and coverage of monitoring. This will be affected by the geographical location and number of sentinel sites, the number and characteristics of individuals sampled, prior treatment history, the incidence of the target pathogen and the methods of detection [19]. In LMICs it is highly likely that sites involved in surveillance are primarily tertiary care hospitals or regional hospitals; with secondary care and primary care centres poorly represented [27]. In these countries, the majority of tertiary care hospitals are national referral centres and cater to patients from different regions, without specific population catchment areas [27]. Coverage by the AMR surveillance systems differ for the different laboratories and hospitals at a local, national and global level. Some may only focus on a small number of hospitals and apply a weighting or ratio when reporting to be able to generalise the burden of disease in a larger area of country. Achieving high coverage is also a challenge in LMICs [19]. Differences in the frequency and distribution of sampling among physicians, institutions, and countries, and the inclusion of screening isolates instead of the inclusion of only clinical isolates, undermine how representative the data are. As a One Health approach is established ensuring appropriate coverage and representativeness of must be considered including environmental testing.

Identifying and ensuring appropriate infrastructure for surveillance needs. This includes laboratory infrastructure, staff capacity and training, availability of consumables, diagnostics and reagents, quality assurance, and funding. There are decision support tools which have been developed for some surveillance applications to help countries assess their surveillance needs and current capacity, and identify areas for development, for example those provided by the Food and Agricultural Organisation of the United Nations (FAO) for AMR surveillance in the food and agricultural sector [16]. Options can be suggested based on the circumstances of the country, it is likely that LMICs and HICs will have different requirements.

Parameters for surveillance applications. Some newer areas of surveillance are slow to develop due to a lack knowledge of which indicators need to be monitored to meet AMR public health needs, for example in environmental surveillance strategies. Coordinated efforts to determine these indicators will enable identification and testing of the most suitable approaches for monitoring AMR whether these be phenotypic, molecular, or NGS based. For example, the JPIAMR funded project *Towards Developing an International Environmental AMR Surveillance Strategy* aims to provide information for environmental

surveillance, and will assess the utility of all types of tools, both phenotypic and targeted, metagenomic and WGS sequencing technologies [17].

2.3.3 Sampling (or AMR susceptibility testing)

Sample source is a crucial aspect in estimating the burden of AMR [18]. While some surveillance systems will consider data on AMR collected from all types of clinical samples, others will only consider a specific subset of clinical samples e.g. blood. Considering information from as wide a range of clinical samples as possible provides details on AMR carriage and allows evaluation of the relationship between AMR carriage and infection [18]. In some settings, sample collection for further AMR analysis is considered best practice only for the more severe infections or those not responding to first-line treatment. In these cases, rates of AMR might be inflated, and use of these data could lead to an inappropriate therapy selection and increase resistance and health-care costs [20]. Conversely, under-reporting of healthcare-associated infections and AMR might occur if samples are not routinely collected, and reliance on laboratory-based surveillance underestimates the incidence of clinically relevant healthcare-associated infections [20]. Additional considerations of sample sources are needed for other setting such as environmental, wastewater or veterinary. Careful documentation of isolate sources should be emphasised, particularly as One Health surveillance approaches are implemented.

Susceptibility testing methods. What susceptibility testing methods are being used for national surveillance programmes is often not clear, however it can be assumed that the majority of susceptibility testing is currently being done using traditional phenotyping and serotyping methodologies, which may not be standardised across countries and laboratories.

Quality Management. Currently GLASS recommends national reference laboratories take responsibility for quality management of laboratory tests. Selection of an appropriate system and consistency across them can improve reporting of AMR to surveillance systems. Pharma-led networks typically do not involve LMIC laboratories in EQA programmes and send all isolates to a central laboratory for confirmatory testing [19]. Global surveillance programmes for AMR in TB, HIV, influenza and gonorrhoea all had proficiency testing programmes delivered via supranational networks of reference laboratories [19]. Among the networks for AMR surveillance in bacteria, the Latin-American network, Red Latinoamericana de Vigilancia de la Resistencia a los Antimicrobianos (ReLAVRA) has been running an EQA scheme (LA-EQAS) since 2000 and provides proficiency testing services at no cost to participating laboratories. The Central Asian and Eastern European Surveillance of Antimicrobial Resistance (CAESAR), the non-EU European network, have used the UK National External Quality Assessment Service (UK-NEQAS) for EQA. WHO-sponsored EQA efforts for AST included the discontinued WHO EQAS AST (1998–2001) and the WHO-AFRO/NICD-SA EQAP for countries within the WHO-AFRO region.

2.3.4 Standardisation

Inadequate standards. A very common observation is that there is a lack of harmonised standards for the reporting of resistance in surveillance programmes [18-21]. Some standards for AMR surveillance do exist but are inadequately implemented or followed [12, 13, 35]. Standardisation of epidemiological definitions, samples and data collected, culture media used, microbiological testing methods, criteria for characterisation and publication of reports years after data collection cause challenges with consolidating data from these varied networks. Data sharing policies can also be obstacles to reliable and informative collaborative surveillance nationally and internationally [20] and there are also no recognised standards for the composition and activities of AMR surveillance networks [19].

Different definitions for the same phenotype. For example, some AMR surveillance systems use cefoxitin, oxacillin and/or flucloxacillin to define MRSA [18]. The Dutch national

antibiotic resistance surveillance system (ISIS-AR) includes cefoxitin results to define MRSA and, if this antibiotic is not available, oxacillin and/or flucloxacillin is used. The Canadian Antibiotic Resistance Surveillance System (CARSS), however, uses methicillin, oxacillin and cefazolin to define this same phenotype.

Different criteria for interpretation. This can make it difficult to compare findings across surveillance systems [18]. For example in the US, NARMS uses the CLSI interpretation criteria, and in Europe EARS-Net uses the European Committee on Antimicrobial Susceptibility Testing (EUCAST) criteria. It has been shown that a change in interpretation criteria from CLSI to EUCAST increases the number of strains classified as MDR including *K. pneumoniae* (by 2.2%), *Enterobacter cloacae* (1.1%), *P. aeruginosa* (0.7%) and *E. coli* (0.4%) [36]. For some bacteria–antibiotic pairs, such as *E. coli*–ciprofloxacin and *K. pneumoniae*–ciprofloxacin, the agreement between CLSI and EUCAST was 77.8% and 61.5%, respectively [37].

2.3.5 Data

Data quality. An integral component of AMR surveillance is the use of technology/databases for collection, storage and analysis of surveillance data. The accuracy of AMR burden estimates depends on the quality and availability of this input data. Data are often self-reported, heterogeneous, and based on few isolates from a handful of surveillance sites or projects. The quality of data generated from AMR surveillance networks is dependent on laboratory practices (use of internal and external quality assurance and control, quality management systems, and accreditation), clinical sampling methodology, and consistent use of microbiology laboratories for infectious disease diagnostics (it is ideal to obtain cultures prior to the administration of antimicrobial therapy, but in LMICs, it is common practice to utilise diagnostic microbiology services only after patients fail to improve on broad-spectrum antibiotic therapy, a practice that could inflate AMR rates) [27]. Practices that influence AMR surveillance data quality include reporting on key pathogen-drug combinations, defining multi-drug resistance, the inclusion of appropriate specimens, and reporting clinically inappropriate pathogen-drug combinations. Variability in these areas results in difficulties in data interpretation and comparison [27].

Metadata collection. Within healthcare this includes patient clinical data which is unavailable in most AMR surveillance systems [18]. An ideal AMR surveillance system would include both clinical and laboratory data, although many surveillance systems remain predominantly laboratory focused. This is partly due to regulatory provisions of countries such as in Europe where privacy laws are an important consideration. Additionally, to obtain clinical information it is useful to have a unique identifier for each patient, which can be difficult for countries that do not have a medical information system in place.

Data sharing and access within and between countries. This includes ensuring that all sample sources within a country are considered for a surveillance strategy, and all data is uploaded to the same database within a country and preferably internationally. This data should then be made accessible for global research and surveillance efforts. If there are multiple producers of data e.g. national laboratories, private laboratories, or direct from sample/benchtop sequencing then consideration should be given on how to coordinate this data. In some cases having a centralised structure responsible for processing and uploading all data sharing may be the most feasible option [38].

Collection of relevant explanatory data for AMR trends that can be applied globally. Surveillance data is not useful unless it is used to inform and monitor interventions to reduce the burden of AMR, for example by reducing use of antimicrobials, strengthening hygiene practices, or interrupting transmission chains. There is a need to identify indicators needed to explain and interpret AMR sequencing data, and collect data on them for surveillance purposes. These could include levels of AMU either reported or by direct measurement, indicators of sanitation, indicators of population health and total numbers of AMR infections

etc. These indicators need to be defined for each situation and measured in a systematic way that can be applied globally, to enable global comparison of AMR.

2.4 Implementation of surveillance programmes

The considerations mentioned above outline the numerous challenges affecting the implementation of effective and sustainable AMR surveillance systems. These challenges have been described in further detail elsewhere [39-49]. Access to effective antimicrobials is an essential component of all health systems, however the situation is complex in terms of expanding access while avoiding excess use, which can be hampered by minimal laboratory diagnostic capacity, leading to inappropriate use and risk of treatment failure [5]. Surveillance is only one element in this complex system. Limitations of the surveillance systems are related to a lack of international co-ordination, inadequate standardisation and publication of results years after data collection [20]. Tacconelli et al have highlighted key points on surveillance of AMR in Europe that are relevant for all surveillance programmes, in summary these are [20]:

Short-term priorities:

- Discussion between a range of stakeholders from e.g. academia, health systems, pharmaceutical industry, to agree on surveillance goals, and definition and measures of antimicrobial resistance
- Development of data sharing policy(-ies) encouraging and enabling surveillance systems to provide barrier-free and timely access to key national data on antimicrobial resistance

Long-term priorities:

- Creation of collaborative platform(s) to optimise surveillance efforts involving a range of stakeholders
- Political support for investment in surveillance
- Implementation of surveillance systems that link clinical, epidemiological, radiological, and microbiological data
- Further development of AMR surveillance systems in animals and the food chain
- One Health approach to surveillance systems

A successful AMR surveillance network should generate up-to-date comparable, representative, high-quality data on pathogens of concern from the target population(s). It should be able to detect and track unexpected events including outbreaks in real time, have rapid, effective mechanisms for communication and reporting, and have a responsible data-sharing policy [19]. A network needs strong leadership and coordination, and it should influence guidelines and policy and ultimately impact on human and animal health. Challenges in current AMR surveillance approaches, programmes and networks will impact on the implementation of sequencing for this purpose, especially in countries with limited resources [50].

2.4.1 Moving forward

There are protocols and programmes that aim to improve national and transnational surveillance efforts. Examples include (summarised in Table 4):

Protocol for Enhanced Isolate-Level Antimicrobial Resistance Surveillance in the Americas Primary Phase: Bloodstream Infections which describes the steps and procedures to establish/enhance AMR surveillance in Latin America and the Caribbean [51]. It provides technical guidance to integrate patient, laboratory, and epidemiological data to monitor AMR emergence, trends, and effects in the population. It also provides the necessary elements to move from aggregated data to isolate-level data surveillance, starting with blood isolates. It facilitates uniform data collection processes, methods, and tools to

ensure data comparability within the region of the Americas. Finally, it builds on over a decade of experience of the Latin American AMR surveillance network—ReLAVRA by its Spanish acronym—and its procedures are aligned with the GLASS methodology, enabling countries to participate in the global GLASS AMR surveillance.

The **Tricycle protocol,** which takes a One Health approach [52] and is a standard protocol for integrated multisectoral surveillance. The model targets monitoring one indicator, the extended spectrum beta-lactamases- (ESBL) producing *Escherichia coli* across the human, animal and environmental sectors. This protocol includes standard methodologies in the human, food chain and environmental sectors to be implemented in low resource settings to facilitate the establishment of the integrated multisectoral surveillance on AMR. Countries can build from this approach a complete national surveillance system that involves other cross cutting pathogens, resistance mechanisms and expand the implementation in different cities and provinces in the country to get more evidence of the spread of the antimicrobial resistance in the different sectors and allow the implementation of interventions in a holistic way to contain AMR.

The Tricycle protocol has also been piloted in LMICs such as Indonesia, Malaysia, Pakistan, Bangladesh, Sri Lanka, Nepal, India, Madagascar and Ghana, as well as countries like Jordan [53, 54]. The “Network for Enhancing Tricycle ESBL Surveillance Efficiency” (NETESE) involves 15 institutions from nine LMICs at different stages of implementation of Tricycle and three EU countries that have been instrumental in its development. The countries involved include: France, Belgium, the Netherlands, Pakistan, Indonesia, Malaysia, Madagascar, Senegal, Cameroon, Burkina Faso, Democratic Republic of the Congo, and Ivory coast [55]. A publication in 2021 noted that they are expected to shortly publish an optimised protocol [56].

Africa CDC’s Antimicrobial Resistance Programme, which plans to establish cross-border networks for strengthening disease control at the continental level [21]. The AMR programme is a cross-border network that will enable health institutions and experts to coordinate AMR surveillance and control activities in Africa while also providing a platform for high-level policy engagement. Hosted by the African Union, AMR Programme members will include national public health institutes (NPHIs), and the platform’s activities will be implemented by the regional collaborating centres in collaboration with NPHIs and other organizations. The AMR Programme will link GLASS and other initiatives, national health ministries, the AU’s Inter-African Bureau for Animal Resources (AU-IBAR), the AU’s Pan-African Veterinary Vaccine Centre (AU-PANVAC), the UN Food and Agriculture Organization FAO), and other development partners and NGOs.

WHO Monitoring and evaluation of the global action plan on antimicrobial resistance: framework and recommended indicators, which is an initiative to establish a richer dataset on AMR policies than what is currently being collected through the GLASS surveys; it has been projected to take five to ten years to compile [57]. This new framework of monitoring and evaluation will remain dynamic so that it will adapt to changes in our understanding of AMR, new techniques and technologies, such as molecular genetics, electronic patient records and big data analysis. As knowledge on AMR and related measures improve, and lessons emerge on what works in different countries and contexts, the indicators, and the framework itself, are likely to evolve substantially. Hence, the framework will be revised after a period to reflect the lessons learned from its implementation and to incorporate emerging evidence about AMR and any new tools or technology. We can anticipate this will include better information on laboratory methodologies being used to identify AMR.

The **Fleming Fund** brings evidence and people together to encourage action against drug resistance [58]. They support low- and middle-income countries to generate, share and use data to improve antimicrobial use and encourage investment in AMR. Such funding should

support surveillance infrastructure-building efforts that result in sustainable surveillance system.

The **Mérieux Foundation** is an independent family foundation with an interest in combating infectious diseases that affect LMICs, for example by building clinical laboratory capacity. They support numerous AMR initiatives including the AMR intensive course which builds capacity for decision-making in a range of areas, particularly in LMICs [59].

SEQAFRICA has objectives to develop, expand and support WGS and bioinformatics capacity for AMR surveillance across Africa [60]. Their centres provide WGS and analysis services to surrounding countries in West, East and Southern Africa and support investigations of outbreaks, unusual resistance phenotypes, and/or delineation of the flow of organisms/genes across human, animal, agricultural and aquaculture sectors. They have established and currently support a consortium of three regional sequencing centres in: Nigeria (University of Ibadan, UI), Tanzania (Kilimanjaro Clinical Research Institute, KCRI), and South Africa (National Institute for Communicable Diseases, NICD) as well as a national centre in Ghana (Noguchi Memorial Institute for Medical Research, NMIMR). SEQAFRICA has also developed extensive virtual training for a range of audiences going from novel and non-users to more experienced staff.

Tool development to further assist in the surveillance of AMR is also underway and includes:

- The Indian Council for Medical Research (ICMR) has developed the Antimicrobial Resistance Surveillance system (i-AMRSS), a promising tool for global antimicrobial resistance surveillance [61] developed in India which provides a standardised data collection tool. The main aim of the ICMR Antimicrobial Resistance Surveillance and Research Network (AMRSN) was to develop a hospital network to track the patterns in the antimicrobial susceptibility profile of medically significant human health-restricted bacteria and fungi. Since no previously available tools were able to meet the requirements of the network, the data management team at ICMR developed a web-based online AMR data entry system, named i-AMRSS for capturing storage and analysis of AMR data. It is currently being used to capture human testing and antibiotic consumption data, piloted in ICMR's AMR Network of 31 hospitals and laboratories across India since 2016. The developed tool has collected more than 280000 patient records to date. Some of these hospitals/centres are multispecialty sites with good laboratory and hospital information systems generating more than 500 patient records in a day. With a large patient burden and manual data entry, the data entry process has become a significant bottleneck for the data inflow. A platform-independent web interface capable of integrating data from multiple sources has been developed to mitigate these challenges. The designed app named *i-DIA* (Data Import App) has been integrated with i-AMRSS [62]. It can also be extended for AMR surveillance using a One Health approach.
- For low-resource settings, Médecins Sans Frontières has developed the Mini-Lab to provide simplified bacteriological diagnosis and AMR surveillance in challenging settings [47]. They are also addressing the difficulties of obtaining necessary AMR data in LMICs, as well as the role that stakeholders outside public medical systems can play in the collection of this information.

Table 4. Summary of example protocols and programmes (discussed above) that aim to improve national and transnational AMR surveillance efforts.

Protocol / Programme	Description	Countries / region
Protocol for Enhanced Isolate-Level AMR Surveillance in the Americas Primary Phase: Bloodstream Infections	This protocol describes the steps and procedures to establish/enhance AMR surveillance. Providing technical guidance to integrate patient, laboratory, and epidemiological data to monitor AMR emergence, trends, and effects in the population.	Latin America and the Caribbean
Tricycle protocol	WHO integrated global surveillance on extended spectrum beta-lactamases-(ESBL) producing Escherichia coli across the human, animal and environmental sectors. This protocol includes standard methodologies in the human, food chain and environmental sectors to be implemented in low resource settings.	Global. Piloted in: Indonesia, Malaysia, Pakistan, Bangladesh, Sri Lanka, Nepal, India, Madagascar and Ghana. Implementation through NETESE in: France, Belgium, the Netherlands, Pakistan, Indonesia, Malaysia, Madagascar, Senegal, Cameroon, Burkina Faso, Democratic Republic of the Congo, and Ivory coast
Africa CDC's Antimicrobial Resistance Programme	Disease surveillance, emergency preparedness, and outbreak response: A Situational Assessment and Five-Year Action Plan for the Africa CDC Strengthening Regional Public Health Institutions and Capacity for Surveillance and Response Program. 2021	Eastern and Southern Africa
WHO Monitoring and evaluation of the global action plan on antimicrobial resistance: framework and recommended indicators	The framework aims to be robust and practical – to provide a manageable system that can facilitate the generation, collection and analysis of standardized data to assess the success of the GAP, and inform operational and strategic decision-making on AMR for the next 5–10 years at the national and global levels.	Global
SEQAFRICA	SEQAFRICA's main objective is to develop, expand and support WGS and bioinformatics capacity for antimicrobial resistance (AMR) surveillance across Africa. SEQAFRICA has established and currently support a consortium of three regional sequencing centres and a national centre in Ghana.	Nigeria, Tanzania, South Africa, Ghana

2.5 Sequencing and surveillance

A number of NGS approaches can be used to obtain sequencing data for surveillance – WGS, targeted sequencing/panels, or metagenomics. WGS of cultured isolates is currently the most commonly used approach; since it provides the most comprehensive pathogen genome data, it is suited to a wide range of surveillance applications. Examples of targeted sequencing approaches and metagenomic methods for surveillance are usually found in the context of one-off research studies. Metagenomics based approaches have been proposed as a promising surveillance tool, since they can provide sequence data on multiple pathogens simultaneously without prior knowledge of targets. It is also a culture-free method of obtaining WGS data. Metagenomic WGS is still in the development stage and/or or requires further validation as a surveillance tool. For example the 2020 GLASS report on WGS excludes metagenomics and focusses on sequencing of pure cultures [4]. Most examples of sequencing for AMR surveillance described in this report focus on WGS, however many considerations are relevant to other NGS approaches. The specific utility of more future-facing targeted and metagenomics NGS approaches is highlighted in Sections 4.3 and 6.0.

Globally, there are moves towards incorporating the use of WGS in public health surveillance [38, 63-66], which would be of benefit for AMR surveillance. The more widespread implementation of NGS into an already complex and fragmented surveillance landscape, that uses a range of technologies, will require careful consideration and planning [67]. Currently, the use of NGS as a tool to support AMR surveillance is not a requirement or considered standard practice. However, there are efforts to change this, and GLASS is supporting this effort [4]. Others include:

- ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations, 2019 [68]. The ECDC has also published: *COVID-19 surveillance guidance - Transition from COVID-19 emergency surveillance to routine surveillance of respiratory pathogens* [69].
- NIHR Global Health Research Unit on Genomic Surveillance of AMR published *WGS as part of national and international surveillance programmes for AMR: a roadmap* [44]. Within this they include programmes such as the ‘train the trainer’ initiative [70].
- The Tricycle protocol – which includes WGS and provides an overall view of implementing a One Health strategy [52].

It is currently difficult to determine how frequently sequencing is or has been used for AMR surveillance as part of national surveillance programmes, due to a lack of data on laboratory methods used to collect AMR data. Where it has been identified that sequencing has been used for AMR surveillance, it is commonly in highly resourced settings that have well established functional national surveillance systems in place.

AMR surveillance sequencing data is frequently collected through specific studies, single-centre or population-based surveys that provide a ‘snapshot’ of a particular situation where AMR surveillance was done for a targeted purpose. These efforts are primarily projects funded for a short time period, focused on a specific time point, disease, gene, drug, population or region. Whilst studies like this provide valuable information, they tend to be lacking in terms of longitudinal data that allows for monitoring of AMR.

One example illustrating the low usage of sequencing is demonstrated in a global landscape review of serotype-specific surveillance of invasive pneumococcal disease (*S. pneumoniae*) across 75 surveillance sites [71]. The sites reported using a range of serotyping methods, however of these only seven reported using ‘other serotyping methods’, which included ‘NGS and WGS’. Six of these sites were in Europe and one in the US. This study indicates

how few laboratories use sequencing for surveillance, for one of the GLASS priority pathogens, and that use is restricted to laboratories in HICs.

Additionally, it appears a large proportion of the work is being done outside of national surveillance programmes. For example, a scoping review designed to determine the use of WGS in the surveillance of *Enterococcus* spp. found that many hospitals and laboratories have surveillance systems in place for specific organisms independent of national surveillance programmes [72]. In addition, just under half of the 72 studies identified were not associated with a specified surveillance group and could have been one-off studies. The remaining articles were associated with government funded programmes, within hospital screening or surveillance programmes, or private/industry funded surveillance programmes. This demonstrates the fragmented nature of surveillance data, particularly when looking at sequencing efforts.

2.6 National surveillance systems using sequencing

Sequencing, in particular WGS, is being used for AMR surveillance. However, as described above, there is limited use of sequencing within national surveillance programmes, or the reporting of results to the appropriate surveillance programmes such as GLASS. Countries that have developed the use of sequencing in national surveillance programmes include:

United Kingdom: The UK Health Security Agency (UKHSA, formerly Public Health England) have been routinely sequencing all referred presumptive *Salmonella* isolates since 2014 which has transformed their approach to reference microbiology and surveillance [38]. They began reporting results derived from WGS analysis routinely for surveillance purposes from April 2015. Public Health England's (PHE) Gastrointestinal Bacterial Reference Unit (GBRU) receives approximately 10,000 *Salmonella* isolates each year from diagnostic microbiology laboratories, private laboratories and food, water and environmental laboratories for confirmation of identity and typing. Of the average 8,500 individual case reports of salmonellosis in England and Wales annually, ~95% of clinical diagnostic isolates are sent to the reference laboratory for confirmation and further typing. A small number of isolates are still being fully phenotypically serotyped to support validation of novel sequence types. Implementation of WGS has transformed reference microbiology services both in terms of improved accuracy of results, and reduced turnaround times by ~50% [38]. The UKHSA concluded that the integration of routine WGS as a replacement for traditional microbiological methods has revolutionised reference microbiology and impacted real-time surveillance of gastrointestinal pathogens for improved public health outcomes. UKHSA have now implemented routine WGS methods for *Salmonella* [73], *Shigella* [74, 75], *Campylobacter*, *Escherichia* [75, 76], *Listeria* [77], *Vibrio* [78], and *Yersinia* species [79]. It is envisioned that WGS methods will be implemented for all gastrointestinal bacterial pathogen services in England within the next few years. The use of WGS has now extended to the neighbouring nations of Wales and Scotland [80].

United States (US) of America PulseNet and GenomeTrakr laboratory networks work together within the Genomics for Food Safety (Gen-FS) consortium to collect and analyse genomic data for foodborne pathogen surveillance (species include *Salmonella enterica*, *Listeria monocytogenes*, *Escherichia coli* (STECs), and *Campylobacter*) [81]. The GenomeTrakr network was created in 2013 and is the first distributed network of laboratories to utilise WGS for pathogen identification [82]. It consists of 15 federal laboratories, 36 state health and university laboratories, one US hospital laboratory, two other laboratories located in the US, 21 laboratories located outside of the US, and collaborations with independent academic researchers. The network is regularly sequencing over 12,000 isolates each month. An economic evaluation of the GenomeTrackr programme, looking at three pilot pathogens (*E. coli, Listeria, and Salmonella*), found that each additional 1,000 WGS isolates added was associated with a reduction of approximately six illnesses per whole pathogen genome sequenced, per year [83]. Further analysis concluded that by 2019, annual health benefits are estimated at nearly USD500 million, compared to an approximately USD22

million investment by public health agencies. Even under conservative assumptions, the programme likely broke even in its second year of implementation and could produce increasing public health benefits as the network matures [83].

PulseNet International is a global laboratory network dedicated to bacterial food-borne disease surveillance, comprised of the national, regional and subregional laboratory networks of Africa, Asia Pacific, Canada, Europe, Latin America and the Caribbean, the Middle East, and the US; 86 countries in total [84]. The vision of PulseNet International is for WGS to be used in all public health laboratories to identify, characterise and subtype food-borne pathogens, largely replacing existing phenotypic and molecular methods in support of preparedness and response to food-borne illness at the local, national, regional and global levels.

Colombia, India, Nigeria, and the Philippines have all implemented sequencing as part of AMR surveillance through the **National Institute for Health Research (NIHR) Global Health Research Unit (GHRU) on Genomic Surveillance of AMR.** The GHRU recently undertook an exercise to integrate scalable genome sequencing into microbiology laboratories in Colombia, India, Nigeria, and the Philippines [67]. This was done to establish global AMR surveillance using WGS. They developed a laboratory implementation approach within a partnership of national and regional reference laboratories, academic centres, and private organisations [67]. Units in each country took part in a process that included an operational setup that verified the identity of all organisms referred from collection sites, testing their susceptibility against a panel of antimicrobials, and detecting previously published virulence factors. Practical considerations on the implementation of WGS within these settings, as well as a catalogue of laboratory setup challenges and solutions implemented as part of the process are documented [70]. For more information on the laboratory establishment in the Philippines, there is a publication that provides detail on the integration of WGS within the National AMR Surveillance Programme [85]. In addition, they have described how data bottlenecks for genomic pathogen surveillance across the four countries were overcome [86]. India has also opened a WGS reference laboratory in 2021 at the National Centre for Disease Control (NCDC) [21, 83, 87]. However there is limited information available on this laboratory and it is not clear if it is linked to the NIHR project.

Argentina participates in a number of regional and global AMR surveillance networks, which include: WHO-NET, SIREVA network (National Network for Surveillance of Serotypes and AMR in Invasive Isolates of *S. pneumoniae, H. influenzae* and *N. meningitides)*, GONOCOCO, and the Latin American Network for surveillance of AMR (ReLAVRA, PAHO [88]). There are numerous laboratories within these networks and the region that demonstrate WGS capabilities and are using it to monitor AMR [89-92].

ECDC Member States have been submitting WGS data for *Salmonella* to TESSy since 2019, which has enabled EU-wide surveillance and cross-sector comparison [93]. Out of the 20 laboratories that participated in EQA-11 for *Salmonella*, 14 (70%) performed molecular typing-based cluster analysis using WGS. Ten of the participating laboratories only used WGS and no other methods. The number of laboratories performing WGS has stabilised as no new laboratories performed WGS in this year and over time, and the use of PFGE and MLVA has become less frequent. The performance among the 14 participants using WGS was very high, as 13 (93%) correctly identified the cluster of closely related isolates. Countries that participated in EQA-11 were: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Luxembourg, Norway, Romania, Slovakia, Slovenia, Spain, Sweden and the Netherlands.

Some regions have future plans for the implementation of AMR surveillance using WGS:

- **East African Community (EAC).** The EAC is a regional intergovernmental organisation of six partner states; Burundi, Kenya, Rwanda, South Sudan, Tanzania and Uganda, with its headquarters in Arusha, Tanzania. The Federal Government of Germany has funded it through the German Development Bank (KfW), a three year

second phase of the EAC Regional Network of Public Health Reference Laboratories for Communicable Diseases Project which is set to further strengthen the capabilities of the EAC Partner States [94, 95]. This will be to improve responses to outbreaks of infectious diseases, including bacterial diseases and cross-border epidemics. In addition, the project will also have a special focus on AMR surveillance. Currently the project focuses on mobile, container laboratories for diagnosis of communicable diseases. For AMR surveillance it will require expanding to include mobile BSL3 container laboratories, equipped with bacterial culture facilities and the capacity to carry out WGS.

- **Arabian Peninsula** There is a clear understanding of the importance of WGS in monitoring AMR and there are plans for a state of the art WGS service for the region which will be developed at King Abdullah International Medical Research Center, Riyadh, Saudi Arabia [96].

Other regions have shown interest in using WGS and have demonstrated some capabilities, but as of now do not appear to be using it as part of regular surveillance systems. Some examples include:

- **Caribbean Public Health Agency (CARPHA).** A pilot project to monitor AMR through phenotypic resistance measurements combined with WGS was set up focused on *Klebsiella pneumoniae* [97]. Unfortunately, funding for this project has ceased along with CARPHA-based AMR surveillance.
- Within **South Africa** AMR is a major focus area of the South African Department of Health. The Group for Enteric, Respiratory and Meningeal Disease Surveillance in South Africa (GERMS-SA) performs national surveillance, and covers the majority of GLASS pathogens. They do appear to have capabilities to carry out WGS, but it is not routine yet [98, 99].
- In **Thailand** the National Antimicrobial Resistant Surveillance Center (NARST) published a study on sequencing of carbapenem-resistant *Enterobacterales* isolates. However only a limited number of isolates were sequenced due to resource accessibility [100].

Many further examples of this nature demonstrate the fragmented nature of WGS use in AMR surveillance, mainly occurring via academic studies that have limited funding. Funding is the main reason for the lack of sustainability of these endeavours, as well as lack of leadership and coordination on what is needed [19, 20].

2.7 Case study: Surveillance of *Mycobacterium tuberculosis* complex using whole genome sequencing

M. tuberculosis complex are a group of bacteria that cause the disease tuberculosis (TB), which in 2020 affected around 10 million people worldwide and caused 1.5 million deaths, making it the second leading infectious disease killer after COVID-19 [101]. TB is treatable and curable by a standard six-month course of four antibiotics. However, multi-drug resistant (MDR) TB, which is resistant to one or more of these first-line drugs, is a persistent and increasing problem. Second-line drugs are available to treat these infections, but they are more expensive, have worse side effects, and require up to two years of treatment. The best estimate of the proportion of people diagnosed with TB for the first time who have MDR-TB is 3-4% of cases and in those who have been previously treated for TB 18-21% cases are MDR [102]. Further resistance to second-line treatments has led to the development of extensively drug-resistant TB (XDR-TB), which require complex treatment regimes. Definitions of MDR- and XDR-TB are available on the WHO website [103]. *M. tuberculosis* is not included in the GLASS priority pathogen list since it is already a globally established priority which urgently requires innovative new treatments to combat MDR- and XDR-TB.

The biology of mycobacteria, rise of drug resistance and the technical complexities of phenotyping played a role in the development of sequencing workflows. The bacteria can be

time-consuming and technically challenging to grow in culture, with turnaround times of up to 12 weeks required for comprehensive phenotypic drug susceptibility typing (DST) of more complex cases. Additional genotyping methods support species identification and epidemiological studies. NGS approaches, including WGS, provide information to: support diagnosis; develop a drug susceptibility/resistance profile; allow for more precise analysis of transmission; and enable the categorisation of emerging drug resistance mutations. Projects such as FIND's Seq&Treat are working to support the adoption of targeted NGS for affordable, scalable and rapid TB drug susceptibility testing, in countries impacted by the spread of drug resistant TB [104].

In the past 10 years, WGS has been increasingly employed as a tool to support the diagnosis, clinical management and surveillance of tuberculosis [105]. For example, sequencing is currently under evaluation by the WHO for genotypic drug resistance testing [106]. To support laboratories worldwide in the interpretation of genome sequencing results, the WHO have produced a catalogue of *M. tuberculosis* mutations and their association with phenotypic drug resistance [107]. The use of WGS in the surveillance of TB is relatively advanced compared to many other pathogens and provides a useful case study in terms of the implementation of sequencing for AMR surveillance.

WGS is already used in England and the Netherlands for DST for first line drugs. In the Netherlands the tuberculosis National Reference Laboratory within the National Institute for Public Health and the Environment (RIVM) carries out WGS on TB isolates, this was due to replace genotyping in 2019 [108]. Evidence from this laboratory suggests that WGS is a more reliable tool than VNTR genotyping in terms of predicting epidemiological links between cases, which makes contact tracing investigations more efficient [109]. Implementation of WGS for diagnostics and surveillance of TB in England started in March 2017 and is now routinely used on all isolates for first-line drug DST and to inform understanding of transmission, with phenotypic DST also used if resistance to first-line drugs is suspected [110].

In Europe, the European Centres for Disease Control (ECDC) coordinates the Tuberculosis Disease Network, which includes the European Reference Laboratory Network for TB (ERLTB-Net) and the European Tuberculosis Surveillance Network (which includes experts from all 53 countries belonging to the WHO's European region and Liechtenstein) [111]. These initiatives aim to provide a framework for strengthening TB prevention and control in EU countries. Use of WGS for routine surveillance and/or outbreak investigations in the EU/EEA is increasing, from eight countries 2015 to 13 in 2017 [112].

In 2021, the results of an ECDC pilot study on the use of WGS for TB surveillance in Europe (EUSeqMyTB) was published [113]. The study evaluated the systematic use of WGS for surveillance of TB in EU/EEA countries and provided information on the drug resistance profiles and cross-border transmission patterns of MDR-TB, which differed between Western and Eastern European countries In the Western countries, a much lower proportion of MDR-TB cases occurred in individuals born in the country of notification, suggesting higher TB transmission in the migrant population. In Eastern countries, more transmission occurred between individuals born in the country of notification. While the pilot study was successful and facilitated equal access of all participating countries to the WGS data, a number of challenges were identified, which had an impact on the ability of the study to provide timely and accurate data for real-time surveillance. These included delays in culture and referral of samples to the sequencing laboratories, incomplete clinical and epidemiological data, and the need for data standards and establishment of criteria for triggering an international epidemiological investigation.

Another European-based study used publicly available MDR- and XDR-TB sequence data and metadata from the NCBI-Sequence Read Archive and compared it to WGS data and metadata of MDR- and XDR-TB isolates collected in Germany in 2012-13 [114]. The goal was to perform additional analyses to see if the publicly available data could improve TB

surveillance. Using the public data, they identified several clusters from multiple countries, suggesting international transmission, which they would have missed if looking at data from Germany only. While there were challenges with the study, such as some phenotype-genotype discordance and incomplete metadata, considering international publicly available data allowed them to identify previously unknown transmission between patients.

Many countries, particularly those with a disproportionate burden of MDR-TB, recognise the value of improved detection and surveillance. In South Africa, the national TB drug resistance survey, conducted between 2012 and 2014 showed that 4.6% of TB cases had some form of drug resistance to one or more first-line and second-line drugs [115]. While targeted genotypic assays such as the GeneXpert are extensively available to support diagnosis, novel mutations or mutations outside targeted regions cannot be detected [116]. NGS technologies are recognised as useful tools that could support more accurate diagnosis of drug resistant TB in South Africa. However, one of the major challenges is the lack of expertise and personnel in terms of handling and analysing sequencing data at scale; automated workflows, standardised reporting and user-friendly bioinformatics pathways are also required. Delays with culture highlight the opportunities for culture-free, direct from sample sequencing, and while momentum is building behind this approach it has not yet been used to support large-scale diagnosis or genetic epidemiology efforts [117, 118]. Further work is also needed to fully understand the costs versus benefits of using sequencing in this manner and to increase accessibility for LMICs [116].

One study has provided a proof of concept study on the implementation of WGS for TB in the Kyrgyz Republic, which is a LMIC with a high burden of MDR-TB [119]. The key lessons they highlight are that: sequencing costs can be significantly higher for LMICs; procurement and building of sequencing capacity can take a longer time than planned; infrastructure requirements should be considered from the beginning; quality assurance needs specific solutions; transitioning to WGS requires careful planning; ongoing support from experts is needed to ensure continued success.

Best practice will also differ between HICs and LMICs, due to the availability of infrastructure, personnel, and computer power. While HICs will have more access to bioinformatics expertise, end-to-end solutions that require minimal bioinformatics expertise from the end user are likely increase the availability of these approaches to a wider range of users. Many pipelines exist, which presents challenges for standardisation, comparison and validation, which is having an ongoing impact on implementation. In LMICs, cloud based solutions are likely to play an important role, however extensive validation and standardisation efforts are required to support the use of WGS for TB in high-burden countries [105, 120].

While progress has been made with the use of WGS for TB management, including surveillance, it is not yet in mainstream use in the majority of countries. This highlights the challenges inherent in implementing standardised, validated, easy to use WGS laboratory procedures and pipelines for infectious diseases in various resource settings.

3 Next generation sequencing platforms

There are a range of sequencing platforms available to perform NGS that each have different advantages and disadvantages depending upon the surveillance needs and context in which sequencing is performed. All NGS platforms can be used for WGS as well as other sequencing approaches detailed in Section 4.3. There are a number of factors that will influence decision making around which NGS platform to use. NGS comprises a workflow with multiple constituent parts and for laboratories, decisions made at each of the workflow stages are likely to be informed by availability, cost and training of staff, as well as wider considerations for infrastructure, resources and consumables. Currently there is a lack of consensus around NGS workflows, and this variation in NGS technologies and surveillance needs means that there is not a 'one size fits all' workflow for AMR surveillance.

With this in mind, this chapter will examine NGS technologies, and the following chapter contains an overview of NGS workflows and any key considerations for each step of the workflow in the context of AMR surveillance.

Sequencing is the process by which nucleic acids in a sample are converted into data that can then be analysed. There exist a number of different ways that this process can be achieved and a number of different technologies that can facilitate it. Broad categories of sequencing technologies are referred to in generations: first (Sanger and others), second (high-throughput) and third (long-read) generation sequencing technologies.

The term 'next generation sequencing' (NGS) is used to refer to sequencing techniques and technologies belonging to the second and third generations. These techniques include:

- Sequencing by synthesis (Illumina)
- Ion torrent semiconductor sequencing (ThermoFisher Scientific)
- DNA nanoball sequencing (Beijing Genomics Institute and MGI Tech)
- Single molecule real-time sequencing (Pacific Biosciences)
- Nanopore sequencing (Oxford Nanopore Technologies)

Each of the above techniques, which are broadly aligned to specific companies, can be performed on a range of instruments produced by the above-named developers. This is an active area of commercial research and development and it is likely that more sequencing platforms from new providers may become available in the future.

3.1 Illumina

Sequencing by synthesis represents the basis of the most widely used NGS methods. Sequencing using Illumina systems provides high throughput short-read sequencing which is widely used including as part of AMR surveillance programmes.

Sequencing by synthesis nucleotide identification occurs as modified nucleotides are incorporated into newly forming DNA. Fluorescently tagged (modified) bases are detected as they are incorporated. Unlike early chain-termination methods (Sanger sequencing), these fluorescently tagged bases do not cause DNA synthesis to stop. Each time a base is incorporated, the attached fluorescent tags are washed away after detection, allowing for more modified bases to be added after this point. The process is repeated until the maximum number of cycles (and therefore sequence length) is reached.

Illumina produces a range of platforms which cover a large array of sequencing applications; these differ in size, capacity, and cost. Illumina's key sequencing platforms are described below (Table 5).

Table 5: Information on Illumina sequencing platforms

Sourced from illumina.com

	iSeq 100	**MiniSeq**	**MiSeq series**	**NextSeq 550 series**	**NextSeq 1000 & 2000**	**NovaSeq 6000**
Run time	9.5–19 hr	4–24 hr	4–55 hr	12–30 hr	24-48 hr	13 - 44 hr (flow cell dependent)
Maximum sequence data output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	300 Gb	6Tb
Maximum read length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp	2 x 250bp
Description	Lower cost, lower capacity, smallest form benchtop sequencer available from Illumina. Released 2018.	Lower cost, lower capacity benchtop sequencer. Released 2016.	Mid-range benchtop sequencer providing longest reads available on Illumina platforms. Released 2011.	Mid-range benchtop sequencer, providing greater throughput than cheaper devices. The related *NextSeq 550 Dx* is diagnostic tool for specific clinical applications only [121].	Higher cost, high capacity benchtop system. Released 2020. As with the smaller capacity NextSeq, a *Dx* version of this machine is also available.	Highest and massive capacity, very large form free-standing system. Capable of sequencing multiple human genomes in one run. Released 2017.
GLASS priority pathogens			*E. coli* [122] Multiple, Including *K. pneumoniae and E. coli* [123] Salmonella spp. [122, 124]	MRSA [125] *Salmonella spp* [126] Enterobacter spp. [127]		K. pneumoniae [128, 129] *Shigella spp*. [130]

Illumina NGS sequencing has several advantages, primary amongst these is high throughput. The highest throughput system, the NovaSeq, can produce around 6Tb of sequence data per run. However, this sequencer is currently only suitable for use in higher resource settings with the appropriate physical infrastructure, which includes cooling systems required whilst the sequencer is active.

Use of Illumina technologies and available tools

Illumina sequencing platforms have been used extensively to support pathogen detection during outbreaks with a significant number of papers demonstrating their utility over conventional methods to track outbreaks, particularly in clinical settings [131]. WGS on Illumina platforms can be used for characterisation or assembling genomes of novel organisms, completing genomes of known organisms or comparing genomes across multiple samples [132]. Many of the available bioinformatics tools for AMR gene identification have been developed using (and for the analysis of) short-read sequencing data, in particular generated using Illumina sequencing platforms [133]. Additionally, the high resolution sequencing data produced by Illumina sequencing platforms allows identification

of mutagenesis, directed evolution, spatial and temporal dynamics of epidemics and mechanisms of disease transmission.

A significant volume of research has explored the use of Illumina sequencing for pathogen typing. One recent study compared the use of WGS data for outbreak surveillance generated on Illumina NextSeq or MiSeq to conventional pathogen typing techniques, including binary typing, PCR serotyping, multi-locus serotyping (MLST), multi-locus variable copy numbers of tandem repeats analysis (MLVA), and pulsed-field gel electrophoresis (PFGE) [134]. Besides being highly concordant (>99%) with results of binary typing, MLST, and serotyping, WGS enabled the identification of separate nested clusters among isolate groups that were undetectable using conventional methods. WGS was applied in routine epidemiological surveillance over a 12-month period enabling higher resolution to link point source outbreaks than would not have been possible with conventional typing.

Illumina MiSeq DX and Illumina NextSeq 550 DX are US Food and Drug Administration (FDA) regulated and CE-IVD marked NGS instruments for *in vitro* diagnostic use [135]. The availability of instruments targeted for diagnostic applications may be significant in the future if sequencing-based diagnostic tools are developed for AMR gene detection and resistance profiling. These tools are likely to be developed to inform clinical decision making and in the future could support clinical diagnostics. These machines can be used to multiplex up to 96 samples and for both targeted NGS and WGS applications.

Advantages of Illumina sequencing

- Comparatively low-cost sequencing at high throughput, where many samples are being processed
- High raw read accuracy and read depth generating high accuracy data
- One of the more commonly used systems for high resolution genomic analysis allowing collaborative development of expertise and advances. Many genetic or research laboratories already possess these systems, and bioinformatics pipelines are relatively well-established
- High levels of sample multiplexing are possible, meaning a high number of samples can be run at once
- Availability of FDA regulated, CE-IVD marked NGS instruments for *in vitro* diagnostic use
- Option of targeted and WGS approaches

Limitations

- Longer sequencing runs
- Most platforms are large and costly to purchase, some require specialised infrastructure for safe use
- Short reads limit accuracy in complex genomic regions and opportunity for identification of the genomic context of mobile genomic elements

A table summarising advantages and disadvantages of all sequencing platforms can be found in Appendix 8.4.

3.2 ThermoFisher Scientific Ion Torrent

ThermoFisher supplies semi-conductor sequencing through its Ion range of systems. Sequencing utilises a semi-conductor chip and a bead-based system which provide a physical platform for sequencing through DNA synthesis. This system relies on the detection of hydrogen ions released as nucleotides are incorporated.

The Ion GeneStudio S5 is the newest and highest capacity NGS system in this range. Older systems include the Ion Proton system and Ion Personal Genome Machine (PGM). An automated ‘specimen to report’ system, the Ion Torrent Genexus, provides rapid sequencing

with reduced sample contact [136]. A number of sequencers in the Ion range are described below.

Table 6. Information on Ion sequencing platforms

Sourced from thermofisher.com

	Ion PGM system + Ion 318 chip	**Ion Proton system + PI chip**	**Ion GeneStudio S5 System + Ion 540 chip**	**Ion GeneStudio S5 Prime System + Ion 550 chips**	**Ion Genexus**
Total sequencing and analysis time at max. throughput	7.2 hr (400bp)	4 hr	19 hr	6.5 hr	24 hours
Max. throughput/day	2Gb	10Gb	15 Gb	50 Gb	Unknown
Read length	400bp	200bp	200bp	200bp	Unknown
Description	Large benchtop sequencer providing the longest reads available from the Ion sequencing range. Older system with low throughput.	Fast, modest throughput benchtop sequencer.	Lowest capacity and throughput of the Ion GeneStudio range of sequencers. Low capacity in comparison to many other technologies. Small to medium profile desktop machine.	Fast and highest capacity Ion sequencing system.	Specimen to report automated system, claiming 10 minutes of 'hands-on' time. Permits variable throughput.
GLASS Pathogen examples	*ESKAPE pathogens incl. S. aureus, K. pneumoniae, A. Baumannii* and *P. Aeruginosa* [137] *Shigella spp. [138]* *E. coli* [139]				

Use of Ion Torrent technologies and available tools

The Ion PGM system is currently the one CE-IVD marked Ion Torrent sequencer commercially available. The PGM is equivalent to the Illumina MiSeq in terms of price, however the cost per sample is higher [135]. The run time of Ion Torrent sequencing is typically shorter and produce longer reads with a smaller data output for analyses. However, reads are only single-stranded resulting in an increased error rate particularly in homopolymer regions. In addition, read coverage in AT- or GC-rich regions has been noted to be poor quality compared to other short-read sequencing platforms limiting the accuracy of WGS data in these difficult to sequence regions [135].

ThermoFisher Scientific has developed culture independent targeted NGS approaches (Section 4.3) for specific sets of genes, such as 16S rRNA, using Ion AmpliSeq technology

for simple and fast library construction. This method uses ultrahigh-multiplex PCR from as little as 1ng of input DNA. This can be used for pathogen typing and to identify known AMR genes within the sample. Ion Torrent sequencing has been used in research using archived samples for disease surveillance and disease aetiology determination. Downstream of whole genome microbial and targeted microbiome sequencing, data analysis methods provided for Ion Torrent include de novo and reference-guided assembly, antimicrobial resistance detection, and typing of microbial strains.

Advantages of Ion Torrent sequencing

- Relatively inexpensive at low throughput
- Comparatively short sequencing runs enable faster return of results
- Low substitution error rate
- Some systems facilitate a highly automated workflow for easy adoption and consistent application of sequencing
- Availability of FDA regulated, CE-IVD marked sequencer
- Option of targeted and WGS approaches
- Longer individual reads

Limitations

- Lower throughput in comparison to other NGS technologies, therefore comparatively expensive at high throughput.
- Higher sequencing error rate
- Higher cost per sample
- Shorter reads than are possible with other NGS technologies able to perform paired-end sequencing
- Fewer bioinformatics tools built for Ion Torrent-generated data

3.3 BGI and MGI Tech DNA nanoball sequencing

DNA Nanoball sequencing (DNBSEQ) platforms produced by MGI Tech are available through the Beijing Genomics Institute (BGI). MGI's proprietary DNBSEQ technology enables flexible, high throughput, short-read sequencing performed on one of a range of instruments.

DNBSEQ utilises circularised reads which are repeatedly amplified using rolling consensus amplification to create a single long strand of DNA. A barcode and primers, which also enable circularisation, are attached to the target sequence during library preparation. The sequence is then massively amplified forming what are known as DNA 'nanoballs' (DNBs). The DNBs are then loaded onto a flow cell with embedded wells which facilitate detection of nucleotide integration through light detection in a similar manner to sequencing by synthesis. The repeated nanoball sequence is read to generate high accuracy consensus sequence data.

Table 7. Information on BGI and MGI Tech sequencing platforms

Sourced from bgi.com

	DNBSEQ-G50 (MGISEQ-200)	**DNBSEQ-G400 (MGISEQ-2000)**	**DNBSEQ-T7 (MGISEQ-T7)**
Run time – dependent upon several different parameters	10-66 hr	13-109 hr	20-24 hr (including loading and base calling)
Max. throughput/run	150 Gb	1.44Tb	6Tb
Max. read length	2 x 150bp	2 x 200bp	2 x 150
Description	Medium throughput desktop system, suitable for smaller genome sequencing and panels	Very high throughput, flexible, desktop sequencing system.	Massive throughput, fast sequencing system, Very large form factor. The latest system to be released from MGI.
GLASS pathogen examples	None currently available		

Use of BGI technologies and available tools

Only very limited information available for how BGI platforms have been applied for AMR surveillance. BGI platforms have been used in food and health settings to determine the presence of AMR genes [140-149]. Much of the literature using BGI and MGI technologies for pathogen sequencing has used the BGISEQ-500 (currently not listed on bgi.com) [140-149]. Subsequent platforms were launched, described in Table 7. BGI provide sequencing services allowing combined sequencing of DNBSEQ platforms with sequencing platforms provided by other providers. DNBSeq has been used primarily in East Asia including for sequencing of SARS-CoV-2 genomes. BGI sequencing is conducted through sequencing service providers with establishment in multiple countries, including a high-throughput human genome sequencing centre in South Africa [150].

Advantages of BGI and MGI technologies

- Flexible sequencing including range of run times, reads lengths and output.
- High throughput
- Linear amplification reduces error accumulation during amplification
- CE-IVD regulated in several Asian-Pacific countries

Limitations

- Shorter reads than are possible with other NGS technologies
- Highest throughput systems are very large

3.4 Long read sequencing for AMR surveillance

Microbial genomes are normally smaller and less complex than human genomes. However, while short read sequencing generate high quality data, it can be difficult to assemble small fragments into complex genomic regions, such as tandem repeats and GC-biased regions or regions that contain several copies of the same mobile genetic element [4]. The two main providers of non-synthetic long read technologies are Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) (ONT). Long read single molecule sequencers use distinct base technologies to read longer contiguous strands of between 10,000 – 100,0000 base

pairs of DNA, significantly longer than other NGS sequencing platforms [151]. Alongside advantages and disadvantages associated with the specific sequencing platforms, there are inherent advantages to producing longer sequence reads. As read lengths increase, it is more likely that a read will be distinct from other reads. This allows them to be computationally reassembled with less ambiguity. Long-read analysis of WGS data has facilitated the identification of mobile genetic elements that contain AMR determinants and revealed the combination of various AMR determinants co-located on the same mobile element [6]. This is particularly useful for sequencing of highly polymorphic or highly repetitive genomes. Some systems also offer the option to perform amplification-free sequencing, potentially removing some amplification bias and facilitating the examination of epigenetic modifications.

3.5 Oxford Nanopore Technologies

Oxford Nanopore Technologies (ONT) is a UK-based company that produces a range of sequencing systems based on nanopores. ONT's systems are designed to be relatively mobile, generate ultra-long reads and be more accessible to those with less experience and expertise. The systems are relatively low cost and are provided primarily through equipment purchase and customer support. Additionally, equipment can be rented from ONT.

Extracted nucleic acids are prepared for sequencing by ligation of a motor protein and adapter sequence at the ends of each strand. RNA may either be amplified and converted to cDNA prior to sequencing as occurs with other sequencing systems or the RNA can be read directly (direct RNA sequencing) without prior amplification or conversion to cDNA.

During sequencing, tagged, single stranded DNA or RNA molecules are fed through a membrane-bound protein pore – a 'nanopore' – by a motor protein. As each DNA or RNA nucleotide is fed through the nanopore, it interrupts the electrical current that exists across the pore, these signals are detected by the sequencing system. This pattern of disruption can be read to determine the base sequence of the molecule.

ONT produces systems capable of producing very long sequencing reads, up to around 800,000 bases in length. Sequencing is also fairly rapid: one nucleic acid can be read by each pore at a given moment, and each molecule is read at approximately 400 bases per second. ONT produce a range of sequencing systems which provide different capacity, throughput and mobility, and cover a wide range of price points.

Table 8. Information on ONT sequencing platforms

Sourced from nanoporetech.com

	Flongle	**MinION Mk1B**	**MinION Mk1C**	**GridION Mk1**	**PromethION 24**	**PromethION 48**
Run time – flexible dependent upon data required	1 min - 16 hrs	1 min - 48 hrs	1 min - 48 hrs	1 min - 48 hrs	1 min - 72 hrs	1 min - 72 hrs
Maximum output	2Gb	50 Gb	50 Gb	250 Gb	5.2 Tb	10.5 Tb
Read length	Dependent on length of target molecule Max. to date > 2Mb.					
Description	Lowest cost, reduced capacity adapter for MinION sequencer	Low cost, mobile, long read sequencer	Mobile long read sequencer and analysis platform in one unit	Medium capacity desktop long read sequencer with onboard analysis platform	High capacity desktop long read sequencer	Higher cost, high capacity desktop long read sequencer
GLASS pathogen examples		*Acinetobacter spp [152]* *Acinetobacter pittii* [153] *N. gonorrhoeae* [154]				

Use of ONT systems and available tools for AMR Surveillance

Early limitations in throughput have been mitigated by improvements in hardware and reagents [151]. ONT's sequencing systems, primarily the MinION, have been used extensively for microbial sequencing. The portability of the MinION and associated equipment has facilitated the development of the 'lab in a bag' used during Ebola and other outbreaks for surveillance and, more recently, for diagnostic testing [155]. It has also seen use in hospital settings for monitoring the spread of nosocomial infections [156, 157]. Utility of ONT Nanopore sequencing has also been demonstrated for outbreak surveillance and identification of AMR genes [152-154, 158].

Nanopore has had significant success in a wide range of applications for infectious disease including *de novo* bacterial assembly, pathogen typing, metagenomics and identification of AMR genes [6]. One example of the value of long read sequencing to resolve complex genomic regions and identify resistance genes was the application of ONT MinION nanopore sequencing to resolve the structure and chromosomal insertion site of a composite antibiotic resistance island in *Salmonella typhi [159]*. For this method, nanopore sequencing data was used to create a scaffold for an assembly from short-read Illumina data, an example of combination sequencing (discussed in more detail in section 4.6).

Oxford Nanopore's MinION platform has been applied to detect AMR genes' taxonomic origins and to explore their genetic organisation with mobilisation markers. Understanding the architecture of these markers could enable more targeted measures in order to mitigate the risks of AMR genes transferring among sites and improve biosecurity practices in

hospitals and other environments. A study of AMR within a veterinary hospital identified a relative abundance of AMR genes (from a variety of pathogens) co-localised to the laundry trolley and mop bucket samples compared with the intensive care unit (ICU) cages, suggesting amplification of AMR genes was most likely occurring in the collection of hospital waste [160].

One critical advantage of nanopore technologies is the real-time data availability. This approach is of particular significance where rapid interpretation is required. This approach has been applied for pathogen identification, possible after 10 minutes of sequencing, with all pre-defined AMR-encoding genes and plasmids previously identified from monoculture experiments detected within one hour using raw nanopore sequencing data [161]. The long read capabilities of ONT systems make them especially well-suited to *de novo* assembly of reference genomes from outbreak strains.

Advantages of Oxford Nanopore Technologies sequencing

- Rapid and flexible - particularly useful for sequencing smaller genomes. Sequences read in real time – allowing for termination when user determines enough reads have been generated
- Smaller sequencing units can be purchased at low cost
- Relatively inexpensive at low throughput
- Mobile sequencing – the small size and high portability of some systems means that these can be used in the field
- Some reagents do not require cold storage meaning they can be safely stored in environments where refrigeration is not possible or unreliable
- Simple user interface and analysis platforms – although knowledge is still required, the need for expertise for many applications of this technology is not required
- Simultaneous examination of methylation possible using direct RNA sequencing

Limitations

- Limited barcoding means this approach is more expensive than other approaches for sequencing at high throughput (a high number of samples). Currently the mobile sequencing units are not capable of providing the same level of multiplexing as other next generation sequencing technologies.
- Some techniques use reagents requiring cold storage meaning these approaches can only be used with reliable cold storage
- Raw signal output files are very large – this makes files difficult to store. As software and pipelines for analysis evolve rapidly, it is useful if not essential for these files to be available for subsequent analysis of the data. This could hinder data deposition on databases.
- Higher error rate in homopolymeric regions
- Lower read coverage than short-read platforms
- Higher base-call error rate and overall error rate

3.6 Pacific Biosciences

Pacific Biosciences' (PacBio) sequencing technologies use single molecule real-time (SMRT) sequencing to produce high accuracy long-read sequence data.

Target nucleic acid molecules are individually immobilised in microscopic pits called zero-mode wave guides (ZMWs) embedded in SMRT flow cells. The many embedded ZMWs contain a fixed DNA polymerase and are open to a pool of free labelled nucleotides. Sequence information is collected through the high-precision detection of miniscule emissions of light which are produced as complementary nucleotides are incorporated into the target molecule.

PacBio offer SMRT sequencing on their newer Sequel II system and older Sequel system. The Sequel II system enables high-throughput, high fidelity sequencing, with accuracy reported to exceed 99% and output of around 160GB per SMRT cell. PacBio's systems offer high fidelity (HiFi) sequencing, achieving higher accuracy than other long read systems.

PacBio sequencing is conducted through sequencing service providers [162] which exist in many countries including India (1 provider), China (5 providers) and Saudi Arabia (1 provider).

Table 9. Information on PacBio sequencing platforms

Sourced from PacB.com

	Sequel System + 1M SMRT cell	**Sequel II System + 8M SMRT cell**
Max run time per SMRT Cell	20 hrs	30 hrs
Average read length	Up to 30kb	Up to 15kb (high fidelity reads)
Max output per SMRT cell	20Gb	160Gb
Description	Large, mid-range long read sequencing platform capable of producing very high accuracy reads	High cost, large high-throughput long read sequencing platform capable of producing very high accuracy reads
GLASS pathogen examples	*S. pneumoniae* [163]	*P. aeruginosa* [164] *K. pneumoniae* [165] *Enterobacteriaceae* [71]

Use of PacBio technologies and available tools for AMR surveillance

Unlike other sequencing platforms, PacBio sequencers do not require amplification making these the fastest sequencers on the market and highly valuable for outbreak analyses [131]. Long read capabilities of PacBio systems make them especially well-suited to *de novo* assembly of reference genomes from outbreak strains. For example, PacBio was used for comparative analysis of MDR *Acinetobacter baumannii* outbreak at a hospital in North Carolina [166]. In this study, a PacBio instrument was used to sequence the genome of an isolate from the first detected case and assemble a draft genome in order to compare isolates against an outbreak-specific reference genome. By utilising the high sequencing speed and long reads, it was possible to generate a case-specific reference genome to analyse the phylogeny and transmission events. It also avoided the use of a generic reference genome which could have masked small evolutionary differences between the outbreak isolates.

PacBio was the sequencing platform used to generate a baseline reference genome to compare results from participating laboratories for the Genomics for Food Safety (Gen-FS) coordinated proficiency testing effort for genomic foodborne pathogen surveillance between US PulseNet and GenomeTrakr laboratory networks [81]. The resultant data was used to perform *de novo* assembly to generate assemblies for both chromosomes and plasmids and generate a consensus sequence. This methodology proposes using the longest reads as a scaffold to recruit other shorter reads to construct highly accurate preassembled reads without the need for short-read sequencing for error correction [167].

Advantages of PacBio sequencing

- Capable of high throughput, equivalent to that of Illumina sequencing platforms
- Capable of producing very high accuracy consensus reads – HiFi sequencing reads around 15,000 bases in length at over 99% accuracy
- Produces long reads
- Errors are random, not systematic, and can therefore be overcome with deeper sequencing
- Sequences read in real time – allowing for termination when user determines enough reads have been generated

Limitations

- Relatively expensive to run compared to other NGS technologies
- Systems are large and more costly than some alternatives
- Reagents are more costly than alternatives
- High vulnerability to DNA fragmentation with a high error rate

3.7 Choice of sequencing technology

There are a range of factors that will have an impact on choice of sequencing technology. Overarching considerations will depend upon the sequencing objectives, which involve a trade-off between accuracy, efficiency, time to result and cost. NGS instruments should be selected according to what best suits a given laboratory workflow without stretching a laboratory's resources; they can choose to set-up in-house platforms or out-source NGS services [135, 151].

Availability of CE-IVD marked machines could also be a consideration if sequencing information is used to support clinical decision making. This is currently limited to short-read sequencers and may have implications for the future as more targeted panels are developed for AMR surveillance using NGS technologies. Appendix 8.3 includes a summary comparison of a range of available sequencing machines.

A summary review of a sample of the literature on WGS for AMR surveillance for each of the GLASS priority pathogens indicates that Illumina is the most commonly used technology provider. One example that illustrates this is a systematic review carried out on the use of WGS for surveillance of *Enterococcus* spp. [72]. Of the studies reviewed, 89% used Illumina technology for sequencing, of these, approximately half used the MiSeq platform.

More detailed decision making around choice of sequencing platform will consider:

- Technical characteristics of the platform:
    - Throughput levels, including minimum samples and pooling / barcoding possibility (sample capacity per run and per kit size)
    - Yield and quality per sample
    - Run time and time for analysis
    - Type of sequencing – chemistry and read length considerations
    - Scalability and flexibility
- Established personnel and facilities for troubleshooting, protocols, and pipelines for analysis
- Infrastructure needed to implement the platform – including available technologies, additional technology needed and personnel requirements
- Costs related to equipment (sequencer and associated equipment), establishment of workflow, reagents, other consumables, personnel time, bioinformatics resources, analysis, and maintenance. Costs will occur both in set up and on an ongoing basis – which can be cost per sequencing run, per genome sequenced, or per megabase of output data

- Proposed sequencing application(s), such as ongoing or real-time surveillance, diagnostics support, vaccine development, research, including the number of samples expected and the quality required for these applications
- Short-read versus long-read technologies, for example short-read platforms tend to have more accurate base-calling than long-read, which has an impact on accuracy.
- Where sequencing is being deployed – e.g. in a mobile or central laboratory
- Anticipated sequencing output – this will be impacted by the number of samples being processed. Sample batching can reduce the cost per sample but might have an impact on the amount of data generated per sample. The volume of data outputs will have an impact on downstream bioinformatics analyses.

While the cost of sequencing technologies (including upfront and ongoing costs) is an important question, it is a challenging issue to address for a number of reasons:

- Full economic cost-benefit analyses of applying NGS for pathogen genomics in public health has not been done and any such analysis will be country and region specific
- Estimates of the cost of consumables from general laboratory equipment (e.g. pipettes, centrifuges), plastics (e.g. tips, tubes), reagents and computational requirements (e.g. laptops, software) are variable and will depend on the specific needs of the laboratory and the jurisdiction in which it resides. These consumable costs are often higher in LMICs because of higher costs of shipping, customs formalities and taxes, and profit margins for local companies and distributors. They also have to accommodate for fluctuating exchange rates to establish costs. It can be expensive and time consuming for LMICs to import the reagents and other equipment required for sequencing.
- Reagents may have a relatively short shelf life (<3 months). Importation procedures in several LMICs can be long and bureaucratic processes. As such, reagents received in the laboratory may have already expired
- Most sequencing platforms rely on kits that contain proprietary or patented reagents, and it may not be possible to find or develop lower cost alternatives. These kits can be categorised as closed or open systems – closed systems require reagents from a specific producer, typically ensuring high quality reagents but creating dependence on a single supplier.
- The availability and cost of appropriate cool or cold chain shipping and storage may significantly affect the ability to maintain the quality of sequencing reagents. The reagents also have a specific shelf life, sometimes hindering stockpiling.
- The current status of sequencing facilities in the country in question – establishing services from scratch will have different cost implications to augmenting existing facilities.

3.8 Future uses: combined sequencing approaches

One final consideration is around combining sequencing technologies. Different NGS technologies are known to have different respective strengths and weaknesses and it has been suggested that the higher error rates of long-read sequencing could be overcome by performing short-read sequencing in parallel [4]. A recently developed method of hybrid assembly of both short- and long-reads was able to resolve complex bacterial genomes and mega-plasmids with high accuracy. However, there are questions around the extra cost of conducting both long-read and short-read sequencing and whether the question at hand requires the additional information provided.

The majority of sequencing is currently performed using Illumina MiSeq instruments [168]. Reads produced on Illumina machines typically span around 250bp and the quality of base calls is very high with Phred quality scores (Q-scores) above 40. However, short read sequencing comes with the limitation that some regions of the genome are significantly more

difficult to sequence, such as tandem repeats and GC-biased regions or regions that contain several copies of the same mobile genetic element [4]. Mobile genetic elements are typically identified as fragments of DNA that encode virulence or resistance determinants, as well as enzymes for integration into new host DNA [169]. Plasmids are one type of MGE able to integrate directly into the chromosome or survive as extrachromosomal material in the cytoplasm of bacteria [170].

For AMR surveillance, identification of the location of AMR determinants on the chromosome or mobile genetic elements may have significant implications for surveillance efforts and any infection prevention control decision making. Long-read sequencing is able to generate long continuous reads, capturing the complete genetic architecture, including the ability to localise AMR genes on these mobile genetic elements. However, reads produced by the ONT MinIon can be greater than 100kbp in length but the Q-scores are often below 10 [168].

Researchers are combining long and short read technologies in a hybrid approach to maximise the utility of these respective technologies. In a study comparing the use of an Illumina MiSeq, ONT's MinION and in combination, the raw data produced by the MiSeq and assembled using the bioinformatics package SPAdes was the most accurate when compared to the reference genome, but the genome assemblies were fragmented due to the short length of the reads [168]. In comparison, the data produced by the MinION and assembled using the bioinformatics package Canu resulted in the least accurate genome assemblies. However, combining MiSeq and MinION read data to create hybrid genome assemblies with Unicycler resulted in contiguous assemblies that covered more of the reference genome than assemblies generated using MiSeq data alone.

A leading challenge in AMR surveillance is the ability to accurately define the genetic architecture of plasmids, capable of transferring AMR genes between bacterial isolates. Plasmids are difficult to assemble from short-read WGS data because they contain regions that are challenging to sequence [6]. Berbers et al. resolved the genetic architecture of a plasmid in *Bacillus* using long read sequencers (ONT Nanopore and PacBio Sequel) and a short read sequencer (Illumina MiSeq) [171]. This approach resulted in an assembly of one chromosome and one plasmid, each with several AMR determinants of which five are against critically important antibiotics. The sequencing data was able to characterise all AMR genes, their exact location and to bridge the gaps of repetitive regions.

Additionally, this hybrid assembly was used to verify existing PCR and qPCR assays for this plasmid and determine the phenotype using AST. Due to the concern of spread of these AMR genes when present on mobile genetic elements, it is crucial to characterise their location with as high sensitivity as possible, which is not possible using short read sequencing alone. This study demonstrates the value of combined sequencing methods to improve the quality of assembled genomes to support AMR characterisation and inform future AMR surveillance efforts by improving AMR characterisation and molecular diagnostics [171].

A number of further studies have employed combined sequencing methods to address challenges around mobile genetic elements [172, 173], resolve complex genomic architecture [174] or define the AMR gene profile of a phenotypically resistant bacteria [175]. However, in most circumstances hybrid assembly may not be worth the additional cost. However, the added costs of MinION sequencing for hybrid assembles may be warranted when the additional detail is needed to compare highly related isolates or to characterise the complete genetic content of an isolate.

During outbreak investigations it is often necessary to demonstrate that bacterial isolates from clinical samples, the environment, or food are genetically identical or closely related, especially when pursuing public health action.

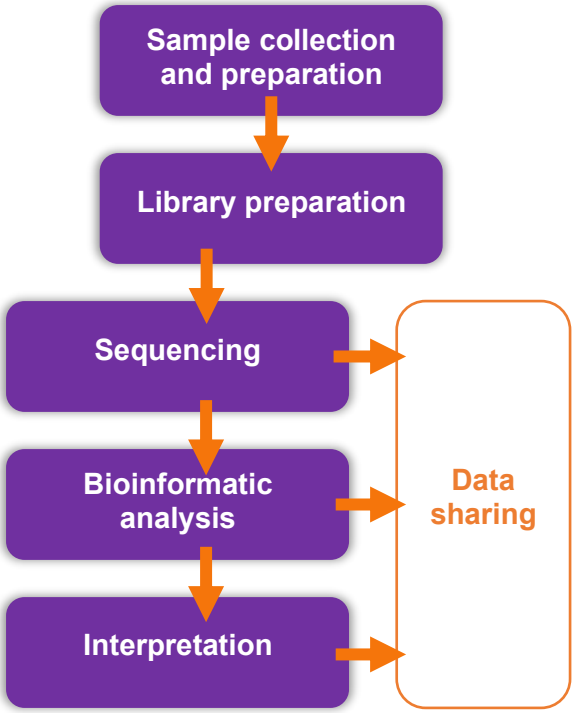
The use of complete hybrid genome assemblies would allow for the most accurate comparison between the isolates. There are a number of outstanding challenges that need

to be overcome, including the need to establish and validate this method, as well as to determine for which pathogens and for which surveillance objectives or clinical scenarios this approach may be useful. Some barriers to implementation will include determining the current infrastructure for combined sequencing, and considerations for additional resource use in terms of staffing, all associated costs and bioinformatics intensity.

4 Sequencing workflows

Producing readable genomic data from a physical sample requires completion of several steps. Each of these steps impacts upon the type or quality of information retrieved from the original sample, and the ability to interpret something useful from that information. Sequencing workflows can be highly flexible - the steps may be specified in one complete or part-complete workflow or tool (dependent upon how the technology is developed), or produced as individual parts that can be combined with other technologies or processes at different stages of the workflow to provide something highly customised.

Figure 1: Schematic of a sequencing workflow



The complexity of the steps varies depending upon a number of factors, including: sequencing equipment selected, sample type and information required. These can also be decided by choice of the sequencing approach (targeted, whole genome or metagenomic) that will be used. Although data analysis and interpretation is key to drawing appropriate conclusions from data, the importance of preparatory steps for assuring good quality data and appropriate subsequent interpretation should not be underestimated.

The steps towards interpretation and prior to sequencing are substantially variable – there are many options for how this can be conducted, and many tools and consumables exist with broad or exclusive application to the different types of sequencing and sequencing equipment. Logistical and environmental limitations also impact upon the steps that can be taken, where optimal conditions and resources are not available, there are often alternative ways of completing a specific step.

As a result there is currently no consensus regarding the optimal workflow for obtaining NGS data using the different platforms. Sequencing has been applied within the context of AMR surveillance for a number of GLASS pathogens using different sequencing technologies (see Tables 5-9). A wide range of workflows have been described for performing AMR surveillance, using sequencing approaches such as targeted NGS of specific genes, metagenomic sequencing of entire microbial communities, and WGS of pure microbial isolates (the most widely used approach).

There is significant variation between laboratories as to what reagents, sequencing platforms and bioinformatics pipelines have been used to report sequencing data. An example of this is in the recent external quality assessment scheme for *Salmonella* by the European Centre for Disease Control (ECDC), where 14 of the 19 laboratories reported using WGS as part of their AMR surveillance [93]. Each of these laboratories used a different combination of library preparation reagents and sequencing platform. However these laboratories predominantly used Illumina platforms with kits provided by Illumina for library preparation.

Key steps in a basic NGS workflow can be summarised as follows (see also Figure 1):

- **Sample collection and preparation** - Includes the steps from collection of a sample through to the storage and transportation of that sample, isolation of the pathogen using a culture-based method if appropriate, to the extraction of DNA (or RNA) prior to further processing. For certain sequencing approaches, extracted DNA may undergo further amplification.
- **Library preparation** - Transforms the retained nucleic acid portion of a collected sample into a prepared sample library ready for sequencing. This may include fragmenting or size selection of nucleic acids (dependent upon application), addition of sequencing adapters and quantification and quality control of resulting libraries.
- **Sequencing** – The process by which the sequence of bases in a series of nucleic acids is detected by one of a number of methods to provide readable data – raw sequence reads. This process covers entry of a prepared sample library into a sequencing system to retrieval of raw sequence information. Different sequencing approaches will capture different amounts of sequencing data. Broadly, these sequencing approaches are described as targeted, whole genome or metagenomic.
- **Bioinformatic analysis** - This includes the processing and conversion of raw data that is produced during sequencing into one of several formats that is suitable for ongoing analysis or interpretation.

Data sharing may also be considered part of this workflow. The approaches adopted towards data sharing, including the databases used, file type, and attachment of metadata will impact upon the ability of others, outside of the original sequencing group, to effectively use the information.

4.1 Sample collection and preparation

GLASS has recommended that laboratories looking to establish NGS services should have adequate infrastructure for isolating organisms from clinical samples, culturing them and extracting DNA from cultured isolates. Currently, most sequencing is performed using culture-dependent methods, from samples which have first been cultured to isolate the microbe of interest. Application of culture-independent and metagenomic approaches is discussed in Chapter 6. AMR surveillance also signals the emergence and transmission of mobile genetic elements containing AMR determinants among animals or the environment and between animals and humans, underpinning the need for a ‘One Health’ approach. As a result, samples may be collected from a wide range of environments including, for example, clinical environments, animal and food samples.

DNA extraction is a critical step for ensuring high quality sequencing data. Extracted genomic material must be of sufficient yield, purity and integrity and additionally, the quantity and quality requirements are dependent on the library preparation kit to be used and the desired sequencing application [135]. The success of most NGS applications depends on high quality and high quantity of input DNA to ensure that the effects of DNA contaminants, such as human DNA in clinical samples, are minimised. This ensures there is sufficient DNA available to generate sequencing reads with high depth and coverage for downstream sequencing analysis. For example, culturing the target bacteria increases the DNA yield,

alternatively, culture-independent target enrichment steps can be used to amplify target DNA.

To prepare samples, DNA needs to be extracted and a library prepared with key quality assurance steps in each of these processes [13]. Quality control (QC) is essential to guarantee the accuracy and precision of results of any laboratory with sequences of poor quality resulting in failure to reveal AMR genes or mutations. Prior to sequencing, quality assurance aims to evaluate the DNA quality and quantity as well as determine contamination which may originate from upstream handling of bacterial isolates and DNA purification, as well as from the preparation and running of the DNA samples. QC metrics for NGS data are well established and widely available, however, there is a lack of international standards for accepted minimum QC-thresholds. There are ongoing efforts to reach consensus and ensure comparability of WGS data [50, 81, 176-178].

4.2 Library preparation

During library preparation, nucleic acids are prepared for sequencing by the addition of identifiers and adapters that allow molecules to adhere to the sequencing flow cell. Library preparation is an essential step prior to sequencing using most systems. The type of library preparation tools and techniques applied are normally closely tied to the sequencing platform to be used, and the steps included in these workflows depends upon how the sequencing itself is conducted. Some steps are common to library preparation methods across several systems, as described below.

Library preparation may also include fragmentation or size selection of extracted nucleic acids (dependent upon approach and sequencing system to be used), amplification (involving the conversion of RNA into cDNA if this is required), quality control, and sample quantitation steps.

Quality assessment and quantitation of nucleic acid samples can be performed using one of a range of methods, including qPCR. These steps help to ensure the correct amount of sample can be loaded onto the sequencer and determine whether potentially disruptive contaminants are present that could impede sequencing.

Many library preparation kits are sequencing-platform-specific and sold by companies producing associated sequencing equipment. For example, Ion Torrent sequencers use emulsion PCR library amplification prior to sequencing which can be a complicated process. However, there is the option of a separate, automated library preparation system performed using the Ion Chef system for library preparation, which makes the speed of this process comparable to Illumina library preparation. However, a range of kits are also available from other suppliers and some of the preparation may be performed without commercial kits. The extent and type of sample and library preparation required depends upon several factors, including sequencing platform and type of sequencing; amount and quality of starting material; the biocontainment level of available facilities; and time taken.

Selection and execution of appropriate sample and library preparation can subsequently impact upon overall sequence quality, genomic coverage and uniformity, error rate, selection of bioinformatic pipelines, and variant interpretation.

Example: Illumina library preparation kit

One of the most widely use library preparation kits for pathogen genomics is the Nextera XT DNA library preparation kit workflow. This is one of the earliest library preparation methods and this library preparation method has been shown in several studies to have GC bias resulting in uneven sequence coverage in extreme AT and GC rich regions which impacts on the quality of downstream bioinformatics processing [179-183]. Subsequent methodologies have been developed to overcome this limitation.

A recent study comparing different enzymatic DNA library preparation kits (Illumina DNA Prep, KAPA HyperPlus, NEBNext Ultra II FS and QIAseq FX) did not identify any significant GC-bias [184]. Choice of library preparation kit is likely to be most dictated by availability, local pricing and individual workflow considerations. For example, the study identified that the Illumina DNA Prep provided the most streamlined and fastest workflow however, the NEBNext Ultra II kit was the lowest cost, $32 per sample. For all kits evaluated, the total number of genes including the accessory genome was adequate to perform whole genome multi-locus sequence typing (MLST) and the quality of data produced exceeded minimum quality requirements set forth by the quality assurance programme of PulseNet and GenomeTrakr networks for high quality SNP analysis [81, 184]. Currently the Nextera XT DNA library preparation kit is recommended by the Genomics for Food Safety (Gen-FS) consortium in the WGS SOP [81].

4.3 Sequencing approaches

NGS technologies may be applied as follows (Table 10):

- Whole genome sequencing – captures the entire genome sequence of an organism, most often from cultured pathogen isolates
- Targeted panel tests – specific sections of the genome are sequenced to capture data on particular genetic loci of interest
- Metagenomics – captures all genomic information in a sample without prior culture typically from a range of microbes (chapter 6)

These approaches rely upon the same basic NGS workflow and the three applications may be run on the same NGS instruments. However the sample type input requirements and processing steps can vary widely depending upon the method chosen.

Targeted sequencing approaches typically have the option of using amplicon-based sequencing or bait-capture/target enrichment methods. Amplicon sequencing relies on the prior amplification of regions of the target genome in order that these regions exist in greater concentrations in the processed pre-sequencing sample than the original raw sample. Bait capture/target enrichment involves nucleic acid capture through the use of ‘bait molecules’ facilitating the selection and targeting of specific regions of the genome. These methods both rely upon prior knowledge of the target sequences of interest.

Targeted methods can be used to amplify certain genomic targets of interest, sufficient to allow identification of key pathogens or key genetic features of AMR. However, the principles of targeting genomic regions could also be applied to obtain WGS data, without the need for prior culture of pathogens. For example, throughout the COVID-19 pandemic, a targeted amplicon WGS approach, using targeted overlapping amplicons designed to provide coverage of the complete viral genome, has been a popular method to obtain SARS-CoV-2 WGS data.

Targeted NGS approaches, both of regions of interest and to obtain WGS data, have not been widely employed for the surveillance of antimicrobial resistance, other than in specific examples such as for *M. tuberculosis*. Most current NGS surveillance approaches rely on culture of the pathogen under investigation to increase the amount of DNA available for sequencing, followed by WGS using unbiased sequencing methods not tailored to any specific region of the pathogen genome (section 4.1). However, there is a growing interest in the potential for targeted gene panel solutions for AMR surveillance, where only specific genetic regions of interest are monitored. These panel tests are discussed in more detail below. Targeted culture independent approaches with the specific intention to obtain WGS data of specific pathogens are also available and in development; these are discussed in more detail in Section 6.3

There is also increasing research and investment into the use of culture-independent methods such as metagenomics, which can provide information on a range of pathogens as

well as more complete genomic data for pathogens of interest. These methods are not used routinely in surveillance efforts, which are mostly focussed on particular pathogens of interest in clinical samples and isolates, but could fulfil specific surveillance requirements in the future. Targeted culture independent approaches with the specific intention to obtain WGS data of specific pathogens are also available and in development; these and metagenomic approaches are also discussed in more detail in Chapter 6.

A broad overview of the advantages and disadvantages of WGS of specific microbes, targeted NGS panels, and metagenomics is given in Table 10.

Table 10. Comparison of different NGS sequencing techniques for surveillance

	Advantages	Disadvantages
Targeted panels	Less expensive and faster than other sequencing methods Could be easier to use and interpret results in non-expert settings High sensitivity and specificity for targets of interests Panels could be custom made for specific scenarios Potential use for direct from sample analysis	Limited to known targets so cannot be used to identify novel microorganisms or AMR genes Data cannot be reanalysed to retrospectively identify features outside of the areas included in the original panel Can only be used to identify a defined number of targets Difficult to link AMR genes to pathogen source More limited data when trying to understand genomic context and enable monitoring of transmission dynamics
WGS of specific microbes	Provides detailed genomic information that can be used in epidemiological studies to monitor transmission dynamics Provides taxonomic and genomic context of AMR genes e.g., which pathogen contains which AMR genes, where on the genome they are located (e.g. on core genome or mobile plasmids), and whether multiple AMR genes are present conferring multidrug resistance Can be used to identify novel AMR genes Potential for direct from sample sequencing using non-targeted or targeted approaches The entire genome can be reanalysed to retrospectively identify features of interest	Most methods currently require culture to obtain bacterial isolates, which is slow and not always possible Difficult to apply to multiple organisms simultaneously Requires high level of expertise Requires sufficient infrastructure and this alongside the techniques used make it typically more expensive to establish and run, especially in LMICs with less ready access to resources
Metagenomics	Can be used to identify and monitor multiple organisms and AMR genes simultaneously and at scale, broadening scope of detection and providing insights of risk of AMR gene transfer within bacterial communities Can be used to identify novel AMR genes Can be used directly on samples, avoiding the need for culture. Ability to be used across multiple surveillance scenarios and purposes to provide different levels of genomic information, allowing data to be comparable across different surveillance applications and facilitating a One-Health approach. The data available can be reanalysed to retrospectively identify features of interest, though the utility of this will depend on the genome coverage obtained in the original experiment	Currently difficult to quantify the actual levels of different pathogens, for direct correlation with infection risk Does not take into account pathogen viability, which could lead to overestimation of infection risk [185] Can be hard to accurately identify closely related pathogens May not be economically and computationally feasible to obtain the genome coverage needed to provide taxonomic and genomic context of AMR genes e.g. for WGS, or to link AMR genes to specific pathogens Results very susceptible to external parameters, requiring protocol standardisations Relatively expensive sequencing to perform, requiring similar resources to WGS of single isolates

4.3.1 Current use of targeted NGS panel tests for AMR surveillance

Currently there are very few targeted sequencing applications that have been developed for AMR surveillance. This is in part explained by the challenge of generating reliable and representative databases necessary for the development of these targeted sequencing applications. There are some pathogen specific approaches being developed, such as for GeneXpert for MDR-TB. The SARS-CoV-2 pandemic has emphasised the challenges of respiratory diseases and the need for diagnostic solutions to support a wide range of clinical questions. Two examples developed for Illumina and ThermoFisher Scientific Ion Torrent sequencing solution are discussed below.

However, these approaches only identify known mechanisms of resistance and have limited opportunity for identification of novel AMR genetic mechanisms. WGS also has the advantage that the pathogen genome generated may be reanalysed, so can be used as part of retrospective surveillance following identification of novel AMR genes. Reanalysis of NGS panel data is limited to changes within the specific genomic targets used. Additionally, targeted panel approaches are not as useful in outbreak scenarios where the broader amount of genomic data obtained by WGS can be used to inform phylogenetic and taxonomic approaches to trace the origin of an outbreak.

Many panels that do exist are designed for the purpose of diagnosis, though in theory they could be applied to surveillance. As such, their design has not taken surveillance into consideration, or been demonstrated to be a validated surveillance tool.

WGS has the advantage that the entire pathogen genome generated may be reanalysed, so can be used as part of retrospective surveillance following identification of novel AMR genes. Reanalysis of NGS panel data is limited to changes within the specific genomic targets used. Additionally, targeted panel approaches are not as useful in outbreak scenarios where the broader amount of genomic data obtained by WGS can be used to inform phylogenetic and taxonomic approaches to trace the origin of an outbreak. Lack of understanding and/or consensus on which targets should be monitored in different surveillance environments is potentially limiting the use of panel tests. More research is required to understand the utility of these tools for providing actionable surveillance information, compared to other methods.

Commercially available AMR panel tests

Targeted AMR panel tests designed for use with certain types of samples and to detect specific AMR genes are commercially available for research purposes, and have been individually developed by academic groups. Two examples of commercial panels provided by Illumina and Thermo Fisher are provided below. Recently, the company Ares Genetics in collaboration with academics at the University of Vienna, developed the ARESdb AMR panel, and NGS target-enrichment panel for the detection of 9218 AMR markers (7312 ANR genes and 1906 genetic variants). The panel has undergone preliminary validation based on its ability to detect reference samples of *K. quasipneumoniae, E. faecium, S. aureus* and *E. coli*, with sensitivity appearing higher than using metagenomics methods. The method could also be combined with 16S ribosomal RNA sequencing to profile bacterial diversity [186]. Another example of a commercial kit designed to target a specific pathogen is the DEEPLEX® MYC-TB assay, which combines targeted deep sequencing with automated data analysis. It is designed to detect 18 *M. tuberculosis* drug resistance-associated gene targets combined with targets for mycobacterial species identification and MTBC strain genotyping [187].

Example: Illumina platform targeted NGS panel approaches to detect AMR genes

Respiratory pathogens are an appropriate target for panel sequencing due to concern around the number of different pathogens (bacterial, viral and fungal) which may result in severe disease with similar clinical presentations. Each disease has distinct treatment implications and therefore the ability to use molecular diagnostics to identify the pathogen

causing disease and tailor treatments would be advantageous. The recent COVID-19 pandemic has increased awareness of this challenge for clinicians.

Illumina have partnered with IDbyDNA to develop the Respiratory pathogen ID/AMR enrichment panel, launched in October 2020 [188]. This research use only (RUO) product is a 24-hour, sample-to-result workflow [189]. This panel targets 1,500 markers to identify known respiratory pathogens (187 for bacteria, 53 for fungi and 42 for viruses) and perform concurrent profiling of 1,218 AMR markers for pathogen characterisation in one assay from respiratory samples [190]. This panel includes targets for five of the eight pathogens highlighted by the GLASS initiative, as well as *Pseudomonas aeruginosa*. This workflow has been designed to consider the presence of viruses, such as SARS-CoV-2, with an RNA genome and therefore the workflow includes preparation with Illumina RNA Prep with Enrichment to generate enriched DNA and cDNA libraries. To perform data analysis, FASTQ sequencing data can be interpreted using the IDbyDNA Respiratory pathogen ID/AMR panel platform accessed in BaseSpace™ Sequence Hub, which generates a report for each selected sample. IDbyDNA is currently running clinical trials to evaluate their analysis platform with the Healthy Volunteer Study: Next-Generation Sequencing-based Analysis of the Urine of Asymptomatic Individuals [191].

In addition, Illumina have developed an AmpliSeq for Illumina Antimicrobial Resistance Research Panel which detects targets in AMR genes associated with resistance for 28 different antibiotic classes. The panel contains two pools of a total of 815 amplicons to assess the presence of 478 AMR genes. This panel cannot be used to perform pathogen typing. One example of its use was in a study evaluating the resistance profile from faecal samples from clinically healthy women and infants in Australia as part of a descriptive pilot study [192].

Example: ThermoFisher Scientific targeted NGS panel approaches to detect AMR genes

ThermoFisher Scientific have also developed targeted panels for their Ion torrent sequencers [89]. The AmpliSeq AMR research panel targets 815 amplicons to detect 478 AMR genes which indicate resistance to 28 different antibiotic classes. This panel has been developed to enable monitoring of environmental and biological samples to understand treatment efficacy. In comparison, the Ion AmpliSeq Pan-bacterial research panel enables detection of 21 species and 364 AMR genes related to 31 different antibiotic classes, as well as 24 amplicons for 16S profiling using the Greengenes database [89]. Health acquired infections are a significant cause of mortality, so this panel has been developed to be a quick, accurate and cost-effective tool to detect bacterial organisms at genus and species level. Additional panels have been developed to pathogen type SARS-CoV-2, Ebola and for AMR surveillance of eight known AMR genes for TB.

4.4 Bioinformatics analysis, interpretation and data sharing

Bioinformatics analysis for WGS data has several key stages [4], including:

- Quality control
- Assembly
- Sequence annotation
- Comparison of genomes
- Confirmation of species identify
- Subtyping of isolates
- Identification of genetic determinants of AMR

Next generation sequencing data analysis requires a high volume of computing resources to generate, analyse, store and manage the sequencing output data, and this can be one key limiting factor for successful implementation of NGS technologies [135]. Each step of the bioinformatics pipeline requires specialised tools, of which there are many options with

different strengths and weaknesses [4]. The majority of available bioinformatics software also requires some knowledge of programming in order to write code capable of running automated data manipulation and analyses. There is an abundance of tools which are continually under development, refinement and packaging together as bioinformatic pipelines, and as a result bioinformatics approaches for AMR surveillance remain highly variable [193].

For each of the stages outlined multiple tools have been developed. One major limitation is that many of the institutions developing and releasing tools and databases for microbial WGS data processing, annotation and analysis, are also releasing protocols defining suggested tools and pipelines. Research comparing these tools in different surveillance environments would be needed to define which of these tools is most appropriate for a particular setting and any key limitations.

4.4.1 Pre-processing and quality control

Pre-processing and quality control of sequencing data is essential to ensure the accuracy and precision of any downstream analysis [50]. Steps include evaluating the raw sequence data, trimming reads, identifying contamination and setting quality control (QC) parameters for draft genome assembly [50]. Poor quality sequencing can result in major errors in AST by failing to detect AMR genes and mutations, or incorrectly classifying them as demonstrating susceptibility or resistance when compared to phenotypic methods. Additionally, contamination of the DNA or erroneous data handling may also introduce errors. Quality metrics for NGS are widely available [194], however, there is no international consensus on minimum performance standards to ensure good quality NGS data processing, or to harmonise analytical approaches and interpretation criteria when using NGS-based approaches to predict AMR. To enable international data sharing of AMR surveillance data, consensus on minimum QC metrics is required to ensure that results are interpretable and comparable [4]. This may depend upon the sequencing technology and the organism being sequenced. There have been some attempts to perform analytical performance validation of NGS based clinical microbiology assays using K-mer analysis workflows in line with Clinical Laboratory Improvement Amendments (CLIA) performance requirements for pathogen typing and AMR gene detection [195].

Genome assembly

Typically, the first step of bacterial AMR surveillance workflows will be to assemble shorter fragments into a complete sequence, either by assembling the *de novo* sequence using overlapping or by mapping reads against a known reference genome [151]. A set of QC parameters for draft genome assembly and their explanation has been listed by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) committee [194].

- **Assembly based methods** generate short overlapping sequences, known as contigs, which can then be used to reconstruct the genome sequence. This method is advantageous when identifying known or novel resistance genes which share low similarity with the reference database and this method captures more data on genomic context, particularly mobile element sequences. However, this method is computationally expensive and time-consuming.
- **Read-based methods** work by aligning sequencing reads to the reference database. This method is fast and less computationally demanding. This approach is limited by the completeness of the reference database and nearby genes and genomic context may be missed due to spurious mapping.

Each approach has its own advantages and disadvantages depending on the need for the analysis and it is not clear if one approach is superior to the other. The choice of tool is likely to be a trade-off based on available resources and the objective of the study.

Sequence annotation involves extraction of biological information from genomic or metagenomic contigs typically to identify protein-coding regions and associated regulatory sequences [196]. This process is important to enable comparison against AMR databases using pairwise alignment between the query sequence and AMR reference sequences to characterise these contigs [133].

4.4.2 Analysis of bacterial WGS data

There are different approaches which may be employed when analysing WGS data. Comparing the assembled genome with reference strains facilitates many different inferences, such as pathogen identification, high-resolution strain typing, and prediction of important phenotypic characteristics, such as virulence or antimicrobial resistance [151]. Each step—assembly, strain typing, phenotyping, and clustering—requires different bioinformatics tools that must be harmonised into a consistent workflow [151]. Many of the tools discussed have been developed to identify antibiotic resistance in bacteria and laboratory and bioinformatic requirements for other pathogens may be different.

Analysis for AMR surveillance. To identify AMR, analysis of NGS data is dependent upon curated AMR databases linking the genetic determinant to the antimicrobials they confer resistance against [133]. Currently there are multiple different databases available to perform this analysis (Table 11, Chapter 5). These databases may be generalised or specialised and will vary in terms of the scope of resistance mechanisms that they cover and in the type of information they provide. Two databases used for sequence analysis widely described in the literature are the Comprehensive Antibiotic Resistance Database (CARD), a database that aims to collect information on a wide range of AMR genes and Resfinder, a web-based and standalone tool for detecting AMR genes from sequenced or partially sequenced bacterial isolates. Unlike CARD and other AMR databases, Resfinder requires contigs as an input file with a focus on curation of mobile genetic elements demonstrated to be horizontally transferred between bacterial species. Frequently these databases and associated tools will be limited to a set of specified pathogenic microorganisms. Species specific databases can be useful when considering that some pathogens will have intrinsic resistance to some antimicrobials. Additionally, species-centric databases enable rapid and effective curation of new AMR genes and chromosomal mutations.

A key challenge is that tools within AMR databases lack efficient and sustainable curation pipelines [133]. Frequently, databases will only receive active maintenance for a few years before becoming outdated. Curation efforts are further hindered by variation in nomenclature resulting in differences in gene names and synonyms across databases, limiting comparability. A number of different sequence identity-based systems exist for assigning new resistance genes with different cut-offs that may not be in consensus with the reference. Establishment of a universal database for AMR interpretation could help to resolve some of the complexities these challenges introduce around application of bioinformatics tools to analyse sequence data for AMR surveillance [10]. Further challenges relate to the difficulty of distinguishing assembly errors from biologically relevant genetic changes when comparing sequencing data to AMR database reference sequences. The challenges associated with AMR databases are discussed in Chapter 5.

AMR resistance genomic data represents an expanding data source, so an active data sharing approach to enable regular updates to AMR databases is critical to the refinement of these tools. There are also research efforts underway to explore the use of machine learning to improve antimicrobial resistance prediction using genomic, transcriptomic, phenotypic and other information available about organisms. While some studies have shown proof of principle of machine learning approaches [197] these techniques are currently not ready for mainstream use [198]. Challenges include difficulty in obtaining genome sequencing data with the associated phenotypic information, availability of data to develop and then train models, the representativeness of the data in terms of the species and genetic diversity covered and determining which machine learning approaches are most suitable in particular

contexts. There is also interest in the commercial sector e.g. Ares Genetics (an OpGen Group Company) is developing DNA-based drug response prediction using a machine learning based approach [199]. GenomeKey, part of the CARB-X AMR initiative (Appendix 8.5), is developing a machine learning tool to diagnose bacterial infections and determine the resistance profile within hours of receiving a sample [200].

Pathogen typing and comparative genomics. Genomic surveillance has been identified as one of the key areas where genomics can reduce the complexity of existing laboratory workflows, such as replacing complicated serotyping protocols for *Salmonella* with WGS [38, 135, 201]. Comparative genomics uses inference of the phylogenetic relationship between bacterial strains and provides insights to enable tracking of outbreak sources and for the identification of clonal strains [182]. Mapping based phylogeny compares sequences using a pseudo-alignment where each sample has a base relative to the reference sequence to identify differences in the sequence and generate a maximum likelihood tree.

When monitoring intra-clonal variation, for example in an outbreak situation, core genome multi-locus sequence typing (cgMLST) data is typically used. Core genome MLST is defined as the set of genes that are found in nearly all strains of a species to determine genetic relatedness [202]. Defining the genes which comprise the core genome requires consensus for any given species. However, achieving consensus has been challenging and a cgMSLT has only been defined for a limited number of bacterial species. In addition, some research has explored the use of whole genome (wg)MLST, using both the core genome and accessory genes from plasmids, for monitoring transmission dynamics, analysing both the core and accessory genome; however, using additional genomic data from a wgMLST does not provide any additional sensitivity over the cgMLST [201].

Interpretation. Interpretation of AMR surveillance to determine resistance or susceptibility to different antimicrobials involves the comparison of genotypic predictions with data from AST, epidemiological and clinical data to confirm AMR [4]. Critically, as new AMR genes are identified by linking together these data, there needs to be an effective reporting mechanism. AMR surveillance will be enabled by the availability of publicly accessible databases curated to describe causes of AMR and associated phenotypic and clinical data. Wider implications of sequencing data include the ability to develop specific, rapid real-time PCR tests to assist in the management of patients, for example, a rapid differential diagnostic assay against non-typhoidal *Salmonella* and azithromycin resistant infections in the UK as an in house assay [38]. Sequencing data underpins the development of reliable molecular diagnostics which have the potential to simplify interpretation to support AMR surveillance in a wide range of settings.

Considerations for AMR surveillance. There is an inherent bias of databases towards human-associated organisms reflected in prediction outputs, so choosing the most appropriate databases to compare assembled contigs is important. Additionally, many tools have been developed to interrogate AMR in specific pathogens. For example, Pathogenwatch offers a web-based platform for AMR analysis and phylogeny generation of *Campylobacter*, *Klebsiella*, *Neisseria gonorrhoeae*, *Staphylococcus aureus*, and *Salmonella* Typhi [203]. This tool is advantageous for those wanting to run consistent pipelines with this online platform offering capacity without requiring more in-depth bioinformatics expertise.

The challenge is that the variability in databases and the level to which AMR gene identification has been interrogated in different pathogens impacts on the interpretation of sequencing data. While WGS data has the potential to be interrogated over time with the identification of new AMR genes, this is not possible with targeted methods and as a result requires continuous redevelopment. For metagenomics approaches, this is further complicated by the volume of data generated and bias in databases will significantly limit interpretation of results (discussed further in Chapter 6).

Additionally, many of the challenges for AMR surveillance relate to bioinformatics and WGS analysis including data collection and integration, sequence analysis and training [203].

These challenges are not unique to AMR surveillance, but shared for all NGS workflows. Support for implementation of sequencing and associated infrastructure, including bioinformatics is crucial for sequencing to become more globally representative for monitoring AMR. Good data sharing networks will enable greater clarity of the state of play of AMR globally. These reference databases and availability of sequencing data with associated metadata will be a core element when determining the concordance of antimicrobial sensitivity testing and genome-derived resistance, as well as identifying novel high-risk clones and discovering novel mechanisms of resistance [48].

The Centre for Genomic Pathogen Surveillance affiliated with the University of Oxford, UK are attempting to overcome these barriers by developing a suite of protocols and bioinformatic tools to enable WGS and AMR surveillance [201, 203]. An example of a tool developed is Epicollect5, a free-to-use mobile data-gathering platform including a form builder and geotagging of collected data. They have further tools developed for data visualisation and genomic epidemiology, AMR prediction and cluster analysis, and data integration pipelines. Many of these tools have been funded and developed in collaboration with and for use by multiple public health organisations including the US Centre for Disease control and Prevention (CDC), ECDC and UKHSA.

AMR surveillance in the commercial sector. There are companies working to curate data on AMR genes to enable development of commercial products to identify AMR genes and inform decision making of the product users. These may have utility for AMR surveillance if developed with this purpose in mind. There are already molecular diagnostic solutions for AMR surveillance [7].

One company working to curate AMR genes is OpGen who have developed an AI powered DNA testing method combining NGS with ARESdb – a proprietary database containing thousands of whole genome sequences from clinical isolates collected from more than 200 centres globally. ARESdb claims to combine broad resistance profiles with high quality genetic information and to enable pathogen identification with up to >99% accuracy and antimicrobial resistance detection with up to more than 98%. This dataset has been used in research to develop a targeted panel using AMR markers curated in ARESdb (section 4.3.1) [186]. Limited information is provided describing which pathogens have been included in this database, which antibiotics or other antimicrobials have been surveyed, or which AST methods were used to quantify these phenotypic traits. Most AMR surveillance is performed as part of public health initiatives and without the ability to scrutinise this data, it will be challenging to evaluate the utility of these products as part of AMR surveillance initiatives. OpGen have just received FDA approval for Acuitas, an RT-qPCR panel to detect AMR genetic markers in Enterobacterales [204].

4.4.3 Proficiency testing

One leading challenge for AMR surveillance is how to ensure sufficient quality of WGS data. Integrating data in a One Health approach and ensuring sufficient data sharing between countries and key environments requires good quality data. Introduction of key principles of quality management systems including quality control and proficiency testing will be central to the goals of ensuring that any sequencing data generated is of a consistently high standard and quality. These principles need to be considered at all stages of the workflow, however, many of the challenges and disparities identified relate to inconsistencies and variability of bioinformatic analysis pipelines [81]. Introduction of standard metrics and minimum thresholds for evaluating raw sequencing data will ensure that all data generated has sufficient coverage to reliably detect genes and variants, particularly where the goal is to make genotypic-phenotypic predictions [4, 50, 194]. Ultimately, where the sequence data does not correlate with phenotypic resistance or susceptibility, agreement on how to indicate and report these “indeterminate” results will be important for ongoing research and curation of AMR genes in databases.

One of the leading demonstrations of implementation of proficiency testing is in laboratories in the US PulseNet and GenomeTrakr networks (see Section 2.6). These two networks have attempted to standardise the collection and analytical requirements for genomic data for foodborne pathogen surveillance (*Salmonella enterica, Listeria monocytogenes, Escherichia coli* (STECs), and *Campylobacter*). In 2017, they started harmonising their respective proficiency test exercises agreeing on distributing a single strain set and followed the same standard operating procedure for genomic data collection running jointly coordinated annual proficiency exercise. This includes the establishment of minimum quality requirements set forth by the quality assurance programmes of PulseNet and GenomeTrakr for high-quality SNP analysis [81].

All participating laboratories undertake proficiency testing using the Gen-FS harmonised SOP which specifies all stages of the NGS workflow for DNA extraction, library preparation and DNA sequencing [81]. The GenomeTrakr proficiency testing (GTPT) was designed to assess the performance of participating laboratories and to help the FDA team coordinating the effort identify areas for improvement (e.g. sequence quality, data transfer, following an SOP and communications).

Beyond the establishment of proficiency testing within AMR surveillance networks, the CDC has taken an active interest in trying to promote implementation of NGS workflows including providing guidance on implementation of quality management systems for NGS [205]. Additionally, a global consortium, Public Health Alliance for Genomic Epidemiology (PHA4GE) has been established to translate standards developed by the Global Alliance for Genomics and Health (GA4GH) to the microbial world that is actively working to establish consensus standards, document and share best practices, improve the availability of critical bioinformatic tools and resources, and advocate for greater openness, interoperability, accessibility and reproducibility in public health microbial bioinformatics [67]. PHA4GE have a number of objectives, including the establishment of an open source, standards-driven bioinformatics platform/ecosystem for public health. This resource is under development.

4.5 Resources for building workflows

Application of sequencing has been diverse with different applications of technologies and workflows to produce sequencing data to support AMR surveillance efforts. Within national and international surveillance networks laboratories have been using their own protocols. There are a number of stakeholders interested in implementation of sequencing technologies for AMR surveillance including national and international public health institutions, research institutions and commercial providers. There is not yet consensus around the best sequencing workflow to use or the standards and quality controls necessary to ensure sufficient quality to ensure comparability of generated data. Available sequencing platforms are able to generate sufficient quality data and in reality, much of the complexity stems from the process of analysing and interpreting sequencing data.

Efforts to develop standards for harmonisation of NGS workflows have been variable and there are no internationally recognised standards for QC of WGS data for pathogen typing and AMR detection. Key organisations who have been developing sequencing protocols, quality standards for WGS data and supporting implementation of quality management systems include the US Centres for Disease Control and Prevention (CDC), Danish Technical University (DTU) and NIHR Global Health Research Unit (GHRU). These organisations have significant experience implementing WGS for AMR surveillance and as a result may be valuable partners for identifying critical challenges, refining existing protocols and establishing international standards. In the US, proficiency testing as part of the PulseNet and Genome Trakr networks has encouraged some harmonisation with all labs using the Gen-FS harmonised SOP for all stages of the NGS workflow [81].

Agencies and institutions have published some SOPs and guidance for developing NGS SOPs [50, 81, 176-178]. Mostly these SOPs are from cultured isolates, using Illumina

Nextera XT library preparation and the Illumina MiSeq platform. Additionally, not all of these workflows detail the dry laboratory part of the workflow including QC parameters, assembly method, AMR gene and point mutation prediction and any additional analyses [50]. Where WGS has become the predominant method for pathogen typing, AMR gene identification and prediction, such as for tuberculosis or Salmonella, extensive workflows have been made available [38, 135]. There is also work to establish standard validation strategies for microbial WGS bioinformatics workflows [206].

The Centre for Genomic Pathogen Surveillance in the UK have also produced protocols for: Genomics DNA isolation, DNA library preparation, library normalisation and pooling, WGS and DNA sample transport; as well as bioinformatics protocols and training [207]. This training programme has multiple outcomes and focuses on different types of expertise including the ability to use command line and understanding of genomics [203]. These protocols have been developed to consider the following pathogens: *Acinetobacter baumannii, Escherichia coli, Klebsiella pneumonia, Staphylococcus aureus, Neisseria gonorrhoea, Salmonella typhi and typhimurium, Non-typhoidal Salmonella spp., Pseudomonas aeruginosa, Streptococcus pneumonia* and *Vibrio cholerae*. One limitation of the bioinformatics protocols and training is the need to be competent in using command line to perform analysis and interpretation of findings. This is likely to limit the accessibility of this tool and is not unique to this platform.

5 Databases for AMR surveillance

Genome sequences of the pathogens described in this report are available in a number of publicly available databases. The most widely used are those covered by the International Nucleotide Sequence Databases (INSD) consisting of: NCBI GenBank; the DNA Databank of Japan (DDBJ); and European Molecular Biological Laboratory (EMBL); the Sequence Read Archive (SRA) is also included. Together these provide the principal repositories for DNA sequence data with 95% (705 out of 743) of all nucleic sequence data databases directly linked to or able to download nucleic sequence data from the INSD [208].

Information on sequence databases, including an overview of how they are operated and managed, can be found in the recent UN Convention on Biological Diversity report entitled *Combined study on digital sequence information in public and private databases and traceability* [208]. While sharing of sequence data via databases can facilitate research across different specialisms and geographies, there is currently inconsistency in the use of databases between different national and international organisations and between research groups. Sequence databases have different data sharing policies and are also capable of storing different forms of the data, such as:

- Unprocessed, raw electrical signal files
- Raw reads
- Processed data
- Metadata
- Short reads
- Genome assemblies

Submission of data is voluntary but researchers are highly encouraged to share data. Selection of which database to submit data to can be determined by a range of factors, including:

- The regulatory framework
- Sequencing data available
- Preferred form of data to upload
- Familiarity with the database

Each of these factors will have an impact on the ability of researchers and public health authorities in different jurisdictions to optimise use of pathogen sequence data.

The data deposited in public sequence databases, while comprehensive, can contain more errors than curated, but more limited, databases such as the FDA-ARGOS or the FDA Reference Viral database (RVDB) [209]. The genomes available in the FDA-ARGOS database meet the quality metrics defined for providing reference-grade genomes for regulatory use. The availability of high quality reference genomes is important to support alignment of sequence data which facilitates accurate clinical interpretation of results. This is particularly important when considering unbiased metagenomic NGS – if a pathogen(s) does not have a universally accepted reference genome for metagenomic NGS it is challenging to compare assay performance between different laboratories. These databases are also continuously being updated, in order to correct misannotations and increase the representation of organisms – both in terms of numbers (and quality) of sequences per species and also in terms of the number of species that have sequences available.

5.1 Databases for AMR sequence data

There are a wide variety of sequencing databases and resources available that are relevant to AMR (Table 11). Many overlap in terms of their coverage and vary in terms of the type and scope of information that they contain. These databases can be:

- Broader in scope in terms of pathogen species and AMR genes

- Specific to a particular AMR gene or group of genes
- Specific to a pathogen or host
- Suitable for different types of sequencing data e.g. WGS, metagenomics, targeted sequencing
- Integrated with custom bioinformatics analysis, pipelines and/or annotation tools (Section 4.4)
- A resource that brings in data from other databases to provide a central source of information for a specific purpose.

The Joint Programming Initiative on Antimicrobial Resistance, JPIAMR, curates a list of AMR data platforms [210].

Table 11. Examples of available AMR reference databases. Adapted from [211].

Database	Description	Link	Status
General databases			
CARD	Ontology-based database that provides comprehensive information of AR genes and their resistance mechanisms Currently contains >2,200 protein homologues and includes a curated set of resistance-conferring chromosomal mutations in protein-coding genes	https://card.mcmaster.ca/	Active; launched in 2013; updated monthly
Resfinder	Collation of AR genes involved in HGT events	https://cge.cbs.dtu.dk//services/ResFinder/	Started in 2012; regular updates until last update in February 2019 Website currently live
ResfinderFG	Collection of resistance gene variants identified in multiple functional metagenomics studies	https://cge.cbs.dtu.dk/services/ResFinderFG/	Last update in November 2016 Website currently live
Resfams	A profile HMM-based curated database confirmed for AR function	http://www.dantaslab.org/resfams/	Last update in January 2015 Website currently live
MEGARes	Collation of multiple databases (CARD, ARG-ANNOT and ResFinder) to avoid redundancy between entries For high-throughput screening and statistical analysis	https://megares.meglab.org/	Last update in December 2016 Website currently live
NDARO	Collated and curated data from multiple databases (CARD, Lahey, Pasteur Institute β-Lactamases and ResFinder) Contains 4,500 AR sequences	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA313047	Started in 2016 Website currently live
Mustard	Resource containing 6,095 AR determinants from 20 families, including curated sets of AR genes identified in functional metagenomics studies	http://mgps.eu/Mustard/	Last update in November 2018 Website currently live
FARME database	Curated set of microbial sequences functionally screened to confer resistance in various functional metagenomics studies of different habitats	http://staff.washington.edu/jwallace/farme/	Last update in 2017 Website currently live
SARG (v2)	Hierarchical structured database derived from ARDB, CARD and NCBI-NR database Contains >12,000 AR genes; also includes profile HMMs for 189 AR genes subtypes	http://smile.hku.hk/SARGs	Active Website currently live
BLDB	Manually curated database for AR enzymes classified by class, family and subfamily	http://bldb.eu/	Last update in November 2018 Website currently live
CBMAR	Database that identifies and characterizes novel β-lactamases on the basis of Ambler classification	http://proteininformatics.org/mkumar/lactamasedb/	Last update in September 2014 Website currently live
Species-specific databases			
MUBII-TB-DB	Database of mutations associated with AR in *Mycobacterium tuberculosis*	https://umr5558-bibiserv.univ-lyon1.fr/mubii/mubii-select.cgi	Last update in December 2013 Website currently live
u-CARE	User-friendly, comprehensive AR repository for *Escherichia coli*	http://www.e-bioinformatics.net/ucare	Last update in 2016 Website currently live
KlebNET-GSP	A unified genomic surveillance platform with tailored analytics for the *Klebsiella pneumoniae* species complex	https://klebnet.org/	Launched January 2022

5.2 Current challenges for sequencing databases

The variety around the form, format and function of AMR sequencing databases presents a number of issues that will need to be considered:

Recency of data. Databases are continuously under development and are regularly updated with new knowledge on resistant genes and strains. The frequency with which databases are updated and the degree of active curation of data varies. Some databases become inactive and it is valuable to check the latest lists (for example on JPIAMR) of databases that are available and which are most suitable for a particular purpose. Due to the circumstances under which some AMR sequencing databases are established, for example through a research project attached to a specific source of funding, databases can become inactive or no longer curated. In order to optimise future use of sequencing data for AMR surveillance, consideration needs to be given as to the most appropriate mechanisms to ensure the ongoing accuracy, timeliness and longevity of databases.

Data entry. Sequence data can be in the form of reads or genome assemblies. Each have their own advantages and disadvantages. Sequencing reads can provide coverage estimations, and be assembled into contigs that can then be annotated.

Metadata. Improvement in the metadata quality and what is collected with the sequencing data is needed, for example when available MIC with each AMR determinant will help understand the biological context for interpretation of a result. Comparisons and analysis with more species and tools maybe required and could be useful in better interpreting results. Additional metadata such as source of the sample as well as site and location where they originate from can assist in wider understanding of the importance of the AMR identified.

Bioinformatics. A large number of bioinformatics tools have been developed to detect AMR determinants in genomic data (section 4.4). They differ in supported inputs, search algorithms, data parameters, underlying reference databases and output formats. There are particular challenges with the communication of the information generated by sequencing workflows to the users that will act on this information, such as clinicians or public health officials. Many databases are command line – this means there is a requirement to have specialised programming or bioinformatics experts that are capable of carrying out the analysis as well as have an understanding the outputs and how to interpret them.

Variety in outputs. The outputs (the analysis for AMR within a sequenced sample) from many of the AMR databases are not comparable to each other. Some will report closest hit, where others report best estimates. The level of point mutation detection information is not available in all database outputs and will only report the presence of whole genes. The description of the genes in reports is also variable between the databases. Lack of standardisation in the reporting of AMR gene detection greatly hinders the comparison of results across the public health sector. The myriad of options available for this purpose highlights a critical interoperability problem. For this reason databases such as hARMonization have been developed that allow for a single entry point that can run the genomic data in an extensive list of AMR databases, and provide a single output for each of the databases used [212]. Its development included a standardised specification to improve data interoperability. Additional interpretation of results is currently still needed once the outputs from the various sources has been generated.

Reference genomes for pathogens of interest and AMR. AMR genomic data analysis relies on being able to use a reference that is well annotated for the biological mechanism explaining the resistance. Therefore, access to appropriate, well annotated and curated databases with suitable reference genomes are needed. As AMR is changing and pathogens adapt the recency of these reference genome annotation is important.

Genotype-phenotype correlations. For some organisms, there is high correlation between genotype and phenotype, for example *Campylobacter*, *Salmonella* and *E.coli*, with the most recent analysis indicating a >99.7% consistency [213]. For other pathogens, such as *Pseudomonas aeruginosa* [197], it is not as feasible to carry out a phenotype-genotype correlation due to high levels of phenotypic plasticity shown by the pathogen in response to environmental cues, including antibiotics. Predicting phenotypes is challenging from pathogens since finding a resistance determinant does not always mean expression, and the impact of gene expression will differ depending on the clinical context. The environment that a bacterium is found in – such as a biofilm or a chicken for consumption versus a live chicken – can cause unique gene expression and which will have different clinical implications. Current tools are not able to address these challenges. Finally there are still many unknowns, including underlying resistance genes that have not yet been described. On a mechanistic level, for some bacteria a single mutation can lead to resistance whereas in others a number of acquired genes may result in more significant AMR. In some circumstances experimental validation of resistance under different circumstances could provide useful information for the additional annotation of sequences, and support understanding of the changing biological mechanisms of resistance.

Database standardisation. Lack of standardisation across databases is an issue. These range from data collection and entry into the database, storage, the analysis pipeline, to the output and how the data are used. All of this affects the downstream interpretation. There is a need for harmonisation of AMR resources to enable users to compare and contrast data, which can help downstream data use and interpretation.

Tool uses. There needs to be alignment between databases, their tools and the goal of the analysis. The analysis of sequencing data, including tool use, will vary depending on the purpose of the analysis. High accuracy will be required if the analysis is being used to support clinical decision making. Different tools will be needed to support detection of the broader presence of AMR genes in the environment, or to enable discovery of novel resistance genes. Being able to select the appropriate tool also requires an understanding of the tools as well as the purpose of the analysis (Table 12; also section 4.4).

Table 12. Sequencing-based tools for antimicrobial resistance detection. Adapted from [211].

Name	Description	Accessibility	Year	Link
Assembly-based tools				
Resfinder	Tool for detecting acquired AMR genes from sequenced or partially sequenced bacterial isolates	Web and /or standalone	2012	https://cge.cbs.dtu.dk/services/ResFinder/
RGI (resistance gene identifier)	Pairwise comparison of query sequence with CARD (comprehensive antibiotic resistance database). Uses curated AMR detection models to predict intrinsic resistance genes, dedicated resistance genes and acquired resistance from mutations in drug targets	Web and /or standalone	2015	https://card.mcmaster.ca/analyze/rgi
ARIBA	Tool for rapid AMR genotyping directly from sequencing reads using curated public databases	Standalone	2017	https://github.com/sanger-pathogens/ariba
NCBI-AMRFinderPLUS	Tool for identification of acquired resistance genes using NCBI's curated AMR database and curated collection of hidden Markov model	Standalone	2018	https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/
Read-based tools				
SRST2	Tool for direct mapping of reads to curated AMR databases	Standalone	2014	https://github.com/katholt/srst2
ShortBRED (Short, better representative extract dataset)	Tool to profile protein families in the metagenomic data using short peptide marker sequences	Standalone	2015	http://huttenhower.sph.harvard.edu/shortbred
SSTAR	Tool to identify known, putative new alleles and truncated versions of existing AMR genes from WGS data	Standalone	2016	https://github.com/tomdeman-bio/Sequence-Search-Tool-for-Antimicrobial-Resistance-SSTAR-
PATRIC (Pathosystems Resource Integration Center)	Unique resource for studying AMR	Web	2016	www.patricbrc.org

Database selection. Determining which database or tool is of use requires consideration of the following features:

- What pathogens, AMR genes and drugs are covered in the database?
- Is it regularly curated / updated?
- What are the inclusion criteria for AMR determination, AMR genes and point mutations?
- Are only full-length genes included? This is important for identification of best hits.
- How well curated are the genes?
- How are the gene symbols reported? There is variability in gene symbols, confusing comparisons between tools.
- Are there links to the literature or laboratory experiments that can allow for back tracking of information?
- Are possible phenotypes reported?

The FAIR guiding principles are not always used in practice [214]. Moving forward there is a need for standardisation of AMR databases, inclusion of metadata in AMR databases, specialised databases and tools for species, and harmonisation of AMR resources.

5.3 AMR surveillance databases

AMR surveillance databases are distinct from sequence databases. They are required to collect data and information that is relevant for surveillance and traditionally do not include sequence data. Numerous AMR surveillance databases are available and include:

- European Surveillance System (TESSy) platform which supports the European antibacterial resistance surveillance network (EARS-Network) [215] as well as the tool Surveillance Atlas of Infectious Diseases, ECDC-ATLAS [216] a tool that interacts with the latest available data about a number of infectious diseases. The interface allows users to interact and manipulate the data to produce a variety of tables and maps. The information contained in the dataset provided through ATLAS is made available through TESSy.
- Indian Council of Medical Research (ICMR) initiated the design and development of ICMR's Antimicrobial Resistance Surveillance system (i-AMRSS) [61]
- GLASS data capture on the WHO Global Health Observatory [217]
- SENTRY Antimicrobial Surveillance Program (SENTRY Program) is one of the longest running antimicrobial surveillance programmes in the world which ran from 1997-2016 [218, 219].

Others are attempting to bridge the gap between being a sequence database and surveillance database:

PulseNet has been the backbone in the detection and sharing of outbreaks and in the past, PulseNet used pulsed-field gel electrophoresis (PFGE) and multiple locus variable number tandem repeat analysis (MLVA) for DNA fingerprinting. However, the current gold-standard method for PulseNet is WGS to generate DNA fingerprints. PulseNet uses WGS to subtype *E. coli* (O157 and other Shiga toxin-producing *E. coli*), *Campylobacter*, *Listeria monocytogenes*, *Salmonella*, *Shigella, Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Cronobacter* isolates [220]. PulseNet International consists of a network of 83 countries.

Global Microbial Identifier. The genomic epidemiological database for global identification of microorganisms is a platform for storing WGS data of microorganisms, for the identification of relevant genes and for the comparison of genomes to detect outbreaks and emerging pathogens [221].

NCBI Pathogen Detection is a new database that is attempting to address many of the issues highlighted in this chapter. It integrates bacterial pathogen genomic sequences originating in food, environmental sources, and patients. It identifies AMR, stress response, and virulence genes found in bacterial genomic sequences. This enables scientists to track the spread of resistance genes and to understand the relationships between AMR and virulence. It quickly clusters and identifies related sequences to uncover potential food contamination sources, helping public health scientists investigate foodborne disease outbreaks [222].

European Surveillance System (TESSy). Surveillance data, including some basic typing parameters for the isolated pathogen, are reported by ECDC Member States to TESSy [93]. One of the key objectives of ECDC is to improve and harmonise the surveillance system in the EU and increase scientific knowledge of aetiology, risk factors and the burden of food and waterborne disease. In 2012, ECDC initiated enhanced EU-level surveillance by incorporating molecular typing data into reporting of foodborne pathogens. Since 2019, countries have been able to report WGS data to TESSy for *Salmonella* and *Listeria monocytogenes*. The overall aims of integrating molecular typing data into EU-level surveillance are to:

- Foster the rapid detection of dispersed international clusters/outbreaks
- Facilitate the detection and investigation of transmission chains and relatedness of strains across EU/EEA and contribute to global outbreak investigations
- Detect the emergence of new and/or evolving pathogenic strains
- Support investigations to trace the source of an outbreak and identify new risk factors
- Aid the study of particular pathogen's characteristics and behaviour in community of hosts.

Molecular typing-enhanced surveillance gives Member State users access to EU-wide molecular typing data for the pathogens included. It also provides users with the opportunity to perform cluster searches and cross-sector comparability of EU-level data to determine whether isolates characterised by molecular typing at the national level(s) are part of a multinational cluster that may require cross-border response collaboration.

6 Culture-free sequencing

The vast majority of sequencing for AMR surveillance is currently performed on pure cultured isolates of the pathogen under investigation. The advantage of this approach is that it reduces host contamination and concentrates the DNA of the target pathogen being sequenced, which allows more high-quality sequence information to be produced. In addition, as phenotypic AST using pathogen culture is considered the gold-standard tool for making diagnostic decisions, cultures are then available for sequencing. However, there is now a trend for clinical diagnostic laboratories to move towards non-culture based phenotypic and molecular methods to enable more rapid, and in some cases more accurate pathogen identification than current culture-based diagnostics. If these methods become widely used there could be a shortage of cultures from diagnostic laboratories available for WGS based surveillance. As described earlier in the report, the majority of surveillance data is currently obtained from clinical settings, and so a lack of WGS of clinical isolates from diagnostic settings could lead to a lack of AMR surveillance data. New infrastructure and pathways may be required to allow samples to be cultured specifically for surveillance purposes, which may act as a barrier to surveillance [223].

Techniques to enable sequencing directly from samples, without the need to culture, are now under development. If successful, they could provide a number of advantages for surveillance purposes, over culture based WGS:

- Remove or reduce the requirement for pure cultures for WGS based surveillance, as diagnostic laboratories shift to alternative methods for AST
- Provide an alternative option in situations where pathogen culture facilities are unavailable or limited
- Provide information for surveillance more rapidly, which is important for situations such as outbreaks
- Permit surveillance of pathogens which are impossible or difficult to culture
- Reflect the natural genetic diversity of the pathogen more accurately, as culturing may bias towards certain serotypes [224]
- If a non-targeted metagenomic approach is used, allow surveillance across multiple microorganisms in a sample simultaneously

In addition, there may be opportunities in future for WGS to be used to provide diagnostic information, instead of current culture-based phenotypic AST methods. The faster turnaround times could make culture-free sequencing a more attractive option for this than current culture-based WGS methods, if the WGS data produced from culture-free methods is proven to be an accurate and reliable predictor of phenotype. Using the same methods to produce information for diagnostic and surveillance purposes could help make resource and cost-savings, whilst streamlining the process of obtaining surveillance data. The utility of this approach would be highly context dependent based on the pathogen, its prevalence and the level of WGS-based surveillance required.

Methods used to obtain WGS data without the need for pathogen culture can broadly be divided into two main categories – metagenomics and targeted approaches. Metagenomic sequencing methods are inherently culture-free and refer to the sequencing of all genomic material in a sample. Depending on the quantity of pathogen genetic material within a sample, the genetic data produced can then be analysed to obtain high coverage of certain pathogen genomes within a sample. In contrast, targeted approaches make use of methods designed to detect and enrich the entire genome of a pathogen of interest directly from the sample, which can then be sequenced to provide WGS data.

However, while WGS data is useful for AMR surveillance, metagenomic and targeted culture-free methods are currently most often used when the aim is not to provide whole genome coverage of specific pathogens, but to provide other types of genetic information. For example, most metagenomic methods have been developed for surveillance situations

where the aim is to monitor the presence of AMR, specific pathogens, and overall microbial diversity in samples containing potentially thousands of microbes. Similarly, targeted methods are often applied in the context of gene panels, where the aim is to sequence specific known genomic regions of interest to allow key microbes and AMR features to be identified, and not to produce WGS data. Targeted NGS panels are already available for the sequencing of specific pathogen typing and AMR relevant genes directly from clinical samples (section 4.3).

In the sections below, the applications of different culture-free sequencing methods for surveillance purposes are discussed, alongside their current readiness to be applied, with a focus on metagenomics as a promising new sequencing tool. Many techniques are only used in very specific contexts or are still at the research and development stage, with their utility yet to be proven.

6.1 Metagenomics methods

Metagenomics can be defined as the sequencing of all the genomic material present in a mixed community of organisms [225]. It is commonly used in research settings to sequence microbial communities, in order to characterise the total microbial diversity of the sample as well as key features such as AMR genes. For the results to reflect the natural genetic composition of the sample as accurately as possible, metagenomics typically requires direct-from-sample sequencing, as opposed to the use of cultured samples. Whilst metagenomics methods aiming to sequence all the genetic material in a sample could be applied to cultured samples, many would consider this not to be true metagenomics and instead be 'unbiased' sequencing.

While in theory metagenomics refers to the sequencing of all genomic material, in practice the approach may also aim to sequence particular types of microbes, such as just the bacteria or viruses present in a sample, rather than all microbes. In addition, other types of genetic material in a sample, such as human or animal host DNA, may be depleted in order to improve results for the microbes of interest.

Methods considered to fall under the umbrella of metagenomics belong to two main categories: shotgun metagenomics and targeted amplification methods:

Shotgun metagenomics is used to sequence all DNA and/or RNA in a sample, and therefore can be considered 'true' metagenomics. If enough sequencing is performed on a sample (i.e. to a sufficient sequencing depth), it can be possible to get enough genetic information to not only identify the types of microbes present but assemble partial to whole genomes of specific microbes.

Targeted amplification methods focus on amplification of a common gene found in all organisms of interest, but which still has enough variability to allow different species to be identified. For example, amplification of the 16S ribosomal RNA gene is typically used to understand the bacterial diversity of a sample. This is sometimes described as a metagenomic method, but if we define metagenomics as the untargeted sequencing of the entirety of DNA in a sample, in reality it is not.

In this section of the report, non-targeted shotgun metagenomics will be discussed. Targeted 16S ribosomal RNA is discussed later in section 6.2 together with other targeted culture-free sequencing methods. There are several advantages and disadvantages of metagenomic sequencing for surveillance.

Advantages

In addition to the benefits of being a culture-free approach, metagenomics analysis has the following advantages:

- It can be used to detect multiple microorganisms and AMR genes within a sample at the same time, allowing detection of all pathogens and AMR genes present, as well as providing context on the microbial community and diversity within a sample
- As a non-targeted method, target enrichment strategies do not have to be developed for each species of microorganism, it is a pathogen agnostic approach
- An output with the same format is produced across multiple different sample types and/or pathogens, enabling easier comparison of different types of surveillance, facilitating a One Health approach
- Data can be shared electronically in a standard format, facilitating an international One Health approach to surveillance
- Sequencing data can be reanalysed when needed, for example if a new bioinformatics method becomes available, or to carry out retrospective investigations to determine the presence of new pathogens of interest. This provides an advantage over PCR based target specific tests which do not produce sequencing data for reanalysis, or NGS panels only assessing known regions of interest, which can be reanalysed but may not be as applicable to future scenarios e.g. analysis of novel genes or pathogens not originally targeted in the panel.

Disadvantages

- Shotgun metagenomics can be more expensive on a large-scale, compared to PCR testing and targeted methods to assess specific resistance genes which can be used at scale in a high throughput manner at relatively low cost
- Targeted methods are highly sensitive and specific, whereas the low prevalence of AMR genes in a metagenomics sample limits its specificity [226]. As data are generated for multiple organisms, sensitivity is also lower than for genomic analysis of single isolates
- Assembling a genome from metagenomics data is computationally expensive and likely to result in lower genome coverage than WGS of isolated microorganisms
- Sample composition can be easily affected by multiple parameters, from sampling strategy to DNA isolation and sequencing methods, making standardisation particularly important, especially if a key aim is to compare different samples. This may be challenging to achieve nationally and internationally
- For the purposes of pathogen genetic typing, it can be hard to accurately link resistance genes to the microorganisms from which they originate, especially if they are located on mobile genetic elements such as plasmids.

6.1.1 Basic metagenomics workflow for AMR surveillance

As with other sequencing workflows (chapter 4), metagenomics for AMR surveillance purposes involves steps related to sample collection and processing, library preparation, sequencing, library assembly and bioinformatics analyses. Depending on the protocol and depth of sequencing used, different types of information can be obtained, ranging from identification of microbes and AMR genes of interest, to partial and even whole genome assembly of microbes of interest present in the sample. This information can then be used to inform the intended surveillance purposes. The surveillance needs and infrastructure available will determine exactly what type and level of metagenomic sequencing to perform. For example, in some cases it may be useful just to track types of AMR genes present in a certain setting. In other cases, more detailed genomic information may be required to link AMR genes to specific microbes, to enable transmission dynamics of AMR genes between bacteria, or to track specific high risk antibiotic resistant bacteria.

Sample collection. A metagenomics workflow starts with sample collection. An overarching benefit of metagenomics is that samples can be directly or indirectly obtained from a wide range of settings to facilitate a One Health approach to surveillance:

- Humans (e.g. in clinical settings or in the wider 'healthy' population)
- Animals (e.g. in agricultural settings or in the wild)
- Meat and other food products
- Wider environment (e.g. in soil, rivers or urban settings)

The type of sample has to be chosen carefully to ensure it is suitable to reflect the microbial population of interest, and relevant to the intended surveillance purposes. For example, when monitoring farm animals, it may appear most accurate to take a rectal swab directly from the animal. However, this is more invasive and time consuming than taking a sample from faeces in the pen of the animal. Whilst the first strategy may be appropriate for surveillance purposes requiring individual animal level data and 'fresh' microbial samples, the second strategy may be more appropriate for broader surveillance purposes to get a general representative sample of a certain population of animals. Many other parameters, such as environmental conditions, may also affect the microbial composition of the sample and therefore the surveillance strategy. Robust and standardised sample collection procedures will be needed to allow comparisons to be made between multiple different sites either within or between countries.

Sample processing and DNA extraction. The processes by which the samples are processed and DNA extracted can influence DNA quality and quantity and therefore the overall results, and so for surveillance purposes should be standardised as much as possible. A common step for microbial metagenomics is to deplete the relatively high quantities of any host (animal or human) DNA in the sample, to increase the sensitivity of detection for microbial DNA. Different methods may also be required for different types of microbes (i.e. bacteria, fungi and viruses) and when analysis of microbes with RNA genomes is required. Other considerations when choosing a DNA extraction method for metagenomics surveillance include:

- Whether the method suitable for a range of sample types (e.g. faeces, sewage, clinical samples etc)
- How representative the method is of the 'true' content of genetic material in the sample
- Whether the method will produce sufficiently high DNA yield, quality and stability
- Whether there is the possibility for automation to allowing scaling up, if required
- What safety measures are required for potentially hazardous pathogens, as well as is there the possibility to avoid hazardous reagents (such as phenol or chloroform)
- How affordable and time efficient the method is
- Whether further optimisation of the method is required, as well as how best to perform quality assurance (e.g. by spiking known amounts of specific organisms into the sample)

Some studies have compared different sample preparation methods to support these types decisions, it is likely further studies will be useful as the field continues to develop [227].

Library construction and sequencing. As with other types of NGS, a range of library preparation methods and sequencing platforms may be used. The key feature of library preparation methods for metagenomic sequencing is that they enable non-targeted sequencing of all genomic material in the sample, as opposed to target amplification methods. The depth of sequencing required will then depend on the data analysis requirements of the sequencing (see below). Methods used to identify the genomic and taxonomic context of AMR genes and/or to identify novel types of microorganisms require a higher limit of detection and so need a greater sequencing depth, than methods which just aim to identify known pathogens or AMR genes in the sample.

Most metagenomics for surveillance studies have used more conventional and more established short read sequencing which typically has lower error rates, performed on Illumina or Ion Torrent platforms. However long read sequencing (performed on PacBio or Oxford Nanopore platforms) can be used to provide more genomic context despite higher error rates, and may become increasingly used in future. The portable nature of some Oxford Nanopore equipment such as the MinIon also makes it a useful tool in field settings where there is no current sequencing infrastructure. Metagenomics can be combined with other methods if needed to fulfil the specific data requirements of the study. For example, using long read sequencing in addition to short read sequencing to analyse a sample can help provide genomic context as well as high genomic resolution (section 4.6) [226].

DNA sequence analysis. As with all sequencing experiments, a range of bioinformatics tools (section 4.4) are used for quality control of the sequencing data produced, and then to analyse the data.

Read mapping: If the aim is to provide answers on the presence and absence of specific microbes or specific genes i.e. AMR genes, then a read mapping approach can be taken. In this case the sequencing reads produced are compared to known sequences in pathogen sequence databases, to identify the presence (or absence) of different microorganisms and AMR genes.

Assembly and binning: An assembly and binning approach can be used if the aim is to gain more information on the genomic and taxonomic context of the microbes sequenced, for example: which AMR genes are associated with which species of microbes; where the AMR genes are located on the genome, and does a pathogen have multi-drug resistance. This approach aims to assign reads to contigs, a series of overlapping DNA sequences that can be assembled to reconstruct a larger region of DNA, or partial or whole pathogen genomes. As well as providing more genomic context, this approach also helps to provide increased confidence that the pathogens identified are correct as there is a lower risk of false positives, and it can be used to identify novel taxa.

Metagenomic data can provide some particular challenges for sequence analysis:

- Some species or AMR genes may be present at very low quantities in the sample, below the limit of detection of the sequencing method. Spiking in known quantities of specific indicator DNA can be used to establish a limit of detection, so that it can be determined that anything not detected is either absent or below the limit of detection
- It can be challenging to use data to assess the quantities of different microorganisms, as some pathogens may produce more sequence reads than others, for reasons other than their abundance in the sample. For example, those with smaller genomes may receive a higher sequencing coverage, or the sequencing methods used may be biased towards specific genomic regions. Data analysis methods can be used to help overcome some of these challenges, but results should be interpreted with caution
- It is likely that many sequences detected cannot be assigned to a particular species or gene, due to limitations in the sequence data base (Chapter 5), or there are not enough reads present to support decision makgin. Similarly, for assembly and binning approaches, a lack of read assembly could be mistakenly interpreted as absence of the microorganism
- It can be very difficult to confidently assign AMR genes to particular organisms, even using an assembly and binning approach, especially if there is a high diversity of microorganisms in a sample
- For whole genome assembly of specific microorganisms, metagenomics is a comparatively less sensitive technique than culture based and/or targeted WGS of specific microbes, as not all reads produced will belong to one microbe. A higher depth of sequencing can make it harder to assemble an entire pathogen genome

- Assembly and binning methods required to assemble more complete genomics sequences are computationally expensive.

Data Interpretation. The data obtained can then be used to answer specific questions, which requires various methods of statistical and epidemiological analyses. Questions that can be asked include:

Those related to microbial community composition (distribution analysis) e.g.:

- The abundance of different species or specific genes such as AMR genes
- The relative diversity of different species
- The distribution of different species- how similar and dissimilar are they from each other?

Those related to the causes of the microbial composition (determinant analysis) E.g.

- Is there a correlation between AMR genes and levels of certain bacteria?
- Is there a correlation between epidemiological parameters (such as antimicrobial usage, sample site type, or economic indicators) and AMR and/or types of bacteria?
- Can antimicrobial susceptibility be predicted from the sample resistome (all AMR genes present in a sample) potentially using machine learning algorithms?

Answers to questions such as these can be used to inform and monitor the success of interventions to control AMR, for example limiting the use of certain antimicrobials in areas with a high level of specific AMR, or improving sanitation. Metagenomics is not the only tool capable of answering some of these questions – those relating to determinant analysis in particular could be answered using different NGS methods. This is relevant when the pathogen and resistance determinant(s) are known and can be targeted for monitoring. Metagenomics approaches are an option for extensive monitoring of a large number of organisms and AMR in a single sample/test.

6.1.2 Metagenomics for different surveillance purposes

Metagenomics can be used for a range of surveillance purposes, making it a useful tool for One Health strategies, as the AMR data produced is comparable between applications. The level and type of data that it is possible and/or useful to obtain through metagenomic sequencing will depend on the particular surveillance need. For example, for general monitoring of AMR in the environment or in the commensal microorganisms of healthy animal or human populations, it may be sufficient to obtain data to explain the microorganism and AMR gene diversity in the sample, without having to obtain more complete sequence data for particular microorganisms. Alternatively, for analysis of samples suspected or proven to contain AMR pathogens, for example from clinical infections or infected food supplies, the aim may be to obtain a near or whole genome sequence for that pathogen. This provides more detailed data for phylogenetic and epidemiological analysis to track outbreaks and inform control measures.

Metagenomics methods are in various stages of development and for the use of different types of surveillance, three of which are described in more detail below:

- Monitoring AMR in livestock and the food chain
- AMR environmental surveillance
- AMR surveillance in clinical settings

This is not an exhaustive list, and metagenomics could also be useful in other surveillance scenarios, for example monitoring of AMR in wildlife, or in healthy human populations.

The current and future use of metagenomics for these applications and others is determined not just by the availability of metagenomics protocols and tools, but by the extent to which surveillance more broadly is carried out in these different areas. For example, whilst the importance of environmental surveillance is recognised, this is a relatively new area of AMR

surveillance and currently very little used, though this is likely to change in future. In contrast, AMR surveillance in clinical and food production settings is much more established, with demonstrated utility for public health.

6.1.3 Metagenomic AMR surveillance in livestock and the food chain

Current status of livestock and food chain surveillance

High antibiotic usage in livestock when treating veterinary infections or for prophylactic use, as well as the use of other broad spectrum antimicrobials such as feed preservatives, is leading to increased numbers and types of antibiotic/antimicrobial resistance bacteria in animals intended for human consumption. Antibiotic resistance most frequently arises in commensal bacteria that can act as a reservoir, from which resistance genes can be acquired by zoonotic bacteria capable of causing food-borne infections in humans. Resistance can also arise in these food-borne bacteria directly [228].

Surveillance of AMR in both commensal and food-borne bacteria in farm animals and meat is internationally recognised as a priority, as part of a One Health approach to global AMR surveillance [228]. For example, in Europe it is mandatory to actively monitor AMR in zoonotic bacteria *(Salmonella* and *Campylobacter*) and indicator bacteria (*E. coli*) from healthy animals intended for food consumption, and the subsequent food produced [31]. This activity is coordinated by the European Food Safety Authority (EFSA). Together with the European Centre for Disease Prevention and Control (ECDC), they produce an annual joint report on AMR in zoonotic and indicator bacteria from humans, animals and food [229]. By taking a One Health approach to foodborne illness, AMR sequencing data of clinical cases of foodborne illness could be investigated to see if there is a link with AMR prevalence at particular food production or agricultural sites.

In addition, non-animal derived food products also have the potential to be a source of food-borne illness from AMR resistant microorganisms. Fresh vegetables and salad are often eaten raw, and studies have shown that these can be a potential, if rare, source of major antibiotic resistant food-borne bacteria [230].

Example: FARMS-SAFE project: Future-proofing Antibacterial resistance Risk Management – Surveillance and Stewardship in the Argentinian Farming Environment [231].

Dates: 01 August 2019 – 31 May 2023

Location: Argentina

Aim: To provide better surveillance information for AMR and antimicrobial usage, by exploring four key risk areas via an UK/Argentinian research consortium: animal disease as a driver of AMR risk; farm management practices that influence antimicrobial use; farm waste management as a driver of AMR risk; and risk-informed regulatory capacity for AMR. Genomic surveillance of AMR bacteria with potential to affect human health is being performed as part of the investigation into risk from farm waste, and AMR correlated with antimicrobial usage.

Anticipated impact: The results are expected to ‘inform policy making within Argentina, Latin America and the wider world. The research programme's team will also create a surveillance structure and train researchers who can continue to monitor AMR, antimicrobial usage and environmental contamination with antimicrobials and AMR into the future, and who can measure success and failure of strategies employed to reduce this risk’.

Why Metagenomics?

PulseNet International, a global laboratory network dedicated to bacterial food-borne disease surveillance, is aiming to implement WGS for multilocus sequence typing as its standardised method of choice [232]. In this context, the network is mainly focussed on

clinical isolates from humans suffering from specific food-borne infections, rather than monitoring the background level of AMR in the food supply. Whilst WGS can be used for AMR monitoring in healthy animal populations, and provides high genomic resolution, it is currently still dependent on isolates of specific indicator pathogens obtained from cultures. This means that it is likely detecting only a fraction of the original sample's resistome belonging to all the microbes, as well as potentially missing some hazardous pathogens [233].

Metagenomics methods could be a more representative and faster way of providing general AMR monitoring in 'healthy' animal populations, or other food sources. Results from several studies show that metagenomics data can represent AMR abundance in bacterial communities more accurately than current commonly used methods based on indicator species [233]. If a source was found to be contaminated with a pathogenic microorganism, either through reports of illness or routine surveillance, metagenomics could also be used to identify and provide more in-depth analysis of the pathogen of interest, as an alternative to culture-based methods. This may be particularly useful to enable rapid identification and typing of the pathogen where its likely identity is unclear, or it is not amenable to culture.

Current status of metagenomics for monitoring AMR in livestock and the food chain.

Metagenomics is not currently used for routine surveillance of AMR in healthy livestock in either HICs or LMICs. Phenotypic analysis of cultured samples is still used, with some use of PCR and NGS methods. In many countries, surveillance of AMR in agricultural and food production settings is still limited.

Several large-scale pan-country studies have demonstrated the potential of metagenomics and developed methods for livestock surveillance, especially for swine [233, 234], however these have taken place in a one-off research context rather than a routine setting. Further work is needed on ensuring quality control and reproducibility, and demonstration of validity, to support use of metagenomics as a routine surveillance tool. Evidence is also still being generated that will inform how to create an optimal AMR surveillance strategy in livestock.

Research has also been carried out into the use of metagenomics-based WGS from food sources that are at risk from microbial contamination. For example, a protocol has been developed to detect *Salmonella* species from lettuce leaves without the need for culture [235]. Following metagenomic sequencing of all microbial DNA, different bioinformatics approaches were developed to analyse genome markers and perform whole genome assembly for *Salmonella* species specifically, as well as perform taxonomic analysis of other species present in the sample. The method was sufficient to characterise *Salmonella* serovars and AMR genes, and provides an example of how a universal metagenomic rapid sequencing approach combined with tailored bioinformatics analysis could be used to identify bacterial contamination in food. This study was designed with surveillance in mind, however further work will be needed to determine if the techniques developed meets the requirements for AMR surveillance specifically, or if further characterisation by other methods such as WGS of cultured isolates would still be required.

Case study 1: Metagenomics analysis of the resistome of animals intended for the food chain [233]

Title: Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries

Year: 2018

Type of study: Research study

Aim of study: To understand the impact of differing levels of antimicrobial use (AMU) across different European countries on AMR genes in pig and poultry herds, and the subsequent risk of transfer to humans.

Project Summary: As part of the European Union-funded EFFORT project, over 9,000 animals were sampled in 181 pig and 178 poultry herds in 9 European countries, to generate herd level composite samples of over 5000 Gb of DNA [236]. These were sequenced using a metagenomics approach and the abundance, diversity and structure of the acquired pig and broiler resistomes investigated. An association between AMR gene abundance and national veterinary AMU was also analysed.

Sequencing requirements: Library prep: For pooled pig samples, the NEXTflex PCR-Free library preparation kit (Bioo Scientific) was used; for poultry samples, the minimal amplification-based KAPA Hyper kit (Kapa Biosystems) was used. For all samples, the Bioo NEXTflex-96 adapter set (Bioo Scientific) was used. Sequencing: In batches of roughly 60 samples, the libraries were multiplexed and the majority sequenced on the HiSeq3000 platform (Illumina), using 2 × 150-bp paired-end sequencing per flow cell. Analysis: FASTQ reads were analysed using the MGmapper tool, reads were aligned to the prokaryotic RefSeq genomes from the NCBI GenBank and AMR genes present in the ResFinder database.

Key findings and utility for surveillance

- The strategy used was able to detect and quantify over 400 AMR genes across the pig and poultry herds sampled. The livestock resistomes differed within and between countries, potentially due to levels of AMU, with between-country resistomes clustering according to the level and diversity of AMU, especially in pig herds.
- An association was found between crude levels of veterinary AMU and the abundance of AMR in each country. However treatment incidents data from specific farms was less correlated to AMU. Better AMU data and identifying reporting biases between countries would help clarify the association between AMU and AMR.
- The protocol used was not optimised for poultry faeces, meaning lower DNA yields and the requirement for a PCR step in the poultry samples. This emphasises the need for protocols which can be used across different sample types, as different library prep methods mean results cannot be accurately compared.
- Clinically relevant resistance genes were identified, including *mcr-1*, and $bla_{CTX\text{-}M}$, in poultry herds. The protocol was adequate for drawing broad conclusions about AMU and AMR. However sensitivity of metagenomics is still likely lower than phenotypic methods, meaning levels of important resistance genes could be underestimated.

Next steps

This was a research study and provided protocols and useful baseline information, showing this method can be used to evaluate approximate levels of AMR in agricultural settings, which appears to be at least partially influenced by AMU, and potentially by other country specific agricultural practices. This can lay the foundations for further research and generation of a surveillance strategy. For accurate estimation for specific AMR genes and pathogens though, phenotypic methods remain more appropriate.

Relevance to GLASS pathogens: Not pathogen specific, potentially relevant to all pathogens. Unless a further step of genome assembly and binning is taken, this read mapping approach cannot identify in which pathogens the AMR genes exist.

6.1.4 Metagenomics AMR surveillance in the environment, human and animal populations

Current status of environmental, human and animal population surveillance

AMR can arise in microorganisms found in our natural environment, for example those microbes populating soil and water. These microorganisms can act as a potential reservoir of AMR genes, which they may pass on to more harmful pathogen bacteria existing in similar environments. In addition, sampling of selected environments (both natural and artificial) can facilitate the indirect monitoring of AMR in broader populations of animals and humans. For example, the composition of the microbial communities present in water systems exposed to manure run-off from farms can reflect the types and prevalence of AMR present in the farm animals. The risk posed to human populations coming into contact with AMR pathogens present in those water systems can then also be evaluated. Sewage and wastewater microbial sampling could also be considered a form of 'artificial environment' surveillance, as opposed to investigating the natural environment. The aim of this type of surveillance is specifically to monitor AMR in the human populations producing the sewage, to understand the prevalence and sources of AMR outside of humans diagnosed clinically with infections. Whilst this is not strictly surveillance of the environment, from this point it will be referred to an environmental surveillance for simplicity, and to reflect the similar methods used. Other environments may be selected to allow the simultaneous monitoring of multiple potential sources of AMR, such as humans, farmed and wild animals. Regardless of the type of source, AMR control and/or implementation strategies could then be put in place based on the surveillance information.

To date environmental and population surveillance efforts for AMR have been very limited, with no large international coordinated efforts to help implement them as part of a One Health approach. However COVID-19 wastewater surveillance has accelerated progress and established some infrastructure; there is an opportunity to build on this and establish procedures for routine and coordinated environmental surveillance [237]. Sampling wastewater, or water known to be widely polluted from several sources, is an area of particular promise monitoring of AMR from multiple environments such as human, agricultural and industrial environments. The mixing of microorganisms from these different environments increases the risk of AMR gene transfer between different microorganisms, which could then pose a threat to human health [237]. Wastewater sampling provides and additional advantage in that the pathogen data is anonymous, making it possible to indirectly sample human populations without the strict consent regulations required for clinical sample use, though consent rules may still apply in some countries. Drinking water supplies are another key environmental component where the presence of AMR pathogens could pose a direct risk to human health.

Why Metagenomics?

Environmental surveillance approaches of any kind are currently very limited, providing an opportunity to test new surveillance models. Similar to the monitoring of healthy livestock, an approach that can provide broad simultaneous monitoring of the general AMR prevalence across all organisms, can be useful for estimating the overall AMR risk than pathogen specific culture and/or WGS methods. Environmental settings such as drinking water which could contain known high-risk pathogens, may benefit from a more accurate targeted approach, and the high resolution that WGS provides for tracking transmission pathways, especially in an outbreak situation. Targeted molecular panels may also have a useful role to play in monitoring multiple genes and/or pathogens at higher specificity and lower cost. However, by their nature these will always be limited in scope compared to metagenomics

as retrospective reanalysis of data to identify new genomic regions not covered by the pre-defined targets is not possible, and data formats are less easily shared.

Current status of metagenomics for environmental monitoring

Several studies have demonstrated the potential for metagenomics in sewage analysis. For example, a study of urban sewage in Kibera, an informal settlement in Nairobi, Kenya, was able to identify trends in approximate abundance of a variety of pathogenic bacteria, viruses, and parasites over time. For some bacterial pathogens, these changes correlated with changes in abundance of AMR genes. If implemented as a real-time surveillance system, this could be used to provide a background level of AMR data to help guide and interpret the success of AMR prevention measures [238]. Another study (described in case study 2) showed this approach could be extended to compare samples from multiple different countries, to allow surveillance to take place on a global level [239].

Whilst most studies have taken place in a research setting, the methods could begin to be applied in a routine surveillance format. The natural pooling of samples that takes place in sewage waste means that fewer samples are needed to represent a large area of the population, making this a relatively easy surveillance approach to implement compared to surveillance of farms or clinical cases, which require multiple independent samples from many sites. Although no fully evaluated sewage surveillance approach currently exists, there is an opportunity to design a standardised approach suitable for global implementation. Methodologies to facilitate other areas of environmental surveillance are also in development, one example combines culture-based with metagenomic analysis of AMR in wastewater, urban wetlands, beaches, agricultural soil, raw meat, and the transport system within Victoria, Australia [240].

Case study 2: International environmental surveillance based on sewage analysis [239]

Title: Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage

Type of study: Research: Protocol development and developing evidence of utility

Aim of study: To explore the utility of a metagenomic strategy for analysis of bacterial taxonomy and AMR in untreated sewage, to inform global surveillance of AMR and identify opportunities for AMR control interventions.

Project Summary: Metagenomic analysis of untreated sewage was used to characterise the bacterial resistome from 79 sites in 60 countries. Participants were instructed how to collect the urban sewage samples and associated metadata. Only one sample was taken from each site. The samples were sent to the Technical University of Denmark, who performed DNA extraction, sequencing, and data analysis.

Sequencing requirements: Library prep: NEXTflex PCR-free Library Preparation Kit (Bioo Scientific). Sequencing: libraries were multiplexed and sequenced on the HiSeq3000 platform (Illumina). Analysis: trimmed reads were input into the taxonomy-assignment tool MGmapper. Reads were aligned against reference sequence databases for the best hit. An acquired AMR gene database (ResFinder) was used to annotate properly paired reads. Multiple statistical tools used for distribution and determination analyses.

Utility for surveillance

- The authors suggest that this study 'provides the foundation for a flexible, simple, affordable, and ethically acceptable global real-time surveillance of AMR that could be immediately implemented globally also in low- and middle-income countries'. The study design could potentially be adapted for different techniques such as culture and PCR methods
- It was possible to identify systematic differences in the abundance and diversity of AMR genes between Europe/North-America/Oceania and Africa/Asia/South-America
- The contextual data on antimicrobial use, as well as the data on bacterial taxonomy only explained a minor part of the AMR variation observed. However, AMR gene abundance strongly correlated with socio-economic, health, and environmental factors, which the authors use to predict AMR gene abundances in all countries in the world.
- Overall, the authors suggest their findings show that global AMR gene diversity and abundance varies by region and sanitation and health could be important measures in reducing the global burden of AMR, not just limiting antimicrobial use.

Suggested infrastructure

No specific infrastructure detailed in this study, though it was assumed it was a strategy simple and affordable enough to be used in LMICs. It is uncertain if the protocols and analysis tools are readily accessible and easily applied to other studies

Next steps

The authors suggest that this type of methodology could be implemented for other surveillance projects, but could also benefit from further studies to identify the utility of different technologies, as well as to explore different analysis methods e.g., focusing on specific resistance genes. This publication was from the Global Sewage Surveillance Project Consortium, who are continuing to gather data for the project.

6.1.5 Metagenomic AMR surveillance in clinical settings

Current status of clinical surveillance. Despite the importance of a One Health approach being acknowledged, most AMR strategies still focus on human clinical sample surveillance, including those for sequencing such as the GLASS report on WGS [4]. AMR surveillance of human infections can be used to monitor the burden of clinically significant AMR in a particular population, to identify outbreaks and track AMR transmission, and to support outbreak control measures.

Phenotypic AST methods provide an answer on whether a pathogen is resistant or susceptible to a particular drug, and at what concentration a drug inhibits pathogen growth. However, one limitation is that the genetic mechanisms causing resistance are unknown. Sequencing can support surveillance of known infectious disease cases through identification of:

- Single or multiple resistance genes/mutations
- Emerging threats
- Degree of relatedness to other pathogens
- Transmission chains

WGS analysis of pathogens isolated and cultured from clinical samples is being advocated for AMR surveillance in clinical settings, as it provides the most complete and high-resolution data for addressing these additional needs. The extent to which WGS is used in surveillance will vary depending on the precise situation and surveillance need. For example, for some high-risk but relatively uncommon diseases most or all identified isolates are sequenced. This is the case for TB in England, where WGS is used for first line DST and to support surveillance and genomic epidemiology. For most other pathogens, an approach that targets a subset of isolates for sequencing may be required.

Why Metagenomics?

A key reason that metagenomics may prove a useful alternative to current WGS methods is its culture-free nature, which has several advantages over WGS of cultured clinical isolates. As clinical samples typically have a higher load of the pathogen of interest than environmental samples, assembling whole genomes of clinically relevant pathogens from metagenomic data is more feasible. The disadvantages of metagenomics include more limited sensitivity and a higher computational power requirement for genome assembly, and so for each scenario the suitability and benefits of metagenomics over culture based WGS methods for surveillance would have to be evaluated. For clinical cases, there may be a particular need for complete and high-resolution genomics data to enable phylogenetic analysis, for example to track and help prevent outbreaks. Therefore, metagenomics derived whole genome data would have to be suitable to fulfil these surveillance needs.

Whilst WGS provides the most complete data for clinical surveillance, there could also be utility for both diagnosis and surveillance in performing more focused analyses on specific genomic regions, which provide sufficient information to identify pathogen serotypes and the most clinically relevant AMR genes. For example, WGS of cultured isolates may not be available in all settings, and having some surveillance data is preferable to none at all. Targeted panels have been developed for this purpose, however analysis of metagenomic data could also be used to identify key targets and pathogens. Whilst it may be limited in sensitivity compared to targeted methods, metagenomics is not limited by having to select targets in advance, and the data could be reanalysed in future.

Current status of metagenomics for human clinical AMR surveillance. Metagenomics is not routinely used in any clinical context as a surveillance (or diagnostic) tool for AMR. There have been several research studies demonstrating its potential, though most focus on diagnostic use rather than for surveillance [209, 241, 242]. The exception to this is its use as a direct-from sample sequencing tool for TB surveillance. A recent study found that a metagenomics based approach was less accurate than a targeted WGS approach, but likely

to be more affordable and still sufficiently accurate for the diagnosis and surveillance of most TB clinical sputum samples [118] (case study 3). TB surveillance is more advanced than for many other AMR infections, and considerable effort has been spent on developing culture-free protocols, which are still not in routine use. While this is quite a specific example, the principle may be applied to help demonstrate the feasibility and utility of culture-independent sequencing for other bacterial AMR pathogens.

Aside from TB, a study of bacterial lower respiratory infections showed that a 6 hour 'sample to result' nanopore metagenomics workflow was 96.6% sensitive and 41.7% specific for pathogen detection compared with culture. Sensitivity and specificity levels could be increased to 100% with confirmatory PCR and pathogen specific gene analysis. The ability to identify AMR genes belonging to the different pathogens required improvement, however. The study also demonstrated the ability to construct genomes for surveillance purposes, which was deemed suitable for permitting outbreaks analysis for public health purposes [242]. Another study using nanopore technology was able to provide proof of principle that nearly complete *N. gonorrhoeae* genomes could be assembled directly from clinical urine samples, and using spiked in samples demonstrated that AMR gene detection was also feasible [243].

Further development and validation of the methods is still required for surveillance specific purposes, along with studies that compare the utility of this approach to others. There are also new methods in development that could lead to improvements in metagenomic sequencing, e.g. new sample preparation methods to enrich the total microbial DNA available, and the use of long read sequencing over short read, or combining both methods, to help improve genome assembly and provide more genomic context.

One of the major limitations to using metagenomics methods to obtain whole genome sequence data is the restricted amount of high-quality pathogen DNA that can be obtained from clinical samples, requiring new sample preparation and enrichment methods. One example of a company that has developed a technology to allow WGS direct from blood samples using a non-targeted metagenomic approach is Day Zero Diagnostics, described below. Tools such as this, which cover the entire sequencing workflow from sample to analysis, could help enable non-experts to perform sequencing for diagnostic purposes, with the data then feeding into AMR surveillance efforts. In addition, use of non-targeted NGS methods mean that in principle this method could be applicable to multiple pathogens. However, the applicability and validity of data produced from diagnostic sequencing tools for surveillance purposes is still not clear, and will depend on if the genome coverage and data quality is sufficient for surveillance needs. For surveillance to be effective globally, WGS data from diagnostics would also have to be shared for others to access.

Example: A company developing new tools to allow complete whole genome profiling direct from sample, using a metagenomics approach

Company: Day Zero Diagnostics[244]

- Technologies for sample prep to allow WGS direct from blood sample (Blood2Bac and BacDetect)
- Tools for computational analysis based on WGS data Keynome® ID and Keynome® g-AST
- MicrohmDB® - proprietary database to allow development and training of algorithms.

Together these make a sequencing-based rapid diagnostic that identifies, within hours, both the species and the antibiotic resistance profile of a bacterial pathogen.

Development stage: Validation

From a surveillance perspective, it is still to be determined if metagenomic sequencing data from clinical samples using these tools is equivalent to culture based WGS data in its utility, or if WGS would still be desirable in some or all circumstances. It is possible that both

methods may be used in a complementary manner to meet different data needs and circumstances, for example metagenomics helps with identification of novel pathogens or cases where the cause of infection is uncertain, while WGS provides more detailed information on known pathogens.

Case study 3: Metagenomic methods for direct from sample WGS [118]

Title: Whole-genome sequencing of *Mycobacterium tuberculosis* directly from clinical samples for high-resolution genomic epidemiology and drug resistance surveillance: an observational study

Year: 2020

Type of study: Research study

Aim of study: To optimise and implement a direct-whole genome sequencing workflow for WGS of *Mycobacterium tuberculosis*, to overcome technical challenges and assess its suitability for inferring transmission clusters and predicting drug resistance.

Project Summary: Direct whole-genome sequencing was performed on 37 clinical samples from 23 tuberculosis patients. Metagenomic sequencing was performed on samples with more than 10% *Mycobacterium tuberculosis* DNA following DNA extraction from sample (43% of specimens), for the remaining samples *Mycobacterium tuberculosis* DNA was further enriched using biotinylated probes. Following sequencing, 29 samples (85% of smear test positive and 55% of smear test negative) met the quality criteria for downstream analysis. Data from these samples was analysed alongside 780 previously sequenced clinical isolates, to assess regional transmission networks and predict drug susceptibility.

Sequencing requirements: Library prep: Illumina Nextera XT library preparation kit, with further enrichment using biotinylated baits for samples with low *Mycobacterium tuberculosis* DNA. Sequencing: 2×300 bp Illumina MiSeq runs, multiplexing 12 samples per run. Analysis: Reads were quality filtered and a read mapping approach used to identify genomic variants. Samples with over 90% of genomic coverage were used for downstream analysis of transmission clusters. PhyResSE and ReSeqTB catalogues of mutations that are known to confer drug resistance were used to predict drug resistance. IQ-TREE was used to build phylogenetic trees.

Utility for surveillance

This study demonstrated the ability of a workflow to produce data for surveillance from a direct-from sample sequencing strategy, that was useful for surveillance purposes of drug resistance and epidemiological analysis, even though sensitivity was lower than WGS obtained from pure cultures.

Whilst this approach is currently too expensive for many LMICs, costs are expected to be reduced in future. This approach could allow surveillance to be carried out in countries where culture and drug susceptibility testing are not routinely done.

This workflow showed that metagenomics and targeted enrichment culture independent sequencing methods could be used alongside each other to reduce overall costs, with metagenomic sequencing used in samples with low pathogen DNA, and the remaining samples sequenced using a target enrichment approach (discussed further in 6.3.2).

The authors believe that advances in genomic technologies mean that culture independent methods of WGS may be able to match the accuracy of culture base diagnostics in the near future.

Relevance to GLASS pathogens: This study demonstrates the potential of this method for AMR surveillance. However, further research would be required to assess the suitability of similar strategies for GLASS pathogens, and then design and test these strategies. This is a future facing tool for GLASS pathogens.

6.1.6 Requirements for successful metagenomics surveillance

For metagenomics to be a useful surveillance tool, there are requirements both in relation to the technology itself, and also external requirements and factors that govern its utility for surveillance. Other considerations to support use of metagenomics include the availability of appropriate infrastructure, equipment, and contextual data such as levels of antimicrobial use, to allow the necessary interpretation of metagenomics data for the intended surveillance purpose. Here the requirements specific to use of metagenomics in AMR surveillance are highlighted; more general considerations for AMR surveillance and use of sequencing data are described in Sections 2.3 and 5.0.

Further evidence demonstrating the utility of metagenomic sequencing approaches for surveillance, in comparison to other NGS methods. As has been demonstrated throughout this report, there are multiple approaches to NGS based sequencing, each with their own advantages and limitations. For each surveillance application and circumstance, it should be demonstrated which tool is most suitable to meet minimum and optimal requirements. For example, in cases where metagenomic sequencing is not affordable or surplus to basic requirements, it may be possible to use NGS panels instead. Currently metagenomics methods may lack the accuracy required for some surveillance purposes, but it is likely that the technology will continue to improve. There is also a need for funding for surveillance specific pilots to generate evidence of utility, to help move away from the current research focused studies, in those areas such as sewage surveillance where metagenomics-based tools have already been developed.

Standard protocols that can be implemented globally. This is relevant to all sequencing technologies, but especially important for metagenomics analysis of entire microbial populations which can be easily affected at each stage of the metagenomics workflow. Standardisation will help ensure results are comparable and informative globally

Global collaborating communities to set priorities, evaluate methodologies and define tools. In order to meet the requirements above, it will be essential to coordinate and agree on activities, to help ensure a global and integrated approach to surveillance. This includes collaboration between the different areas of surveillance to allow a One Health Approach.

Increasing accessibility to LMICS. Metagenomics is currently an expensive tool, and requires complex analysis and interpretations. There is a need to lower costs and make the technology accessible to those currently lacking resources, infrastructure and expertise. Whilst the cost of sequencing generally is expected to come down, costs may also be lowered by taking a centralised approach when available. For example, for sewage analysis, samples may be sent across the globe to a centralised infrastructure, where sample processing and sequencing may take place, before results are returned to individual countries for use in public health. This approach would avoid the need to invest heavily in sequencing infrastructure. Where this is not a useful strategy, efforts need to be made to allow more easily accessible sequencing tools, for example optimising metagenomics protocols for use on ONT technology.

Ensuring databases allow sufficient interpretation of metagenomic data. Most metagenomics analyses rely on a read mapping approach to identify microorganisms and AMR genes based on databases of known microorganisms. There is a need to ensure these databases require unbiased and representative data to make the most of the ability of metagenomics to detect multiple microorganisms and AMR genes simultaneously.

6.2 Targeted culture-independent sequencing approaches for WGS

Whilst one of the key benefits of metagenomics methods is that they are inherently culture-independent, this approach can also be extended to other types of sequencing tools. When a key requirement for sequencing is for the approach to be culture-free, it is useful to compare the properties of all potential options, either when choosing which tool to use or

developing a new tool. Some methods are already available for specific pathogen and/or AMR detection purposes, or can be readily expanded to new AMR genes and pathogens. Others are in an early research and development stage and need further validation, but may make promising tools in future.

Culture-independent methods can broadly be divided into two main categories

- Targeted sequencing of specific genomic regions i.e. targeted panel tests
- Culture independent WGS using targeted methods

Targeted NGS panel tests for the sequencing of specific regions of interest have already been discussed in Section 4.3, sequencing approaches. This section focussed on targeted approaches to enable WGS of specific pathogens directly from samples.

6.2.1 Targeted approaches for culture-independent WGS

The majority of WGS is currently performed on cultured isolates of the pathogen under investigation. The advantage of this approach is that it reduces host contamination and concentrates the DNA of the target pathogen being sequenced, which allows more high-quality sequence information to be produced. However, shifting to culture independent WGS, if technically possible, could overcome the limitations to culture-based sequencing, as described earlier, whilst still obtaining the detailed genomic information produced by WGS.

As an alternative to the non-targeted metagenomics methods, a whole genome sequence can be obtained independently of culture by selectively enriching for and amplifying DNA belonging to the target pathogen. This is performed in a similar manner to the targeted sequencing methods described above, but in this case the target regions are designed to cover the entire genome. One method is to design primers to produce overlapping amplicons of genomic regions, which together cover the entire genome of interest (tiled amplicons), producing data from which a whole genome sequence can be assembled. This has been the most common method used for the sequencing of SARS-CoV-2 and other viruses [245]. Another approach is to use ‘bait-capture’ target enrichment methods. For example, the genome of a pathogen can be ‘captured’ using biotinylated RNA probes; the RNA probes bind to the pathogen DNA, then are extracted from the sample using a compound that binds to the biotin, and the DNA amplified. The genome can then be assembled from the amplified DNA.

A semi-culture independent approach can also be used, where a sample is cultured on an agar plate to isolate pathogens of interest, but instead of then using the colonies to produce a pure culture, the colonies are sequenced directly using a non-targeted sequencing approach. This approach means that much more DNA is available for sequencing, including that of the pathogen of interest, whilst reducing the time needed to produce a pure culture.

6.2.2 Current and future use of targeted culture-independent WGS for AMR surveillance

For non-AMR purposes, there are multiple examples of where targeted culture-independent WGS has been used to characterise pathogens for surveillance purposes [245], typically in outbreak settings to enable rapid analysis, for example in Ebola, Zika and SARS-CoV-2 outbreaks. As some of these situations may occur locally and sporadically in settings without established links to centralised surveillance hubs, this has often been performed using Oxford Nanopore technologies, with the MinIon in particular suitable for transportation to field laboratory settings. However, for SARS-CoV-2 sequencing, culture independent amplicon based WGS has also been the method of choice for routine surveillance purposes. Notably, these examples of where culture-independent sequencing has been rapidly applied are all viruses, which have much smaller genomes than bacteria allowing a higher sequencing depth to be achieved to obtain sufficient quality whole genome data.

For bacterial sequencing, there have been several studies demonstrating that WGS information can be obtained from commercially available bait capture methods in bacteria. For example, the Agilent SureSelectXT Target Enrichment system has been used to enable WGS of bacteria including meningococcal strains, herpesviruses, *C. trachomatis* and *M. tuberculosis*, directly from clinical samples [246, 247]. For *M. tuberculosis*, these methods show promise in profiling AMR, and may provide a more representative analysis of *M. tuberculosis* genetic diversity than culture-based methods [224, 247, 248]. The utility of these methods for routine surveillance is currently unclear even in the well-researched field of *M. tuberculosis*, both in terms of the data quality produced compared to culture based WGS, and the feasibility in terms of the cost of the approaches. As methods continue to improve and are validated, it may become a useful alternative to culture-based sequencing. So far these methods have been highly target specific and require extensive optimisation of sample preparation and sequencing approaches. Therefore, careful method development is currently needed for each pathogen of interest, and it is possible that it will not be suitable for some pathogens and sample types [223].

A semi culture-independent WGS has been developed and evaluated for MRSA based on sequencing of bacterial colonies rather than pure cultured isolates, as a more rapid method of producing surveillance information for use in outbreak investigation, to allow infection control measures to be implemented faster. This is another approach that could be a useful WGS tool, though it is not truly a ‘direct from sample’ method [249].

7 Cross-cutting themes and conclusions

In this report, we have explored the current AMR surveillance and database landscape, sequencing technologies and their application to AMR surveillance, and sequencing approaches that could support future AMR surveillance efforts, particularly in the One Health setting. There is not a standardised approach to the collection of current AMR surveillance data or to sequencing approaches to support AMR surveillance globally.

While NGS technologies, including WGS, will have an important role to play in future AMR surveillance, there is not a 'one size fits all' approach that will meet the needs of all stakeholders. The sequencing approach and platform will vary depending on the pathogen, laboratory context, surveillance objective, available personnel resource, computing and bioinformatics infrastructure, and mechanisms for data sharing. While sequencing has become the primary tool of choice for the management of some pathogens, e.g. *Mycobacterium tuberculosis*, it will not replace phenotypic or molecular methods for measuring AMR in all pathogens. For many pathogens, more whole genome sequences are required in order to provide a foundation on which to build information gathering and surveillance efforts.

The surveillance settings in which sequencing will be used will also have an impact on the type of technology and sequencing approach deployed. For example, AMR surveillance as part of hospital infection control or for routine analysis of isolates sent to a reference laboratory might deploy short-read sequencing technologies to carry out WGS. On the other hand, environmental AMR surveillance could deploy a range of sequencing technologies to carry out targeted or untargeted sequencing.

Therefore, the use of sequencing for AMR surveillance will need to be considered not just in the context of these other methods, but in terms of the bigger picture of how WGS data can be used for a range of purposes, such as infection control, research, or more widespread monitoring.

There are a number of areas which could benefit from further consideration to optimise the use of sequencing as an AMR surveillance tool:

Scientific and biological understanding of AMR. Accurate characterisation of AMR genotype and association with phenotype is vital for supporting understanding of sequencing data and enabling the reliable detection of AMR. Collaborative efforts on sharing and interpreting sequencing data will be required to facilitate this.

Identification of best practice. A large number of organisations and sectors are invested in AMR surveillance and many face common challenges. A focus on sharing experience and best practice would support the development and validation of standardised workflows, databases and bioinformatics analyses, which will by their nature differ depending upon the type of sequencing data generated and the approach to analysis.

Standardisation and quality control. There is ongoing work to establish appropriate quality controls, reference material and standardisation of NGS for AMR surveillance. Work in this area has been carried out by PulseNet and GenomeTrakr, while Illumina has been working to establish internal developed standards for their sequencing platforms. The development of standards will be required for a range of applications – not just the clinical setting, but also other settings such as the food chain or the wider environment.

Databases. There are a large number of databases available for AMR surveillance data – both sequencing and non-sequencing data. They are fragmented in terms of their content and scope. With AMR genetic reference databases, for some it is unclear what the situation is regarding ongoing curation and maintenance. Focus on building sustainable and comprehensive reference databases will be critical toward the use of NGS for AMR surveillance. Consideration will also need to be given to whether separate AMR surveillance databases should be established, or whether efforts should focus on building the capacity of

existing pathogen genome sequence databases or integration of genetic data into surveillance databases, such is being done with TESSy.

Sequence data standardisation. The type of sequencing data contained within databases will have an impact on the analyses that can be carried out on that data. Agreement around data standards will be an important part of discussions around the best way forward to make pathogen sequencing data available for surveillance.

Data sharing. While schemes to support data sharing are in place, there is large variation within and between different countries as to the degree of data sharing, with many collaborations relying on *ad hoc* networks. Most data sharing networks are predominantly or between HICs. Consideration is needed on how best to enable data sharing internationally, including LMICs, as well as between all the contexts covered by One Health.

Fragmented surveillance using sequencing. There are many efforts underway in surveillance of AMR, yet there is no single resource available to interrogate or investigate all the activities taking place. Identifying which laboratories have the required skills, equipment and resources to be able to undertake AMR surveillance using sequencing could be valuable. A list of laboratories that have either participated in AMR surveillance projects and have underlying experience or capabilities, including which aspects of AMR they have investigated (pathogen, genes or sample sources) could help establish networks and sharing of knowledge.

AMR surveillance remains a critical central element of the global response to AMR. Without further progress and development including the use of sequencing technologies, it will not be possible to contain and improve this significant threat to population health.

8 Appendix

8.1 Abbreviations list

AGISAR - WHO Advisory Group on Integrated Surveillance of Antimicrobial Resistance
AMR - antimicrobial resistance
AMU - antimicrobial use
AST - antimicrobial susceptibility testing
BSL - Biosafety laboratory
CAESAR - Central Asian and Eastern European Surveillance of Antimicrobial Resistance
CARD - Comprehensive Antibiotic Resistance Database
CARPHA - Caribbean Public Health Agency
CARSS - Canadian Antibiotic Resistance Surveillance System
CE-IVD - Conformitè Europëenne (CE) - in vitro diagnostic
CDC - Centre for a Disease Control and Prevention
cDNA - complementary DNA
CLIA - clinical laboratory improvement amendments
CLSI - Clinical and Laboratory Standards Institute
DNA - deoxyribonucleic acid
DR - drug resistant
DST - drug susceptibility test
DTU - Danish Technical University
EAC - East Africa Community
EARS-Net - European antibacterial resistance surveillance network
ECDC - European Centre for Disease Prevention and Control
EEA - European Economic Area
EQA - External Quality Assurer
ESBL - extended spectrum beta-lactamases
EU - European Union
EUCAST - European Committee on Antimicrobial Susceptibility Testing
FAO - Food and Agriculture Organization of the United Nations
FDA - Food and Drug Administration
GAP-AMR - Global Action Plan to tackle AMR
Gen-FS - Genomics for Food Safety
GHRU - Global health research unit
GLASS - Global Antimicrobial Resistance and Use Surveillance System
GTPT - GenomeTrakr proficiency testing
HIC - high income countries

HIV - human immunodeficiency virus

ICMR - Indian Council of Medical Research

ICU - intensive care unit

IPC - infection prevention control

IVD - in vitro diagnostics

JPIAMR - Joint Programming Initiative on Antimicrobial Resistance

KfW - German development bank

LA-EQAS - Latin America EQA scheme

LMIC - low and middle income countries

MDR - multi-drug resistant

MIC - minimum inhibitory concentration

MLST - multi-locus sequence typing

MLVA - multi-locus variable copy numbers of tandem repeats analysis

MRSA - methicillin-resistant Staphylococcus aureus

NARMS - National Antimicrobial Resistance Monitoring System (USCDC)

NAP - national action plan

NCBI - National Center for Biotechnology Information

ng - nanogram

NGS - next generation sequencing

NICD-SA - National Institute for Communicable Disease South Africa

NIHR - National Institute for Health Research (UK)

OIE - World Organization for Animal Health

ONT - Oxford Nanopore Technologies

PacBio - Pacific Biosciences

PCR - polymerase chain reaction

PFGE - pulsed-field gel electrophoresis

PGM - Personal Genome Machine from Thermo Fisher Ion Torrent range

PHE - Public Health England (now known as UKHSA)

QC - quality control

QMS - quality management system

qPCR - quantitative polymerase chain reaction

ReLAVRA - Red Latinoamericana de Vigilancia de la Resistencia a los Antimicrobianos (Latin American AMR network)

RIVM - National Institute for Public Health and the Environment in the Netherlands

RNA - ribonucleic acid

rRNA - ribosomal ribonucleic acid

RT-qPCR - real-time quantitative polymerase chain reaction

RUO - research use only

SMRT - single molecule real-time
SNP - single nucleotide polymorphism
SOP - standard operating procedure
sp. - species
spp. - multiple species
TB - tuberculosis
Tb - terabytes
TESSy - European Surveillance System platform
TrACSS - Tripartite AMR country self-assessment survey
UKHSA - United Kingdom Health Security Agency
UK-NEQAS - United Kingdom National External Quality Assessment Service
US - United States
USD - United States Dollar
VNTR - multiple-locus variable-number tandem repeat
WGS - whole genome sequencing
WHO - World Health Organization
XDR - extensively drug resistant
ZMW - zero-mode wave guides

8.2 Surveillance systems

Abbreviations for Table 1:

ISIS-AR: Infectious Diseases Surveillance Information System for Antimicrobial Resistance;
LabBase2: Health Protection Agency's voluntary surveillance;
ANRESIS: Swiss Centre for Antibiotic Resistance surveillance database;
Euro-GASP: European Gonococcal Antimicrobial Surveillance Programme;
NARMS: National Antimicrobial Resistance Monitoring System;
CARSS: Canadian Antimicrobial Resistance Surveillance System;
CHINET: China Antimicrobial Surveillance Network;
MARSS: Marseille Antibiotic Resistance Surveillance System;
KONIS: Korean Nosocomial Infections Surveillance System;
GERMS-SA: Group for Enteric, Respiratory and Meningeal Diseases Surveillance in South Africa;
EARS-Net: European Antimicrobial Resistance Surveillance Network;
BSAC: Bacteraemia and Respiratory Resistance Surveillance System;
KISS: German National Nosocomial Infection Surveillance System;
SARI: Surveillance of Antibiotic Use and Bacterial Resistance in Intensive Care Units;
ARS: Antibiotic Resistance Surveillance System;
NARST: National Antimicrobial Resistance Surveillance, Thailand;
ARMed: Antibiotic Resistance Surveillance and Control in the Mediterranean Region;
GSSAR: Greek System for the Surveillance of Antimicrobial Resistance;
DANMAP: Danish Integrated Antimicrobial Resistance Monitoring and Research Programme;
ITAVARM: Italian Veterinary Antimicrobial Resistance Monitoring;
FINRES-Vet: Finnish Veterinary Antimicrobial Resistance Monitoring and Consumption of Antimicrobial Agents;
NORM: Norwegian Surveillance System for Antimicrobial Drug Resistance;
BMR-RAISIN: Bactéries Multirésistantes-Réseau d'alerte d'investigation et de surveillance des infections nosocomiales;
EPIMIC: EPIdemiological Surveillance and Alert Based on MICrobiological Data;
ONERBA: National Observatory of Bacterial Resistance Epidemiology;
BulSTAR: Bulgarian Surveillance Tracking Antimicrobial Resistance;
ISKRA: Intersectoral Coordination Mechanism for the Control of Antimicrobial Resistance; Croatia
FiRe: Finnish Study Group for Antimicrobial Resistance;
ARMIN: Monitoring Antibiotic Resistance in Niedersachsen;
AR-ISS: surveillance of antibiotic resistance in Italy;
SNARS: Slovak National Antimicrobial Resistance Surveillance System;
SVEBAR: Swedish Surveillance of Antimicrobial Resistance;
CA-MRSA: CA-MRSA surveillance system;
GLASS: Global Antimicrobial Resistance Surveillance System;
CNISP: Canadian Nosocomial Infection Surveillance Program;
JANIS: Japan Nosocomial Infections Surveillance;
ARSP: Antimicrobial Resistance Surveillance Programme of the Philippines;
JVARM: Japanese Veterinary Antimicrobial Resistance Monitoring System;
MIB: National Surveillance of Invasive Bacterial Diseases;
CARAlert: National Alert System for Critical Antimicrobial Resistances;
AURA: Antimicrobial Use and Resistance in Australia;
NTSS: National Tuberculosis Surveillance System;
CIPARS: Canadian Integrated Program for Antimicrobial Resistance Surveillance;
GISP: Gonococcal Isolate Surveillance Project;
KARMS: Korean Antimicrobial Resistance Monitoring System;
VICNISS: Victorian Hospital Acquired Infection Surveillance System;
WHONET-Argentina: National Argentine Network for Monitoring Antimicrobial Resistance;
ARMOR: Antibiotic Resistance Monitoring in Ocular Microorganisms;
MARAN: Monitoring of Antimicrobial Resistance and Antibiotic Usage in Animals in the Netherlands.

8.3 Table of sequencing platforms

Summary of a selection of sequencing instruments with technical specifications and costs. The costs are based on 2018 figures and given in US dollars. Adapted from [135] and [250]. Since the publication of this information in 2018, a number of sequencing platforms have entered the market, including: Ion Torrent Genexus; Oxford Nanopore Technologies Flongle; BGI DNA Nanoball technology. Abbreviations: bp – base pairs; kb – kilobases; Gb – gigabyte(s); na – not available.

Platform	Read length	Yield (Gb)	Run time	Instrument cost (US$)	Annual contract (US$)	Cost per Gb (US$)	Limitations	Advantages
Illumina MiniSeq	50-150bp	1.6-7.5	7-25 hours	50,000	5,000	200-400	High cost per Gb	Low instrument cost, established technology, low error rate, benchtop size
Illumina MiSeq	75-300bp	0.5-1.5	4-56 hours	99,000	14,000	250-2,000	High cost per Gb	Low instrument cost, established technology, low error rate, scalability, benchtop size
Illumina NextSeq	75-150bp	16-120	15-29 hours	250,000	32,000	33-43	High instrument cost	Low cost per Gb, established technology, low error rate, benchtop size
Illumina HiSeq2500	36-125bp	9-500	7 hours to 11 days	690,000	75,000	30-230	High instrument cost, need for deep multiplexing	Low cost per Gb, established technology, low error rate
Illumina HiSeq2000/4000	50-150bp	105-750	1-3.5 days	740,000-900,000	81,000	22-50	High instrument cost, need for deep multiplexing	Low cost per Gb, established technology, low error rate

Platform	Read length	Yield (Gb)	Run time	Instrument cost (US$)	Annual contract (US$)	Cost per Gb (US$)	Limitations	Advantages
Illumina HiSeq X	150bp	800-900 per flow cell	<3 days	1,000,000	93,000	7-10	High instrument cost, need for deep multiplexing, limited compatibility, requires large lab space	Low cost per Gb, established technology, low error rate
Illumina iSeq	150bp	0.3-1.2	9-17.5 hours	19,900	n/a	n/a	Read length	Low initial investment, short run time
Illumina Nova Seq (5000/6000)	150bp	2000-6000	16-44 hours	850,000-950,000	n/a	n/a	Long run time	Read accuracy
Ion PGM	200-400bp	0.03-2	3.7-23 hours	49,000	5,000-10,000	400-2,000	High cost per Gb, not able to do paired-end sequencing, poor homopolymer performance	Rapid sequencing run, benchtop
Ion Proton	Up to 200bp	up to 10	2-4 hours	224,000	20,000-30,000	80	Not able to do paired-end sequencing, poor homopolymer performance	Low cost per Gb, rapid sequencing run
Ion GeneStudio S5	200-400bp	0.6-8	2.5-6 hours	65,000	9,000-18,000	80-5,000	High cost per Gb, not able to do paired-end sequencing	Rapid sequencing run
Pacific Biosciences RS II	~20kb	~1	4 hours	695,000	84,000	1,000	13% single pass error rate, very high cost per Gb, high instrument cost	Very long read lengths, can sacrifice length for accuracy, rapid run time

Platform	Read length	Yield (Gb)	Run time	Instrument cost (US$)	Annual contract (US$)	Cost per Gb (US$)	Limitations	Advantages
Pacific Biosciences Sequel	~20kb	~5	4 hours	350,000	20,000	1,000	13% single pass error rate, very high cost per Gb, high instrument cost	Very long read lengths, can sacrifice length for accuracy, rapid run time
Oxford Nanopore MK MinION	Up to 200 kb	up to 10	up to 48 hours	1,000	0	100-400	10% single pass error rate, increased indel errors in repeat regions, high cost per Gb	Very low instrument cost, portability
Oxford Nanopore GridION	100,000+	50-100	30 minutes – 48 hours	2,400	n/a	n/a	High error rate	Short run time
Oxford Nanopore PromethION	100,000+	480-960	30 minutes – 48 hours	25,000	n/a	n/a	High error rate	Short run time

8.4 Sequencing platforms – comparison of characteristics

	Illumina	Thermo Fisher Ion Torrent	BGI and MGI tech DNA Nanoball Sequencing	Oxford Nanopore Technologies	Pacific Biosciences
Advantages	Comparatively low-cost sequencing at high throughput High raw read accuracy and read depth generating high accuracy data One of the more commonly used systems for high resolution genomic analysis allowing collaborative development of expertise and advances. Many genetic or research laboratories already possess these systems, and bioinformatics pipelines are relatively well-established High levels of sample multiplexing are possible, meaning a high number of samples can be run at once Availability of FDA regulated, CE-IVD marked NGS instruments for in vitro diagnostic use Option of targeted and WGS approaches	Relatively inexpensive at low throughput Comparatively short sequencing runs enable faster return of results Low substitution error rate Some systems facilitate a highly automated workflow for easy adoption and consistent application of sequencing Availability of FDA regulated, CE-IVD marked sequencer Option of targeted and WGS approaches Longer individual reads	Flexible sequencing including range of run times, reads lengths and output. High throughput Linear amplification reduces error accumulation during amplification	Rapid and flexible - useful for sequencing smaller genomes. Sequences read in real time – allowing for termination when user determines enough reads have been generated Smaller sequencing units can be purchased at low cost Relatively inexpensive at low throughput Small size and high portability of some systems supports use in the field Many reagents do not require cold storage Simple user interface and analysis platforms –the need for expertise for many applications of this technology is not required Simultaneous examination of methylation possible using direct RNA sequencing	Capable of high throughput, equivalent to that of Illumina sequencing platforms Capable of producing very high accuracy consensus reads – HiFi sequencing reads around 15,000 bases in length at over 99% accuracy Produces long reads Errors are random, not systematic, and can therefore be overcome with deeper sequencing Sequences read in real time – allowing for termination when user determines enough reads have been generated

	Illumina	Thermo Fisher Ion Torrent	BGI and MGI tech DNA Nanoball Sequencing	Oxford Nanopore Technologies	Pacific Biosciences
Disadvantages	Longer sequencing runs Most platforms are large and costly to purchase, some require specialised infrastructure for safe use Short reads limit accuracy in complex genomic regions and opportunity for identification of the genomic context of mobile genomic elements	Lower throughput in comparison to other NGS technologies, therefore comparatively expensive at high throughput. High sequencing error rate Higher cost per sample Shorter reads than are possible with other NGS technologies able to perform paired-end sequencing Fewer bioinformatics tools built for Ion Torrent-generated data	Highest throughput systems are very large Shorter reads than are possible with other NGS technologies	Limited barcoding means this approach is more expensive than other approaches for sequencing at high throughput (a high number of samples). Currently the mobile sequencing units are not capable of providing the same level of multiplexing as other next generation sequencing technologies. Raw signal output files are very large – this makes files difficult to store. As software and pipelines for analysis evolve rapidly, it is useful if not essential for these files to be available for subsequent analysis of the data. This could hinder data deposition on databases. Higher error rate in homopolymeric regions Lower read coverage than short-read platforms Higher base-call error rate and overall error rate	Relatively expensive to run compared to other NGS technologies Systems are large and more costly than some alternatives Reagents are more costly than alternatives High vulnerability to DNA fragmentation with a high error rate

8.5 AMR funders

Summary table of organisations and institutions supporting AMR research

Funder	Focus	Value
JPIAMR	Funding calls for i) disrupting drug resistance using innovative design, and ii) diagnostics and surveillance networks	i) Open call, total 19 million Euros available. 27 funding organisations from 18 countries; ii) Call opens April 2022. Funding organisations from 11 countries to date. Up to 50k Euros per call
UK government funding to support international collaborative efforts to tackle AMR. Involves three government departments and the Wellcome Trust.	Wide ranging and varied focus: • Global AMR innovation fund (GAMRIF) • Global health research collaborations • Research into drug resistant infections • Product development partnerships • Applied health research • Newton fund research collaborations	£464.4m invested between 2016 and 2021: • Department of Health and Social Care £94.6 m including GAMRIF • Wellcome Trust £175m (including CARB-X funding) • Department for International Development £161m • Department for Business, Energy and Industrial strategy £33.8m, including Newton fund.
AMR action fund Collaboration between the WHO, European Investment Bank and the Wellcome Trust	Aims to bring 2 to 4 new antibiotics to market in the next decade.	Plan to invest up to US$ 1 billion. More than 20 pharmaceutical companies have pledged funding.
Fleming Fund UK Government aid programme	Support LMICs to generate, share and use data to improve antimicrobial use and encourage investment in AMR.	£265 million, funding projects in 24 countries in Africa and Asia.

Funder	Focus	Value
CARB-X Range of international funders (e.g. US, UK, German governments, US NIH, Wellcome Trust)	Supports 54 projects around innovation in antibiotics, vaccines and diagnostics.	US$ 480 million invested 2016-22
CDC Foundation The AMR fund	Supports projects to improve antibiotic use, including access to lifesaving drugs, invests in innovative ways to prevent and treat these infections. Supports international collaborations.	Varied
EU Horizon Europe Strategic plan for 2021-24.	Tackling AMR is one of the goals of cluster 1, health; tackling diseases and reducing disease burden, as outlined in the strategic plan for 2021-24. Focus on the OneHealth context.	Varied, depending on funding applied for
Repair impact fund. Novo Holdings.	Investment in companies developing therapies to target resistant organisms.	US$ 165 million budget
InnovFin Infectious Diseases Finance Facility. Supported by the European Investment Bank.	Provides financial products to innovators developing or manufacturing vaccines, medicines, medical and diagnostic devices or novel research to combat infectious diseases.	Financing available varies, typically between €7.5 million and €75 million
US government research funding E.g. via CDC and NIH	Varied	Varied
Canadian Institutes of Health research. AMR funding.	Varied	Varied

Funder	Focus	Value
Australian Government Department of Health. Global health initiative,	Initial focus on tackling AMR and DR-TB.	AUS$28.4M over 10 years.
International centre for AMR solutions (ICARS). Funded by the Danish ministry of health.	Partner with low- and middle-income countries (LMICs) in their efforts to reduce drug-resistant infections.	Varied
Resources aggregating online funding information		
Global AMR R&D hub Funded by the German Federal Ministry of Education and Research and the German Federal Ministry of Health.	Online resource. Tracks global investment in AMR research. Searchable dashboard with this information.	n/a
The AMR solutions website. Supported by the Wellcome Trust	Offers a summary of available AMR funding	n/a
UN AMR multipartner trust fund Collates funding pledges from a range of international donors. For delivery of GAP AMR.	Supports work of the tripartite.	n/a

8.6 Acknowledgements

Prof Frank M. Aarestrup, Head of Division, National Food Institute, Research group for Genomic Epidemiology, Technical University of Denmark

Prof William Gaze, Professor of Microbiology at The European Centre for Environment and Human Health, University of Exeter

Dr Miriam Huntley, Co-founder and CTO of Day Zero Diagnostics

9 References

1. *Antimicrobial Resistance: World Health Organization*. 2021; Available from: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.
2. *Review on antimicrobial resistance*. 2016; Available from: https://amr-review.org/home.html.
3. *Global action plan on antimicrobial resistance. World Health Organization*. 2016; Available from: https://www.who.int/publications/i/item/9789241509763.
4. *GLASS whole-genome sequencing for surveillance of antimicrobial resistance. World Health Organization*. 2020; Available from: https://www.who.int/publications/i/item/9789240011007.
5. Cars, O., et al., *Resetting the agenda for antibiotic resistance through a health systems perspective.* Lancet Glob Health, 2021. **9**(7): p. e1022-e1027.
6. Kaprou, G.D., et al., *Rapid Methods for Antimicrobial Resistance Diagnostics.* Antibiotics 2021, Vol. 10, Page 209, 2021. **10**: p. 209.
7. *Molecular methods for antimicrobial resistance (AMR) diagnostics to enhance the Global Antimicrobial Resistance Surveillance System.* 2019, World Health Organization: Geneva.
8. *One Health Basics. Centers for Disease Control and Prevention*. 2018; Available from: https://www.cdc.gov/onehealth/basics/index.html.
9. *GLASS report: early implementation 2016-2017. World Health Organization*. 2018; Available from: https://www.who.int/publications/i/item/9789241513449.
10. *Surveillance. World Health Organization Regional Office for Europe*. 2022; Available from: https://www.euro.who.int/en/health-topics/disease-prevention/antimicrobial-resistance/surveillance.
11. Schwoebel, V., *Surveillance of tuberculosis.* Indian J Tuberc, 2020. **67**(4S): p. S33-S42.
12. *Surveillance standards for antimicrobial resistance. World Health Organization*. 2002; Available from: https://apps.who.int/iris/handle/10665/67426.
13. Filter, M., et al., *One Health Surveillance Codex: promoting the adoption of One Health solutions within and across European countries.* One Health, 2021. **12**: p. 100233.
14. *Global antimicrobial resistance surveillance system: manual for early implementation. World Health Organization*. 2015; Available from: https://apps.who.int/iris/handle/10665/188783.
15. Munkholm, L. and O. Rubin, *The global governance of antimicrobial resistance: a cross-country study of alignment between the global action plan and national action plans.* Globalization and Health, 2020. **16**(1): p. 109.
16. *Library of AMR national action plans. World Health Organization*. 2021; Available from: https://www.who.int/teams/surveillance-prevention-control-AMR/national-action-plan-monitoring-evaluation/library-of-national-action-plans.
17. *Monitoring global progress on antimicrobial resistance: Tripartite AMR country self-assessment survey (TrACSS) 2019-2020. World Health Organization*. 2021; Available from: https://www.who.int/publications/i/item/monitoring-global-progress-on-antimicrobial-resistance-tripartite-amr-country-self-assessment-survey-(tracss)-2019-2020.
18. Diallo, O.O., et al., *Antibiotic resistance surveillance systems: A review.* J Glob Antimicrob Resist, 2020. **23**: p. 430-438.
19. Ashley, E.A., et al., *An inventory of supranational antimicrobial resistance surveillance networks involving low- and middle-income countries since 2000.* Journal of Antimicrobial Chemotherapy, 2018. **73**(7): p. 1737-1749.
20. Tacconelli, E., et al., *Surveillance for control of antimicrobial resistance.* The Lancet Infectious Diseases, 2018. **18**(3): p. e99-e106.
21. World Bank, *Disease surveillance, emergency preparedness, and outbreak response in Eastern and Southern Africa: A Situational Assessment and Five-Year Action Plan for the Africa CDC Strengthening Regional Public Health Institutions and Capacity for Surveillance and Response Program.* 2021.
22. Antimicrobial Resistance Collaborators, *Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis.* Lancet, 2022.
23. *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report: 2021. World Health Organization*. 2021; Available from: https://www.who.int/publications/i/item/9789240027336.
24. Borg, M.A., et al., *Antibiotic resistance in the southeastern Mediterranean--preliminary results from the ARMed project.* Euro Surveill, 2006. **11**(7): p. 164-7.
25. *Summary report on the ninth intercountry meeting of national malaria programme managers from HANMAT and PIAM-Net countries Cairo, Egypt, 24–26 October 2017. World Health*

Organization. 2017; Available from: https://www.who.int/publications/i/item/WHO-EM-MAL-387-E.
26. *Global Influenza Surveillance and Response System (GISRS). World Health Organization*. 2021; Available from: https://www.who.int/initiatives/global-influenza-surveillance-and-response-system.
27. Gandra, S., et al., *Antimicrobial Resistance Surveillance in Low- and Middle-Income Countries: Progress and Challenges in Eight South Asian and Southeast Asian Countries.* Clinical Microbiology Reviews, 2020. **33**(3): p. e00048-19.
28. Merieux Foundation. *The AMR Benchmark: tracking pharma's response to drug resistance*. 2021; Available from: https://accesstomedicinefoundation.org/amr-benchmark.
29. Pfizer. *Antimicrobial Testing Leadership and Surveillance (ATLAS)*. 2021; Available from: https://atlas-surveillance.com/#/login.
30. Schnall, J., et al., *Surveillance and monitoring of antimicrobial resistance: limitations and lessons from the GRAM project.* BMC Med, 2019. **17**(1): p. 176.
31. Mader, R., et al., *Building the European Antimicrobial Resistance Surveillance network in veterinary medicine (EARS-Vet).* Eurosurveillance, 2021. **26**(4): p. 2001359.
32. Plaza-Rodriguez, C., et al., *Wildlife as Sentinels of Antimicrobial Resistance in Germany?* Front Vet Sci, 2020. **7**: p. 627821.
33. Founou, L.L., R.C. Founou, and S.Y. Essack, *Antibiotic Resistance in the Food Chain: A Developing Country-Perspective.* Frontiers in microbiology, 2016. **7**: p. 1881-1881.
34. *European Centre for Disease Prevention and Control. European Food- and Waterborne Diseases and Zoonoses Network (FWD-Net)*. 2020; Available from: https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/fwd-net.
35. *Diagnostic stewardship: a guide to implementation in antimicrobial resistance surveillance sites. World Health Organization*. 2016; Available from: https://apps.who.int/iris/handle/10665/251553.
36. Hombach, M., et al., *Influence of clinical breakpoint changes from CLSI 2009 to EUCAST 2011 antimicrobial susceptibility testing guidelines on multidrug resistance rates of Gram-negative rods.* Journal of clinical microbiology, 2013. **51**(7): p. 2385-2387.
37. Cusack, T.P., et al., *Impact of CLSI and EUCAST breakpoint discrepancies on reporting of antimicrobial susceptibility and AMR surveillance.* Clin Microbiol Infect, 2019. **25**(7): p. 910-911.
38. Chattaway, M.A., et al., *The Transformation of Reference Microbiology Methods and Surveillance for Salmonella With the Use of Whole Genome Sequencing in England and Wales.* Frontiers in Public Health, 2019. **7**(317).
39. Cars, O., et al., *Resetting the agenda for antibiotic resistance through a health systems perspective.* The Lancet Global Health, 2021. **9**(7): p. e1022-e1027.
40. Iskandar, K., et al., *Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture.* Antimicrobial Resistance & Infection Control, 2021. **10**(1): p. 63.
41. Lim, C., et al., *Surveillance strategies using routine microbiology for antimicrobial resistance in low- and middle-income countries.* Clin Microbiol Infect, 2021. **27**(10): p. 1391-1399.
42. Kariuki, S., et al., *Antimicrobial resistance surveillance in Africa: Successes, gaps and a roadmap for the future.* 2018, 2018. **7**(2).
43. Perovic, O. and C. Schultsz, *Stepwise approach for implementation of antimicrobial resistance surveillance in Africa.* African journal of laboratory medicine, 2016. **5**(3): p. 482-482.
44. NIHR Global Health Research Unit on Genomic Surveillance of AMR, *Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap.* BMJ Global Health, 2020. **5**(11): p. e002244.
45. Okeke, I.N., et al., *Leapfrogging laboratories: the promise and pitfalls of high-tech solutions for antimicrobial resistance surveillance in low-income settings.* BMJ Global Health, 2020. **5**(12): p. e003622.
46. Farrar, J. *A proposal to develop an equitable global pathogen surveillance network in 2021 that can prevent and respond to emerging and endemic infectious diseases at speed and at scale*. 2021; Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/990314/Pathogen_Surveillance_report_to_the_UK_G7_Presidency_2021.pdf.

47. Ronat, J.-B., et al., *AMR in low-resource settings: Médecins Sans Frontières bridges surveillance gaps by developing a turnkey solution, the Mini-Lab.* Clinical Microbiology and Infection, 2021. **27**(10): p. 1414-1421.
48. Kirchhelle, C., et al., *Setting the standard: multidisciplinary hallmarks for structural, equitable and tracked antibiotic policy.* BMJ Glob Health, 2020. **5**(9).
49. Veepanattu, P., et al., *Building resilient and responsive research collaborations to tackle antimicrobial resistance-Lessons learnt from India, South Africa, and UK.* Int J Infect Dis, 2020. **100**: p. 278-282.
50. Jayatilleke, K., *Challenges in Implementing Surveillance Tools of High-Income Countries (HICs) in Low Middle Income Countries (LMICs).* Curr Treat Options Infect Dis, 2020: p. 1-11.
51. Pan American Health Organization. World Health Organization. *Protocol for Enhanced Isolate-Level Antimicrobial Resistance Surveillance in the Americas. Primary Phase: Bloodstream Infections*. 2021; Available from: https://iris.paho.org/handle/10665.2/53275.
52. *WHO integrated global surveillance on ESBL-producing E. coli using a "One Health" approach. Global Tricycle Surveillance*. 2021; Available from: https://www.who.int/publications/i/item/who-integrated-global-surveillance-on-esbl-producing-e.-coli-using-a-one-health-approach.
53. Banu, R.A., et al., *Extended Spectrum Beta-Lactamase Escherichia coli in River Waters Collected from Two Cities in Ghana, 2018-2020.* Trop Med Infect Dis, 2021. **6**(2).
54. *Mekong Basin Disease Surveillance. Antimicrobial Resistance (AMR). Tricycle Program. "Cambodia, Lao P.D.R and Myanmar"*. 2022; Available from: https://mbdsnet.org/wp-content/uploads/2021/04/Antimicrobial-Resistance-Tricycle-Program.pdf.
55. *JPIAMR. Network for Enhancing Tricycle ESBL Surveillance Efficiency*. 2018; Available from: https://www.jpiamr.eu/app/uploads/2021/08/Summary_NETESE.pdf.
56. Efsa Panel on Biological Hazards, et al., *Role played by the environment in the emergence and spread of antimicrobial resistance (AMR) through the food chain.* EFSA J, 2021. **19**(6): p. e06651.
57. *Monitoring and evaluation of the global action plan on antimicrobiral resistance. Framework and recommended indicators. World Health Organization*. 2019; Available from: https://apps.who.int/iris/handle/10665/325006.
58. *Fleming Fund* 2021; Available from: https://www.flemingfund.org/.
59. Merieux Foundation. *The second edition of the AMR intensive course puts AMR at the centre of discussions*. 2021; Available from: https://www.fondation-merieux.org/en/news/the-second-edition-of-the-amr-intensive-course-puts-antimicrobial-resistance-at-the-center-of-discussions/.
60. *DTU Food National Food Institute. SEQAFRICA - Extending whole genome sequencing capacity for AMR Surveillance*. 2022; Available from: https://antimicrobialresistance.dk/seqafrica.aspx.
61. Kaur, J., et al., *ICMR's Antimicrobial Resistance Surveillance system (i-AMRSS): a promising tool for global antimicrobial resistance surveillance.* JAC-Antimicrobial Resistance, 2021. **3**(1).
62. Kaur, J., et al., *Design & development of customizable web API for interoperability of antimicrobial resistance data.* Sci Rep, 2021. **11**(1): p. 11226.
63. Kozyreva, V.K., et al., *Validation and Implementation of Clinical Laboratory Improvements Act-Compliant Whole-Genome Sequencing in the Public Health Microbiology Laboratory.* Journal of clinical microbiology, 2017. **55**(8): p. 2502-2520.
64. Struelens, M., et al. *Expert opinion on whole genome sequencing for public health surveillance*. 2016.
65. Gardy, J.L. and N.J. Loman, *Towards a genomics-informed, real-time, global pathogen surveillance system.* Nature Reviews Genetics, 2018. **19**(1): p. 9-20.
66. Grundmann, H. and H. Gelband, *Antimicrobial resistance surveillance with whole genome sequencing in Africa: It's (about) time.* Afr J Lab Med, 2018. **7**(2): p. 761.
67. Kekre, M., et al., *Integrating Scalable Genome Sequencing Into Microbiology Laboratories for Routine Antimicrobial Resistance Surveillance.* Clin Infect Dis, 2021. **73**(Supplement_4): p. S258-S266.
68. European Centre for Disease Prevention and Control. *ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations*. 2019; Available from: https://www.ecdc.europa.eu/en/publications-data/ecdc-strategic-framework-integration-molecular-and-genomic-typing-european.
69. European Centre for Disease Prevention and Control. *COVID-19 surveillance guidance - Transition from COVID-19 emergency surveillance to routine surveillance of respiratory*

pathogens. 2021; Available from: https://www.ecdc.europa.eu/en/publications-data/covid-19-surveillance-guidance.
70. Abrudan, M., et al., *Train-the-Trainer as an Effective Approach to Building Global Networks of Experts in Genomic Surveillance of Antimicrobial Resistance (AMR).* Clinical Infectious Diseases, 2021. **73**(Supplement_4): p. S283-S289.
71. Deloria Knoll, M., et al., *Global Landscape Review of Serotype-Specific Invasive Pneumococcal Disease Surveillance among Countries Using PCV10/13: The Pneumococcal Serotype Replacement and Distribution Estimation (PSERENADE) Project.* Microorganisms, 2021. **9**(4).
72. Rogers, L.A., et al., *The Role of Whole Genome Sequencing in the Surveillance of Antimicrobial Resistant Enterococcus spp.: A Scoping Review.* Frontiers in Public Health, 2021. **9**(656).
73. Ashton, P.M., et al., *Identification of Salmonella for public health surveillance using whole genome sequencing.* PeerJ, 2016. **4**: p. e1752.
74. Chattaway, M.A., et al., *Whole-genome sequencing for national surveillance of Shigella flexneri.* Frontiers in microbiology, 2017. **8**: p. 1700.
75. Chattaway, M.A., et al., *Identification of Escherichia coli and Shigella species from whole-genome sequences.* Journal of clinical microbiology, 2017. **55**(2): p. 616-623.
76. Chattaway, M.A., et al., *Whole genome sequencing for public health surveillance of Shiga toxin-producing Escherichia coli other than serogroup O157.* Frontiers in microbiology, 2016. **7**: p. 258.
77. Elson, R., et al., *Utility of whole genome sequencing to describe the persistence and evolution of Listeria monocytogenes strains within crabmeat processing environments linked to two outbreaks of listeriosis.* Journal of food protection, 2019. **82**(1): p. 30-38.
78. Greig, D.R., et al., *Evaluation of whole-genome sequencing for identification and typing of Vibrio cholerae.* Journal of clinical microbiology, 2018. **56**(11): p. e00831-18.
79. Inns, T., et al., *First use of whole-genome sequencing to investigate a cluster of Yersinia enterocolitica, Liverpool, United Kingdom, 2017.* Journal of medical microbiology, 2018. **67**(12): p. 1747-1752.
80. NHS Glasgow and Clyde. *Whole genome sequencing in the reference laboratory: An Introduction & Overview*. 2017; Available from: https://www.smvn.scot.nhs.uk/wp-content/uploads/2017/09/Whole_genome_sequencing_in_the_reference_laboratories.pdf
81. Timme, R.E., et al., *Gen-FS coordinated proficiency test data for genomic foodborne pathogen surveillance, 2017 and 2018 exercises.* Scientific Data, 2020. **7**(1): p. 402.
82. US Food and Drug Administration. *GenomeTrakr Fast Facts*. 2021; Available from: https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-fast-facts.
83. Brown, B., et al., *An economic evaluation of the Whole Genome Sequencing source tracking program in the U.S.* PLoS One, 2021. **16**(10): p. e0258262.
84. Nadon, C., et al., *PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance.* Euro Surveill, 2017. **22**(23).
85. Argimón, S., et al., *Integrating whole-genome sequencing within the National Antimicrobial Resistance Surveillance Program in the Philippines.* Nat Commun, 2020. **11**(1): p. 2719.
86. Afolayan, A.O., et al., *Overcoming Data Bottlenecks in Genomic Pathogen Surveillance.* Clin Infect Dis, 2021. **73**(Supplement_4): p. S267-S274.
87. *Union Health Minister Inaugurates Whole Genome Sequencing National Reference Laboratory for AMR and New BSL 3 Laboratory*, in *eHealth. The enterprise of healthcare.* 2021.
88. *WHO-PAHO. Biennial Meeting of The Latin American (ReLAVRA) and the Caribbean Network for Antimicrobial Resistance Surveillance*. 2019; Available from: https://www.paho.org/en/node/68930/biennial-meeting-latin-american-relavra-and-caribbean-network-antimicrobial-resistance.
89. Gianecini, R.A., et al., *Genomic Epidemiology of Azithromycin-Nonsusceptible Neisseria gonorrhoeae, Argentina, 2005-2019.* Emerg Infect Dis, 2021. **27**(9): p. 2369-2378.
90. Petroni, A. *Facilities in Microbiological and Genetic Analysis: Insights on Applied AMR Research. National Institute of Infectious Diseases (INEI) ANLIS "Dr. Carlos G. Malbrán", National and Regional Reference Laboratory for Antimicrobial Resistance, Ministry of Health, Argentina - PAHO / WHO* Available from: https://bbsrc.ukri.org/documents/amr-workshop-presentation-2-pdf/.

91. Faccone, D., et al., *Plasmid carrying mcr-9 from an extensively drug-resistant NDM-1-producing Klebsiella quasipneumoniae subsp. quasipneumoniae clinical isolate.* Infect Genet Evol, 2020. **81**: p. 104273.
92. Gagetti, P., et al., *Population genetic structure, serotype distribution and antibiotic resistance of Streptococcus pneumoniae causing invasive disease in children in Argentina.* Microb Genom, 2021. **7**(9).
93. European Centre for Disease Prevention and Control, *Eleventh external quality assessment scheme for Salmonella typing.* 2021, ECDC: Stockholm.
94. *East African Community set to launch second phase of Regional Network of Public Health Reference Laboratories for Communicable Diseases Project*. 2021; Available from: https://www.eac.int/press-releases/147-health/2091-eac-set-to-launch-second-phase-of-regional-network-of-public-health-reference-laboratories-for-communicable-diseases-project.
95. Affara, M., et al., *The East African Community (EAC) mobile laboratory networks in Kenya, Burundi, Tanzania, Rwanda, Uganda, and South Sudan—from project implementation to outbreak response against Dengue, Ebola, COVID-19, and epidemic-prone diseases.* BMC Medicine, 2021. **19**(1): p. 160.
96. Alghoribi, M.F., et al., *The role of whole genome sequencing in monitoring antimicrobial resistance: A biosafety and public health priority in the Arabian Peninsula.* J Infect Public Health, 2018. **11**(6): p. 784-787.
97. Heinz, E., et al., *Caribbean multi-centre study of Klebsiella pneumoniae: whole-genome sequencing, antimicrobial resistance and virulence factors.* Microbial genomics, 2019. **5**(5): p. e000266.
98. Strasheim, W., et al., *Ward-specific clustering of methicillin-resistant Staphylococcus aureus spa-type t037 and t045 in two hospitals in South Africa: 2013 to 2017.* PLoS One, 2021. **16**(6): p. e0253883.
99. Singh-Moodley, A., et al., *Complete Genome Sequence of a Staphylococcus aureus Isolate from a Nasopharyngeal Swab from a Mine Worker in South Africa.* Microbiol Resour Announc, 2019. **8**(50).
100. Paveenkittiporn, W., et al., *Molecular epidemiology of carbapenem-resistant Enterobacterales in Thailand, 2016–2018.* Antimicrobial Resistance & Infection Control, 2021. **10**(1): p. 88.
101. *Tuberculosis. World Health Organization*. 2021; Available from: https://www.who.int/news-room/fact-sheets/detail/tuberculosis.
102. *Global tuberculosis report 2021. World Health Organization*. 2021; Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2021.
103. *WHO announces updated definitions of extensively drug-resistant tuberculosis. World Health Organization*. 2021; Available from: https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis.
104. *FIND. Seq&Treat*. 2022; Available from: https://www.finddx.org/at-risk-populations/seq-treat/.
105. Meehan, C.J., et al., *Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues.* Nat Rev Microbiol, 2019. **17**(9): p. 533-545.
106. *Global tuberculosis report 2020. World Health Organization*. 2020; Available from: https://www.who.int/publications/i/item/9789240013131.
107. *Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance. World Health Organization*. 2021; Available from: https://www.who.int/publications/i/item/9789240028173.
108. *Tuberculosis reference laboratory. National Institute for Public Health and the Environment, The Netherlands*. 2018; Available from: https://www.rivm.nl/en/tuberculosis/reference-laboratory.
109. Jajou, R., et al., *Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing: A population-based study.* PLoS One, 2018. **13**(4): p. e0195413.
110. *Tuberculosis in England: 2021 report*. 2021; Available from: https://www.gov.uk/government/publications/tuberculosis-in-england-annual-report.
111. *Tuberculosis disease network. European Centre for Disease Prevention and Control*. 2021; Available from: https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/tuberculosis-disease-network.
112. *ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations*. 2019; Available from:

https://www.ecdc.europa.eu/en/publications-data/ecdc-strategic-framework-integration-molecular-and-genomic-typing-european.
113. Tagliani, E., et al., *Use of a whole genome sequencing-based approach for Mycobacterium tuberculosis surveillance in Europe in 2017-2019: an ECDC pilot study.* Eur Respir J, 2021. **57**(1).
114. Sanchini, A., et al., *Improving tuberculosis surveillance by detecting international transmission using publicly available whole genome sequencing data.* Euro Surveill, 2021. **26**(2).
115. Ismail, N.A., et al., *Prevalence of drug-resistant tuberculosis and imputed burden in South Africa: a national and sub-national cross-sectional survey.* Lancet Infect Dis, 2018. **18**(7): p. 779-787.
116. Dlamini, M.T., et al., *Whole genome sequencing for drug-resistant tuberculosis management in South Africa: What gaps would this address and what are the challenges to implementation?* J Clin Tuberc Other Mycobact Dis, 2019. **16**: p. 100115.
117. Votintseva, A.A., et al., *Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples.* J Clin Microbiol, 2017. **55**(5): p. 1285-1298.
118. Goig, G.A., et al., *Whole-genome sequencing of Mycobacterium tuberculosis directly from clinical samples for high-resolution genomic epidemiology and drug resistance surveillance: an observational study.* The Lancet Microbe, 2020. **1**(4): p. E175-E183.
119. Vogel, M., et al., *Implementation of whole genome sequencing for tuberculosis diagnostics in a low-middle income, high MDR-TB burden country.* Sci Rep, 2021. **11**(1): p. 15333.
120. Satta, G., et al., *Mycobacterium tuberculosis and whole-genome sequencing: how close are we to unleashing its full potential?* Clin Microbiol Infect, 2018. **24**(6): p. 604-609.
121. *Illumina. NextSeq 550Dx*. 2022; Available from: https://emea.illumina.com/systems/sequencing-platforms/nextseq-dx.html.
122. Iramiot, J.S., et al., *Whole genome sequences of multi-drug resistant Escherichia coli isolated in a Pastoralist Community of Western Uganda: Phylogenomic changes, virulence and resistant genes.* PLOS ONE, 2020. **15**(5): p. e0231852.
123. Sands, K., et al., *Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low- and middle-income countries.* Nature Microbiology, 2021. **6**(4): p. 512-523.
124. Katiyar, A., et al., *Genomic profiling of antimicrobial resistance genes in clinical isolates of Salmonella Typhi from patients infected with Typhoid fever in India.* Scientific Reports, 2020. **10**(1): p. 8299.
125. Chen, H., et al., *Drivers of methicillin-resistant Staphylococcus aureus (MRSA) lineage replacement in China.* Genome Medicine, 2021. **13**(1): p. 171.
126. Xu, X., et al., *Genomic characterization of Salmonella Uzaramo for human invasive infection.* Microbial Genomics, 2020. **6**(7).
127. Peirano, G., et al., *Genomic Epidemiology of Global Carbapenemase-Producing Enterobacter spp., 2008-2014.* Emerging infectious diseases, 2018. **24**(6): p. 1010-1019.
128. Zhu, Z., et al., *Emergence and genomics of OXA-232-producing Klebsiella pneumoniae in a hospital in Yancheng, China.* Journal of Global Antimicrobial Resistance, 2021. **26**: p. 194-198.
129. Liu, C., et al., *Hypervirulent Klebsiella pneumoniae is emerging as an increasingly prevalent K. pneumoniae pathotype responsible for nosocomial and healthcare-associated infections in Beijing, China.* Virulence, 2020. **11**(1): p. 1215-1224.
130. Bengtsson, R.J., et al., *Pathogenomic analyses of Shigella isolates inform factors limiting shigellosis prevention and control across LMICs.* Nature Microbiology, 2022. **7**(2): p. 251-261.
131. Quainoo, S., et al., *Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis.* Clinical microbiology reviews, 2017. **30**(4): p. 1015-1063.
132. *Methods Guide for Microbial Whole-Genome Sequencing*. 2021; Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/gated/10381_Micro_Methods_Guide_WGS.pdf.
133. Boolchandani, M., A.W. D'Souza, and G. Dantas, *Sequencing-based methods and resources to study antimicrobial resistance.* Nature Reviews Genetics, 2019. **20**(6): p. 356-370.
134. Kwong, J.C., et al., *Prospective Whole-Genome Sequencing Enhances National Surveillance of Listeria monocytogenes.* Journal of clinical microbiology, 2016. **54**(2): p. 333-342.
135. *The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide. World Health Organization*. 2018; Available from: https://apps.who.int/iris/handle/10665/274443.

136. *ThermoFisher Ion Torrent Genexus System*. 2022; Available from: https://www.thermofisher.com/uk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-torrent-genexus-system.html.
137. Ragupathi, N.K.D., et al., *Plasmid profiles among some ESKAPE pathogens in a tertiary care centre in south India.* The Indian journal of medical research, 2019. **149**(2): p. 222-231.
138. Sethuvel, D.P.M., et al., *Antimicrobial resistance, virulence & plasmid profiles among clinical isolates of Shigella serogroups.* The Indian journal of medical research, 2019. **149**(2): p. 247-256.
139. Devanga Ragupathi, N.K., et al., *First Indian report on genome-wide comparison of multidrug-resistant Escherichia coli from blood stream infections.* PLOS ONE, 2020. **15**(2): p. e0220428.
140. Chukamnerd, A., et al., *Whole-genome analysis of carbapenem-resistant Acinetobacter baumannii from clinical isolates in Southern Thailand.* Computational and Structural Biotechnology Journal, 2022. **20**: p. 545-558.
141. Liu, Q., et al., *Characterization of Salmonella Resistome and Plasmidome in Pork Production System in Jiangsu, China.* Frontiers in Veterinary Science, 2020. **7**.
142. Melnikov, S.V., et al., *Exploiting evolutionary trade-offs for posttreatment management of drug-resistant populations.* Proceedings of the National Academy of Sciences, 2020. **117**(30): p. 17924.
143. Zhao, Y.-h., et al., *Identification and expression analysis of ceftriaxone resistance-related genes in Neisseria gonorrhoeae integrating RNA-Seq data and qRT-PCR validation.* Journal of Global Antimicrobial Resistance, 2019. **16**: p. 202-209.
144. Zhang, B., et al., *Increased growth rate and amikacin resistance of Salmonella enteritidis after one-month spaceflight on China's Shenzhou-11 spacecraft.* MicrobiologyOpen, 2019. **8**(9): p. e00833-e00833.
145. Liu, B., et al., *Impact of Plasmid-Encoded H-NS–like Protein on blaNDM-1-Bearing IncX3 Plasmid in Escherichia coli.* The Journal of Infectious Diseases, 2020. **221**(Supplement_2): p. S229-S236.
146. Zhou, A., et al., *Whole-Genome Comparative and Pathogenicity Analysis of Salmonella enterica subsp. enterica Serovar Rissen.* G3 (Bethesda, Md.), 2020. **10**(7): p. 2159-2170.
147. Zong, X., et al., *Complete Genome Sequence of Community-Acquired Methicillin-Resistant Staphylococcus aureus Strain RJ1267, Isolated in Shanghai, China.* Microbiology Resource Announcements. **9**(18): p. e00244-20.
148. Paudyal, N., et al., *Persistent Asymptomatic Human Infections by Salmonella enterica Serovar Newport in China.* mSphere, 2020. **5**(3): p. e00163-20.
149. Fang, H., et al., *Genome Recombination-Mediated tRNA Up-Regulation Conducts General Antibiotic Resistance of Bacteria at Early Stage.* Frontiers in Microbiology, 2022. **12**.
150. Glanzmann, B., et al., *Human whole genome sequencing in South Africa.* Scientific reports, 2021. **11**(1): p. 606-606.
151. Gwinn, M., D. MacCannell, and G.L. Armstrong, *Next-Generation Sequencing of Infectious Pathogens.* JAMA, 2019. **321**(9): p. 893-894.
152. Chen, F.-J., et al., *Molecular Epidemiology of Emerging Carbapenem Resistance in Acinetobacter nosocomialis and Acinetobacter pittii in Taiwan, 2010 to 2014.* Antimicrobial Agents and Chemotherapy, 2019. **63**(4): p. e02007-18.
153. Yang, L., et al., *Emergence of ST63 Pandrug-Resistant Acinetobacter pittii Isolated From an AECOPD Patient in China.* Frontiers in Cellular and Infection Microbiology, 2021. **11**(1009).
154. Juma, M., et al., *Antimicrobial Resistance Profiling and Phylogenetic Analysis of Neisseria gonorrhoeae Clinical Isolates From Kenya in a Resource-Limited Setting.* Frontiers in Microbiology, 2021. **12**(1697).
155. Quick, J., et al., *Real-time, portable genome sequencing for Ebola surveillance.* Nature, 2016. **530**(7589): p. 228-232.
156. Xu, Y., et al., *Nanopore metagenomic sequencing to investigate nosocomial transmission of human metapneumovirus from a unique genetic group among haematology patients in the United Kingdom.* J Infect, 2020. **80**(5): p. 571-577.
157. Rhodes, J., et al., *Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen Candida auris.* Emerg Microbes Infect, 2018. **7**(1): p. 43.
158. Hirabayashi, A., et al., *On-Site Genomic Epidemiological Analysis of Antimicrobial-Resistant Bacteria in Cambodia With Portable Laboratory Equipment.* Frontiers in Microbiology, 2021. **12**(1199).

159. Ashton, P.M., et al., *MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island.* Nature Biotechnology, 2015. **33**(3): p. 296-300.
160. Kamathewatta, K.I., et al., *Exploration of antibiotic resistance risks in a veterinary teaching hospital with Oxford Nanopore long read sequencing.* PLOS ONE, 2019. **14**(5): p. e0217600.
161. Taxt, A.M., et al., *Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing.* Scientific Reports 2020 10:1, 2020. **10**: p. 1-11.
162. *Pacific Biosciences Certified Service Providers*. 2022; Available from: https://www.pacb.com/products-and-services/certified-service-providers/.
163. Reslan, L., et al., *The Emergence of Invasive Streptococcus pneumoniae Serotype 24F in Lebanon: Complete Genome Sequencing Reveals High Virulence and Antimicrobial Resistance Characteristics.* Frontiers in Microbiology, 2021. **12**(296).
164. Wang, J., et al., *PAU-1, a Novel Plasmid-Encoded Ambler Class A β-Lactamase Identified in a Clinical Pseudomonas aeruginosa Isolate.* Infection and drug resistance, 2019. **12**: p. 3827-3834.
165. van Dorp, L., et al., *Rapid phenotypic evolution in multidrug-resistant Klebsiella pneumoniae hospital outbreak strains.* Microbial Genomics, 2019. **5**(4).
166. Kanamori, H., et al., *Next-Generation Sequencing and Comparative Analysis of Sequential Outbreaks Caused by Multidrug-Resistant Acinetobacter baumannii at a Large Academic Burn Center.* Antimicrobial agents and chemotherapy, 2015. **60**(3): p. 1249-1257.
167. Chin, C.-S., et al., *Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.* Nature Methods, 2013. **10**(6): p. 563-569.
168. Neal-McKinney, J.M., et al., *Comparison of MiSeq, MinION, and hybrid genome sequencing for analysis of Campylobacter jejuni.* Scientific Reports, 2021. **11**(1): p. 5676.
169. Kosecka-Strojek, M., A. Buda, and J. Międzobrodzki, *Chapter 2 - Staphylococcal Ecology and Epidemiology*, in *Pet-To-Man Travelling Staphylococci*, V. Savini, Editor. 2018, Academic Press. p. 11-24.
170. Foxman, B., *Chapter 3 - Applications of Molecular Tools to Infectious Disease Epidemiology*, in *Molecular Tools and Infectious Disease Epidemiology*, B. Foxman, Editor. 2012, Academic Press: San Diego. p. 23-39.
171. Berbers, B., et al., *Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified Bacillus.* Scientific Reports, 2020. **10**(1): p. 4310.
172. Berbers B, et al., *Development of an NGS-Based Workflow for Improved Monitoring of Circulating Plasmids in Support of Risk Assessment of Antimicrobial Resistance Gene Dissemination.* Antibiotics 2020, Vol. 9, Page 503, 2020. **9**: p. 503.
173. Du, P., et al., *Novel IS26-mediated hybrid plasmid harbouring tet(X4) in Escherichia coli.* Journal of Global Antimicrobial Resistance, 2020. **21**: p. 162-168.
174. Shawa, M., et al., *Novel chromosomal insertions of ISEcp1-blaCTX-M-15 and diverse antimicrobial resistance genes in Zambian clinical isolates of Enterobacter cloacae and Escherichia coli.* Antimicrobial Resistance & Infection Control, 2021. **10**(1): p. 79.
175. Ruan, Z., et al., *Hybrid Genome Assembly and Annotation of a Pandrug-Resistant Klebsiella pneumoniae Strain Using Nanopore and Illumina Sequencing.* Infection and drug resistance, 2020. **13**: p. 199-206.
176. *Laboratory standard operating procedure for PulseNet Nextera XT library preparation*. 2019; Available from: https://www.cdc.gov/pulsenet/pdf/PNL34-PN-Nextera-XT-Library-Prep-508.pdf.
177. *Laboratory standard operating procedure for whole genome sequencing on MiSeq*. 2019; Available from: https://www.cdc.gov/pulsenet/pdf/PNL38-WGS-on-MiSeq-508.pdf.
178. *Laboratory standard operating procedure for PulseNet Nextera XT library prep and run set up for the Illumina MiSeq*. 2016; Available from: https://www.cdc.gov/pulsenet/pdf/pnl32-miseq-nextera-xt.pdf.
179. Sato, M.P., et al., *Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes.* DNA research : an international journal for rapid publication of reports on genes and genomes, 2019. **26**(5): p. 391-398.
180. Seth-Smith, H.M.B., et al., *Evaluation of Rapid Library Preparation Protocols for Whole Genome Sequencing Based Outbreak Investigation.* Frontiers in Public Health, 2019. **7**(241).
181. Li, S., et al., *GC Content-Associated Sequencing Bias Caused by Library Preparation Method May Infrequently Affect Salmonella Serotype Prediction Using SeqSero2.* Applied and Environmental Microbiology, 2020. **86**(18): p. e00614-20.

182. Uelze, L., et al., *Performance and Accuracy of Four Open-Source Tools for In Silico Serotyping of Salmonella spp. Based on Whole-Genome Short-Read Sequencing Data.* Applied and Environmental Microbiology, 2020. **86**(5): p. e02265-19.
183. Haendiges, J., K. Jinneman, and N. Gonzalez-Escalona, *Choice of library preparation affects sequence quality, genome assembly, and precise in silico prediction of virulence genes in shiga toxin-producing Escherichia coli.* PLOS ONE, 2021. **16**(3): p. e0242294.
184. Truong, J., et al., *Comparison of four enzymatic library preparation kits for sequencing Shiga toxin-producing Escherichia coli for surveillance and outbreak detection.* Journal of Microbiological Methods, 2021. **190**: p. 106329.
185. Shen, J., et al., *Toward Accurate and Robust Environmental Surveillance Using Metagenomics.* Frontiers in genetics, 2021. **12**: p. 600111-600111.
186. Ferreira, I., et al., *Culture-Free Detection of Antibiotic Resistance Markers from Native Patient Samples by Hybridization Capture Sequencing.* Microorganisms, 2021. **9**(8): p. 1672.
187. Feuerriegel, S., et al., *Rapid genomic first- and second-line drug resistance prediction from clinical Mycobacterium tuberculosis specimens using Deeplex-MycTB.* Eur Respir J, 2021. **57**(1).
188. *IDbyDNA and Illumina Launch New Rapid Test to Better Detect Respiratory Co-Infections and Identify Antibiotic Resistant Pathogens*. 2020; Available from: https://www.prweb.com/releases/idbydna_and_illumina_launch_new_rapid_test_to_better_detect_respiratory_co_infections_and_identify_antibiotic_resistant_pathogens/prweb17465315.htm.
189. *Comprehensive and rapid detection of respiratory pathogens and associated resistance genes using the MiniSeq™ System*. 2020; Available from: https://emea.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-respiratory-pathogen-panel-app-note-470-2020-011/illumina-respiratory-pathogen-panel-app-note-470-2020-011.pdf.
190. *RPIP Pathogens and AMR Marker List*. 2020; Available from: https://idbydna.com/rpip-pathogens-and-amr-marker-list/.
191. *All Clinical Research - IDbyDNA*. 2021; Available from: https://idbydna.com/clinical-research/.
192. Guernier-Cambert, V., et al., *Diverse Bacterial Resistance Genes Detected in Fecal Samples From Clinically Healthy Women and Infants in Australia-A Descriptive Pilot Study.* Frontiers in microbiology, 2021. **12**: p. 596984-596984.
193. Lau, K.A., et al., *Proficiency testing for bacterial whole genome sequencing in assuring the quality of microbiology diagnostics in clinical and public health laboratories.* Pathology, 2021. **53**(7): p. 902-911.
194. Ellington, M.J., et al., *The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee.* Clinical Microbiology and Infection, 2017. **23**(1): p. 2-22.
195. Lepuschitz, S., et al., *Analytical Performance Validation of Next-Generation Sequencing Based Clinical Microbiology Assays Using a K-mer Analysis Workflow.* Frontiers in Microbiology, 2020. **11**(1883).
196. de Sá, P.H.C.G., et al., *Chapter 11 - Next-Generation Sequencing and Data Analysis: Strategies, Tools, Pipelines and Protocols*, in *Omics Technologies and Bio-Engineering*, D. Barh and V. Azevedo, Editors. 2018, Academic Press. p. 191-207.
197. Khaledi, A., et al., *Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics.* EMBO Mol Med, 2020. **12**(3): p. e10264.
198. Anahtar, M.N., J.H. Yang, and S. Kanjilal, *Applications of Machine Learning to the Problem of Antimicrobial Resistance: an Emerging Model for Translational Research.* J Clin Microbiol, 2021. **59**(7): p. e0126020.
199. *Ares Genetics*. 2022; Available from: https://www.opgen.com/ares/.
200. *GenomeKey*. 2022; Available from: https://genomekey.com/.
201. Mihir Kekre, et al., *Integrating Scalable Genome Sequencing into Microbiology Laboratories for Routine AMR Surveillance.* Research Square, 2021.
202. Simar, S.R., B.M. Hanson, and C.A. Arias, *Techniques in bacterial strain typing: past, present, and future.* Current opinion in infectious diseases, 2021. **34**: p. 339-345.
203. Afolayan, A.O., et al., *Overcoming Data Bottlenecks in Genomic Pathogen Surveillance.* Clin Infect Dis, 2021. **73**(Supplement_4): p. S267-s274.
204. *Acuitas - AMR Gene Panel*. 2021 [cited 2021 16/12/2021]; Available from: https://www.opgen.com/acuitas/amr-gene-panel/.

205. Hutchins, R.J., et al., *Practical Guidance to Implementing Quality Management Systems in Public Health Laboratories Performing Next-Generation Sequencing: Personnel, Equipment, and Process Management (Phase 1).* Journal of clinical microbiology, 2019. **57**(8): p. e00261-19.
206. Bogaerts, B., et al., *Validation of a Bioinformatics Workflow for Routine Analysis of Whole-Genome Sequencing Data and Related Challenges for Pathogen Typing in a European National Reference Center: Neisseria meningitidis as a Proof-of-Concept.* Frontiers in microbiology, 2019. **10**: p. 362-362.
207. *GHRU Protocols - Centre for Genomic Pathogen Surveillance*. 2021; Available from: https://www.pathogensurveillance.net/resources/protocols/.
208. Rohden, F., et al. *Combined study on Digital Sequence Information (DSI) in public and private databases and traceability*. 2020; Available from: https://www.cbd.int/doc/c/1f8f/d793/57cb114ca40cb6468f479584/dsi-ahteg-2020-01-04-en.pdf.
209. Chiu, C.Y. and S.A. Miller, *Clinical metagenomics.* Nature Reviews Genetics, 2019. **20**(6): p. 341-355.
210. *AMR data platforms. JPIAMR*. 2021; Available from: https://www.jpiamr.eu/resources/amr-knowledge-hub/amr-data-platforms/.
211. Boolchandani, M., A.W. D'Souza, and G. Dantas, *Sequencing-based methods and resources to study antimicrobial resistance.* Nat Rev Genet, 2019. **20**(6): p. 356-370.
212. *hAMRonization database*. 2021; Available from: https://github.com/pha4ge/hAMRonization
213. Feldgarden, M., et al., *Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates.* Antimicrob Agents Chemother, 2019. **63**(11).
214. *The FAIR principles*. 2016; Available from: https://www.go-fair.org/fair-principles/.
215. *EARS network. European Centre for Prevention and Disease Control*. 2021; Available from: https://www.ecdc.europa.eu/en/about-us/networks/disease-networks-and-laboratory-networks/ears-net-about.
216. *Surveillance atlas of infectious diseases. European Centre for Disease Prevention and Control*. 2021; Available from: https://atlas.ecdc.europa.eu/public/index.aspx?Dataset=27&HealthTopic=4.
217. *AMR data on the Global Health Observatory. World Health Organization*. 2021; Available from: https://www.who.int/data/gho/data/themes/antimicrobial-resistance-(amr).
218. Fuhrmeister, A.S. and R.N. Jones, *The Importance of Antimicrobial Resistance Monitoring Worldwide and the Origins of SENTRY Antimicrobial Surveillance Program.* Open Forum Infectious Diseases, 2019. **6**(Supplement_1): p. S1-S4.
219. *SENTRY Antimicrobial Resistance Surveillance Program. JMI Laboratories*. 2021; Available from: https://www.jmilabs.com/sentry-surveillance-program/
220. *PulseNet methods. US Centers for Disease Control and Prevention*. 2021.
221. *Global Microbial Identifier*. 2021; Available from: https://www.globalmicrobialidentifier.org/.
222. *Pathogen detection. National Library of Medicine, NIH*. 2021; Available from: https://www.ncbi.nlm.nih.gov/pathogens.
223. Armstrong, G.L., et al., *Pathogen Genomics in Public Health.* New England Journal of Medicine, 2019. **381**(26): p. 2569-2580.
224. Nimmo, C., et al., *Whole genome sequencing Mycobacterium tuberculosis directly from sputum identifies more genetic diversity than sequencing from culture.* BMC Genomics, 2019. **20**(1): p. 389.
225. *Metagenomics*. 2021; Available from: https://www.genome.gov/genetics-glossary/Metagenomics.
226. Ricker, N., et al., *Incorporating the plasmidome into antibiotic resistance surveillance in animal agriculture.* Plasmid, 2021. **113**: p. 102529.
227. Knudsen, B.E., et al., *Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition.* mSystems, 2016. **1**(5).
228. Food and Agriculture Organisation of the United Nations, *Monitoring and surveillance of antimicrobial resistance in bacteria from healthy food animals intended for consumption.* 2021.
229. European Food Safety Authority and European Centre for Disease Prevention and Control, *The European Union Summary Report on Antimicrobial Resistance in zoonotic and indicator bacteria from humans, animals and food in 2018/2019.* EFSA J, 2021. **19**(4): p. e06490.

230. Holzel, C.S., J.L. Tetens, and K. Schwaiger, *Unraveling the Role of Vegetables in Spreading Antimicrobial-Resistant Bacteria: A Need for Quantitative Risk Assessment.* Foodborne Pathog Dis, 2018. **15**(11): p. 671-688.
231. *FARMS-SAFE project in Argentina*. 2021; Available from: https://www.bristol.ac.uk/amr/research/surveillance-of-amr-and-antimicrobial-usage-in-farming-/farms-safe-/.
232. Nadon, C., et al., *PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance.* Eurosurveillance, 2017. **22**(23): p. 30544.
233. Munk, P., et al., *Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries.* Nature Microbiology, 2018. **3**(8): p. 898-908.
234. Munk, P., et al., *A sampling and metagenomic sequencing-based methodology for monitoring antimicrobial resistance in swine herds.* Journal of Antimicrobial Chemotherapy, 2016. **72**(2): p. 385-392.
235. Ogunremi, D., et al., *A New Whole Genome Culture-Independent Diagnostic Test (WG-CIDT) for Rapid Detection of Salmonella in Lettuce.* Frontiers in Microbiology, 2020. **11**(602).
236. *Effort Against AMR. European Union*. 2021; Available from: https://www.effort-against-amr.eu.
237. Pruden, A., et al., *Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments.* Current Opinion in Microbiology, 2021. **64**: p. 91-99.
238. Hendriksen, R.S., et al., *Pathogen surveillance in the informal settlement, Kibera, Kenya, using a metagenomics approach.* PLOS ONE, 2019. **14**(10): p. e0222531.
239. Hendriksen, R.S., et al., *Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage.* Nature Communications, 2019. **10**(1): p. 1124.
240. *Culture-based and culture-independent analysis of samples (wastewater, urban wetlands, beaches, agricultural soil, raw meat, subway escalators) from within Victoria*. 2021; Available from: https://sedric.org.uk/amr-surveillance-projects/?member_projects%5Btopic%5D%5B%5D=genomics&member_projects%5Bregion%5D%5B%5D=&member_projects%5Bpast%5D=1.
241. Gong, L., et al., *Culture-independent analysis of liver abscess using nanopore sequencing.* PLoS One, 2018. **13**(1): p. e0190853.
242. Charalampous, T., et al., *Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection.* Nature Biotechnology, 2019. **37**(7): p. 783-792.
243. Street, T.L., et al., *Optimizing DNA Extraction Methods for Nanopore Sequencing of Neisseria gonorrhoeae Directly from Urine Samples.* Journal of clinical microbiology, 2020. **58**(3): p. e01822-19.
244. *Day Zero Diagnostics* 2021; Available from: https://www.dayzerodiagnostics.com/.
245. Quick, J., et al., *Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples.* Nature Protocols, 2017. **12**(6): p. 1261-1276.
246. Clark, S.A., et al., *Targeted DNA enrichment and whole genome sequencing of Neisseria meningitidis directly from clinical specimens.* International Journal of Medical Microbiology, 2018. **308**(2): p. 256-262.
247. Brown, A.C., *Whole-Genome Sequencing of Mycobacterium tuberculosis Directly from Sputum Samples.* Methods Mol Biol, 2021. **2314**: p. 459-480.
248. Doyle, R.M., et al., *Direct Whole-Genome Sequencing of Sputum Accurately Identifies Drug-Resistant Mycobacterium tuberculosis Faster than MGIT Culture Sequencing.* Journal of clinical microbiology, 2018. **56**(8): p. e00666-18.
249. Blane, B., et al., *Rapid sequencing of MRSA direct from clinical plates in a routine microbiology laboratory.* Journal of Antimicrobial Chemotherapy, 2019. **74**(8): p. 2153-2156.
250. *Whole genome sequencing for foodborne disease surveillance: landscape paper. World Health Organization*. 2018; Available from: https://apps.who.int/iris/handle/10665/272430.