GLASS
Whole-genome sequencing for surveillance of antimicrobial resistance

Global Antimicrobial Resistance and Use Surveillance System (GLASS)

World Health Organization

GLASS
Whole-genome sequencing for surveillance of antimicrobial resistance

Global Antimicrobial Resistance and Use Surveillance System (GLASS)

World Health Organization

GLASS whole-genome sequencing for surveillance of antimicrobial resistance

ISBN 978-92-4-001100-7 (electronic version)

ISBN 978-92-4-001101-4 (print version)

Design and layout by 400 Communications Limited.

Contents

Preface

Antimicrobial resistance (AMR) is an increasing threat to public health and sustainable development. The Global Action Plan on AMR underscores surveillance to strengthen the knowledge and evidence base for informing strategies and monitoring the effectiveness of interventions. GLASS currently monitors human infections due to several priority pathogens with microbiological data derived from phenotypic methods for AMR testing.

Whole-genome sequencing (WGS) provides a vast amount of information and the highest possible resolution for pathogen subtyping. The application of WGS for global surveillance can provide information on the early emergence and spread of AMR and further inform timely policy development on AMR control. Sequencing data emanating from AMR surveillance may provide key information to guide the development of rapid diagnostic tools for better and more rapid characterization of AMR, and thus complement phenotypic methods.

This document addresses the applications of WGS for AMR surveillance, including the benefits and limitations of current WGS technologies. Local, subnational, national and international case studies are included as examples of use of WGS in AMR surveillance. Information is also provided on the requirements for setting up and upgrading laboratories to ensure capacity for WGS and for introducing WGS into AMR surveillance systems.

As for any new technology, the application of WGS has some limitations and raises practical challenges in various settings globally. But innovation and further development of WGS methods will ensure that the scope of this new technology will expand application in the future. The costs associated with WGS are decreasing rapidly, which may enable broader, affordable access to this new technology in all countries. This technical note is intended to assist countries that are considering use of methods for AMR detection and surveillance to increase their capacity.

WHO is grateful for the support of international, regional and national partners that contributed to the development of this technical note on the applications of WGS for AMR surveillance.

Acknowledgements

WHO thanks the following authors and contributors

David Aanensen (Centre for Genomic Pathogen surveillance, University of Oxford, UK), Majed Alghoribi (WHO Collaborating Centre for Infection Prevention Control and Antimicrobial Resistance, Kingdom of Saudi Arabia), Allison C. Brown, Jason P. Folster, Benjamin Park, Dawn Seivert (WHO Collaborating Centre for International Monitoring of Bacterial Resistance to Antimicrobial Agents, Centers for Disease Control and Prevention, United States of America), Alejandra Corso (WHO Collaborating Centre, ANLIS, Argentina), Rene Hendriksen (WHO Collaborating Centre for Antimicrobial Resistance in Foodborne Pathogens and Genomics, Denmark), Ben Howden (Peter Doherty Institute for Infection and Immunity, Australia), Chikwe Ihekweazu, (Nigeria Centre for Disease Control, Nigeria), Roman Kozlov (WHO Collaborating Centre for Capacity Building on Antimicrobial Resistance Surveillance and Research, Russian Federation), Olga Perovic (WHO Collaborating Centre for Antimicrobial Resistance, South Africa), Timothy C. Rodwell (University of California, San Diego and FIND), Magnus Unemo (WHO Collaborating Centre for Gonorrhea and Other Sexually Transmitted Infections, Sweden), Carolin Vegvari (Centre for Genomic Pathogen surveillance, University of Oxford, UK), Neil Woodford (WHO Collaborating Centre for Reference and Research on Antimicrobial Resistance and Healthcare Associated Infections, UK) and the WHO AMR Surveillance and Quality Assessment Collaborating Centres Network.

Contributions by WHO staff

Sheick Oumar Coulibaly at the Regional Office for Africa (AFRO); Nienke Bruinsma, Marcelo Galas and Pilar Ramon Pardo at the Regional Office for the Americas (PAHO); Dana Itani at the Regional Office for the Eastern Mediterranean (EMRO); Danilo Lo Fo Wong and Marcello Gelormini at the Regional Office for Europe (EURO); Aparna Singh Shah at the Regional Office for South East Asia (SEARO); Takeshi Nishijima at the Regional Office for the Western Pacific (WPRO).

Staff from WHO Headquarters

Jorge Raul Matheu Alvarez, Silvia Bertagnolio, Anne Dean, Sergey Romualdovich Eremin, Sapna Manglani, Vaseeharan Sathiyamoorthy, Lorenzo Subissi, Christopher Oxenford, Carmem Lucia Pessoa-Silva, Matteo Zignol, Catharina Van Weezenbeek.

External reviewers

Sayed Fekry Abdelwahab (University Taif, Kingdom of Saudi Arabia), Valeria Bortolaia (Technical University of Denmark), Josefina Campos (INEI-ANLIS, Argentina), Tom Connor (Cardiff University, UK), Antoine Abou Fayad (American University of Beirut, Lebanon), Lamia Fouad (Ain Shams University, Egypt), Christian Giske (Karolinska Institute, Sweden), Nathalie Guessennd (Institut Pasteur, Cote d'Ivoire), Katie Hopkins (Public Health England, UK), Gunnar Kahlmeter (EUCAST, Sweden), Sébastien Matamoros (Academisch Medisch Centrum, Netherlands), Roberto G. Melano (Public Health Ontario, University of Toronto), Patrice Nordmann (University of Fribourg, Switzerland), Robin Patel (American Society for Microbiology, USA), Alejandro Petroni (Malbrán, Argentina), Iruka Okeke (University of Ibadan, Nigeria), Jorge Fernández Ordenes (Instituto de Salud Pública de Chile, Santiago, Chile), Firdausi Qadri (ASCODD, Bangladesh), Adriana Rosato (Institute for Academic Medicine, USA), Mayar Said (NAMRU-3, Egypt), Joergen Schlundt (Nanzang University, Singapore), Sonia Sia (Research Institute for Tropical Medicine, Philippines), Vitali Sintchenko (University of Sydney, Australia), Padmini Srikantiah (Bill and Melinda Gates Foundation, USA), Marc Struelens (European Centre for Disease Prevention and Control, Sweden), Javier Tognarelli (Instituto de Salud Pública de Chile, Santiago, Chile), Tim Walker (Oxford University, UK), Tim Walsh (Cardiff University, UK).

Developer group

David Aanensen, Jorge Raul Matheu Alvarez, Rene Hendriksen, Ben Howden, Chikwe Ihekweazu, Carmem Lucia Pessoa-Silva, Timothy C. Rodwell, Carolin Vegvari and Neil Woodford.

Executive group

David Aanensen, Jorge Raul Matheu Alvarez, Carmem Lucia Pessoa-Silva.

Editing

Elisabeth Heseltine.

Financial support

Bill and Melinda Gates Foundation, USA.

Acronyms

AMR	antimicrobial resistance
AST	antimicrobial susceptibility testing (phenotypic)
ESBL	extended-spectrum β-lactamase
Gb	gigabyte
GLASS	Global Antimicrobial Resistance and Use Surveillance System
HIC	high-income countries
LMIC	low- and middle-income countries
MDR	multidrug-resistant
MIC	minimum inhibitory concentration
MRSA	methicillin-resistant *Staphylococcus aureus*
NRL	national reference laboratory
PCR	polymerase chain reaction
QA	quality assurance
QC	quality control
RRL	regional reference laboratory
SDG	Sustainable Development Goal
TB	tuberculosis
WGS	whole-genome sequencing

Executive summary

The rising prevalence of antimicrobial resistance (AMR) is a threat to public health globally, as emerging AMR mechanisms and multidrug-resistant (MDR) pathogens compromise treatment of microbial infections. WHO Member States recognized this threat by unanimously approving a global action plan to tackle AMR at the Sixty-eighth World Health Assembly (resolution WHA68.7). AMR has also been recognized as a threat to attainment of the Sustainable Development Goals (SDGs) *(1)*. It is estimated that, by 2030, AMR infections resulting in increased morbidity, disability, premature deaths and reduced effective labour will become a significant threat to the global economy if action is not taken *(2)*. Moreover, AMR infections in livestock endanger sustainable food production and food security *(3)*.

Epidemiological data on antimicrobial-resistant pathogens are essential to inform policy and to monitor the effectiveness of interventions. Whole-genome sequencing (WGS) offers a vast amount of information and the highest resolution for identifying and characterizing pathogens. With epidemiological and clinical information, WGS can therefore enhance surveillance capacity to better inform strategies to tackle AMR. WGS has been used successfully in AMR surveillance for pathogens such as in multidrug-resistant tuberculosis (TB) and drug-resistant HIV. This document, written for policy-makers and laboratory directors in countries that are considering including WGS in their AMR surveillance programmes, addresses application of WGS for surveillance of AMR in fast-growing bacteria, discusses the benefits and limitations of current WGS technologies for AMR surveillance and outlines the requirements for strengthening or building new WGS laboratories. New developments in the use of WGS for AMR surveillance are increasingly accessible to a wider range of users.

WGS is used in molecular biology to obtain the complete or nearly complete DNA sequence of an organism. For pathogen surveillance and public health, the DNA sequence of an organism can be compared with a database of AMR genes and mutations in well-studied microbial genomes to draw inferences about important phenotypic characteristics of the organism, such as AMR and virulence factors. Moreover, if the sequence data are of sufficient quality, comparison of entire microbial genomes allows reconstruction of putative transmission networks for both antimicrobial-resistant clones and mobile genetic vectors of AMR and the evolutionary history of newly characterized AMR organisms and disease outbreaks.

WGS is not a substitute for phenotypic methods for detecting AMR for public health or for guiding the clinical treatment of most bacterial infections. WGS data can be used to verify the identity of AMR mechanisms in isolates with relevant phenotypic AMR or with discordant phenotypic AMR. It cannot, however, be used to quantify the level of phenotypic AMR, so that it is unsuitable for routine or predictive AST and therefore cannot replace phenotypic methods. It can, however, complement phenotypic methods by adding information on molecular determinants and mechanisms of AMR and genetic factors that facilitate their transmission in microbial populations. Currently, knowledge is lacking on AMR mechanisms, how they and antimicrobial-resistant strains spread and the precise, concrete measures that could be taken to contain AMR. Linkage of comprehensive genome databases to epidemiological and clinical metadata would be invaluable for public health, medical research and clinical care. WGS can add important, policy-relevant information for global AMR surveillance, including more accurate definition of the geographical distribution of resistance genes. AMR surveillance also signals the emergence and transmission of AMR organisms among animals and between animals and humans ("One Health") *(4)*.

A literature review conducted for this analysis indicated that most of the sequenced pathogenic isolates to date have been provided by researchers in high-income countries (HIC). Sequencing technologies and their cost are, however, changing rapidly, providing an opportunity to narrow the gaps in knowledge and technology in low- and middle-income countries (LMIC) and strengthen basic capability to identify pathogens accurately and use both phenotypic and molecular methods such as WGS for AMR surveillance to enhance public health.

International standards should be set for using WGS to predict AMR in pathogens to ensure that the results from different laboratories are comparable. Laboratories should be able to use "gold standard" quality-assured strains and their own protocols, bioinformatics algorithms and software to identify the species of a bacterial strain and to determine the presence or absence of acquired genes and genomic mutations associated with the phenotypic expression of decreased susceptibility or resistance at a specified accuracy. For this to be possible, public databases of sequences should be better curated. Thus, laboratories should upload WGS data into a sequence data repository only with a report confirming that they fulfil important quality assurance (QA) standards. Suitable standards will have to be agreed upon to ensure inclusion of newly established WGS laboratories in LMIC. Not only should laboratories fulfil QA standards, but individual sequences should be uploaded to a database only if they meet quality standards.

WGS is not the method of choice in all circumstances. There should be consensus on a strategy to achieve the defined objectives of AMR surveillance, including its design, the necessary epidemiological and clinical data and microbiological methods. For example, WGS is a useful addition to phenotypic surveillance if reconstruction of transmission chains is necessary to determine the source of infection, the events that led to the acquisition of AMR and breakdowns in infection control practices, e.g. in hospital outbreaks. WGS also provides information on which AMR genes are present and whether they are located on the bacterial chromosome or on plasmids. Furthermore, WGS provides the highest resolution for quantifying the relatedness of human and animal isolates, which is essential for a One Health approach to AMR surveillance *(5)*.

Despite its benefits, the use of WGS in surveillance of AMR in fast-growing bacteria has a number of limitations for public health, including substantial initial and recurrent investment and incomplete understanding of the molecular mechanisms underlying resistance to some antimicrobial classes (Annex 1). As only known resistance mechanisms can be detected, WGS cannot currently replace phenotypic surveillance of AMR in fast-growing bacteria. Additional challenges that must be met before WGS can be included in surveillance of AMR in fast-growing bacteria include local capacity-building, laboratory infrastructure and techniques, standardization of bioinformatics methods (achievable by outsourcing), storage technology in settings with little or no prior experience in the use of molecular methods and agreement on QA protocols and on protocols for data-sharing and use.

To extend use of WGS for AMR surveillance, the initial focus could be on one or a few selected pathogens of public health importance and on antimicrobial agents for which the mechanisms of resistance are well understood. The Global Antimicrobial Resistance and Use Surveillance System (GLASS) monitors a number of priority pathogens with phenotypic methods (Annex 2) *(6)*. For global AMR surveillance, WGS could be introduced initially for a subset of GLASS priority pathogens and/or for organisms reported under the Emerging Antimicrobial Resistance Reporting system *(7)*.

01 Introduction

The rising prevalence of antimicrobial resistance (AMR) is a threat to public health globally, as recognized by WHO Member States in World Health Assembly resolution WHA68.7. In particular, novel emerging AMR mechanisms and multidrug-resistant (MDR) pathogens threaten treatment of microbial infections.

It is estimated that infections due to AMR will have an indirect economic cost of US$ 1–3.4 trillion in terms of morbidity, disability, premature deaths and reduced effective labour by 2030 if action is not taken to counter the rise of AMR *(3)*. In addition, AMR has been recognized as a threat to attainment of the Sustainable Development Goals (SDGs) *(1)*. A reduction in the frequency of bloodstream infections due to methicillin-resistant *Staphylococcus aureus* (MRSA) and *Escherichia coli* resistant to third-generation cephalosporins has been proposed as an indicator of progress towards SDG 3, "Ensure healthy lives and promote well-being for all at all ages" *(1)*. Moreover, AMR infections in livestock endanger sustainable food production and food security (SDG 2: Zero hunger, and SDG 8: Decent work and economic growth) *(5)*.

Data on antimicrobial-resistant pathogens are essential to inform public health policy and to monitor the effectiveness of interventions. Currently, AMR surveillance relies mainly on microbiological characterization of isolates and phenotypic antimicrobial susceptibility testing (AST). Addition of molecular methods can in some cases provide better understanding of the mechanisms of resistance and the relatedness of strains for investigating the emergence and spread of AMR. Molecular diagnostic tests for AMR are becoming available for use in surveillance, as outlined in a previous WHO document *(8)*. Whole-genome sequencing (WGS) offers a vast amount of information and the highest resolution for molecular subtyping of pathogens. When used to answer specific questions, it could further strengthen AMR surveillance. During the past decade, WGS has transformed biomedical research and could transform epidemiological surveillance of pathogens and aid clinical decision-making on infectious diseases and the treatment of individual patients ("precision medicine"). New microbial genomes are sequenced daily and added to large databases of genome and gene sequences. Improvements in sequencing technologies and analysis have rapidly increased the output and speed of analysis and reduced the overall cost of WGS *(9)*, although price reduction has slowed down, and proprietary tools may still be costly.

Potential public health uses of WGS include:

- identification of high-risk or AMR clones regionally and globally;
- identification of AMR mechanisms and how they arise and are transmitted in human and animal populations and via environmental sources ("One Health");
- identification and tracking of outbreaks;
- identification of new targets for antimicrobials and vaccines; and
- development of point-of-care tests for AMR.

Strengthening research and development of novel diagnostics and treatments is especially important because there is currently a shortage of products in the pipeline that could be used against strains that are resistant to available treatment. Nevertheless, WGS still has some important limitations and poses a number of practical challenges for broad implementation. In addition to strengthening their surveillance systems, countries should build capacity to respond to the results of surveillance. Effective response teams are required to stop the emergence and spread of AMR.

1.1 Purpose of the document

The purpose of this document is to discuss the benefits and limitations of current WGS technologies for AMR surveillance, report the results of an analysis of application of WGS to single bacterial pathogens for AMR surveillance and outline the requirements for building new WGS laboratories or upgrading existing laboratories to ensure capacity for WGS.

WGS can be used for single bacterial isolates, or all the DNA in a composite bacterial sample can be sequenced by next-generation sequencing, a technique also known as "metagenomics". Metagenomics analysis has shown promise, but further standardization of extraction protocols *(10, 11)*, databases, bioinformatics tools, sample collection methods and tools, storage and transport conditions and the coverage of sequencing instruments is required. The greatest hurdle to using metagenomics for detection of AMR is linkage of the AMR gene, which is frequently located on a plasmid, to a specific pathogen. This document focuses on WGS of pathogens isolated in pure culture.

These points are discussed with regard to surveillance laboratories with different capacity, in three principal categories:

- local laboratories,
- subnational laboratories serving regions,
- national reference laboratories (NRLs) and regional reference laboratories (RRLs), which serve all countries in a defined world region.

Laboratories in each category may be newly established or have various levels of experience in using microbiological methods, including WGS and other molecular methods of pathogen typing.

1.2 Whole-genome sequencing

WGS is a molecular biology tool used to obtain the (nearly) complete DNA sequence of an organism. By comparing a DNA sequence to standard reference sequences of well-characterized microbial genomes, inferences can be drawn about important phenotypic traits of an organism, such as AMR. Thus, it is in principle possible to test for the presence of all known AMR genes and mutations associated with resistance (i.e. those that have been uploaded to a specific database) during the same analysis to help predict phenotypic AMR to a broad array of drugs. It is important to emphasize that phenotypic testing will still be required for pathogens in which the presence of a gene does not accurately predict resistance and also to identify the emergence of any new resistance genes and mechanisms.

Relatively few microbial genomes have yet been well studied, and strong concordance between genotypic and phenotypic profiles has been established for only a small number of organisms and antimicrobial classes. WGS also has limitations for AMR surveillance, in that it can be used to identify and interpret only known AMR mutations or genes or novel AMR genes that are similar to known ones. Nevertheless, for some bacterial species, WGS has begun to completely replace other, lower-resolution, less granular typing methods, such as pulsed-field gel electrophoresis, multiple-locus variable number tandem repeat analysis and serotyping (particularly for organisms with large numbers of clinically relevant serotypes, such as *Salmonella*) and may do so for others in the near future *(12)*.

1.3 Review of the literature on use of whole-genome sequencing in surveillance of antimicrobial resistance

A literature search for papers on use of WGS in AMR typing and surveillance was conducted for this publication. Pubmed and Web of Science were searched for papers published up to December 2019 with the search terms: whole-genome sequencing AND ((antimicrobial OR antibiotic) AND resistance) AND surveillance. For inclusion in the review, papers had to:

- include analysis of GLASS priority pathogens;
- describe use of WGS for AMR surveillance or clinical diagnostics of AMR; and
- address human infections (including those transmitted by animals, e.g. infected farm workers, sewage samples from hospitals and foodborne illness).

Papers were excluded if they described only development of new laboratory methods, without application; WGS of microbial pathogens for purposes other than AMR detection; or only infections in animals or environmental samples unrelated to hospitals.

The search showed that use of WGS has increased considerably in the past decade (Fig. 1). While pathogens currently monitored under GLASS were the topic of over half the published studies on WGS, most were conducted in HIC, with 56 from the United Kingdom of Great Britain and Northern Ireland and 105 from the USA (Fig. 2). Furthermore, although more and more studies on WGS for AMR surveillance are published, the sequence data and metadata have not necessarily been made publicly available. Sharing of high-quality data is essential to improve public health decision-making; however, data on pathogen sequences and the associated metadata are considered to be sensitive, as discussed in section 2.4.

Figure 1. Annual numbers of publications on use of WGS for AMR surveillance of GLASS priority pathogens

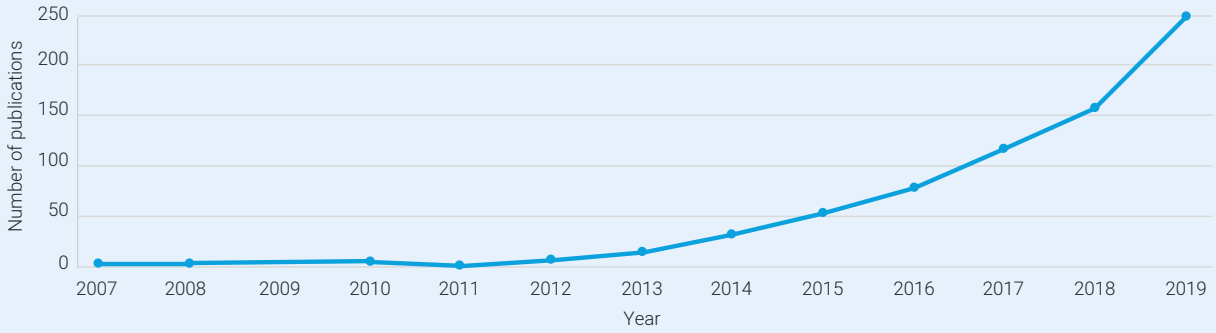
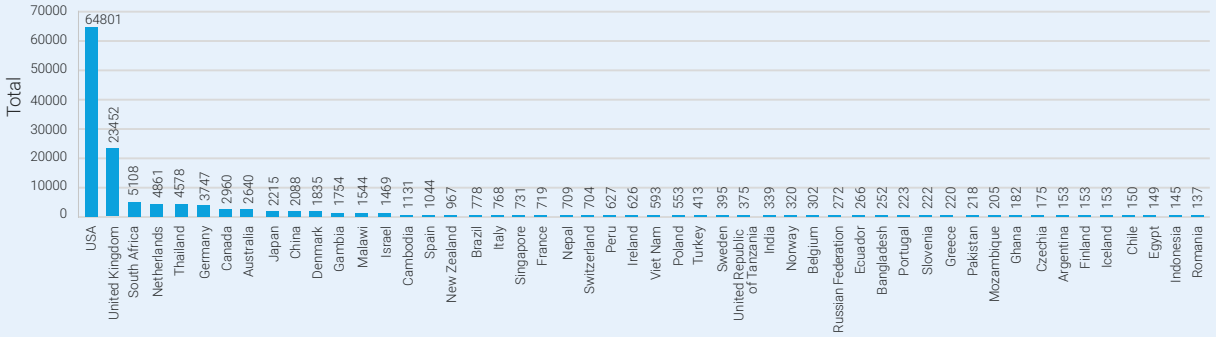


Fig. 2. Numbers of sequenced isolates of GLASS priority pathogens by country of origin in the European Nucleotide Archive.



Only the first 50 countries in terms of isolate numbers are shown. Numbers above bars are the numbers of sequenced isolates. The Archive contained 141 210 sequences of GLASS priority pathogens from 126 countries as of July 2019.

02 Advantages and limitations of whole-genome sequencing in surveillance of antimicrobial resistance

2.1 Phenotypic methods and whole-genome sequencing for characterization of antimicrobial resistance

Surveillance of AMR relies on characterization of antimicrobial-resistant pathogens and their distribution in the population. Phenotypic tests and WGS address different aspects of fast-growing bacteria: phenotypic tests are used to characterize how the bacteria respond in the presence of an antimicrobial, while WGS can be used to characterize the genome of the isolate.

For AMR surveillance, the two types of test can be used in a complementary manner according to the objectives. Both types of test have limitations.

Most AST is routinely performed with standard phenotypic methods, either by reference broth microdilution (standard of the International Standards Organization) or a surrogate phenotypic test, such as disc diffusion or a gradient or semi-automated test. In these tests, bacteria are exposed to different concentrations of antimicrobials, and their ability to grow is tested by estimating the minimum inhibitory concentration (MIC), a zone diameter or a surrogate value, which can then be interpreted against internationally standardized breakpoints to determine whether the pathogen is susceptible or resistant (Clinical and Laboratory Standards Institute and European Committee on Antimicrobial Susceptibility Testing standards).

Like any other method, phenotypic AST has intrinsic limitations, and its application requires continuous improvement. Optimal application of phenotypic methods is described in standards such as those of the Clinical and Laboratory Standards Institute or the European Committee on Antimicrobial Susceptibility Testing. The limitations include potential ambiguity in interpretation if the measured MIC is close to the zone of inhibition, maintaining the appropriate temperature, pH and atmospheric conditions and ensuring the right concentrations of ions in the culture medium with the disc diffusion method. Additional methodological problems for some drugs and bacterial species include the fact that disc diffusion tests are not recommended for some antibiotics, such as colistin, are difficult to use for slow-growing and fastidious bacteria and can be influenced by physical and chemical factors such as incubation temperature and the content and evaporation of growth media *(13)*. Additionally, standard approaches may not be suitable for anaerobic bacteria or rare bacterial species, and there may be no clinical breakpoints. These limitations are not addressed by WGS.

The limitations and challenges of WGS are discussed in section 2.4.

2.2 Consideration of the objectives of antimicrobial resistance surveillance

The first step in deciding which tests should be available for AMR surveillance is to define the objectives, which should advance strategies to tackle AMR. Objectives such as analysis of trends in AMR rates, assessment of the frequency of AMR infections and their impact on human health, data to inform the national list of essential antimicrobial medicines and data to inform treatment guidelines can be fully met by using phenotypic methods. For some other objectives, WGS can complement phenotypic methods but are not essential. For example, the detection and control of the spread of AMR in settings with high selection pressure for AMR, such as health care facilities, have been successfully achieved with phenotypic methods, and the accuracy has been improved by molecular methods such as pulsed-field gel electrophoresis and multiple-locus variable number tandem repeat analysis, which have lower resolution and are less granular typing methods than WGS.

The unique feature of WGS is that it provides nearly complete information on the genome of an isolate, which can be used to understand the genetic basis of AMR mechanisms and differentiate phenotypically identical isolates with the same AST profile. This type of molecular information can be used in the development of novel diagnostics and treatments for AMR. It also allows the location of AMR determinants on the bacterial chromosome or on plasmids, which provides valuable information on the pathways of AMR spread. Comparison of the whole DNA sequences of different isolates can supplement contact tracing and phenotypic antibiograms to reconstruct transmission chains. More benefits of WGS for pathogen surveillance, public health and clinical practice are described below.

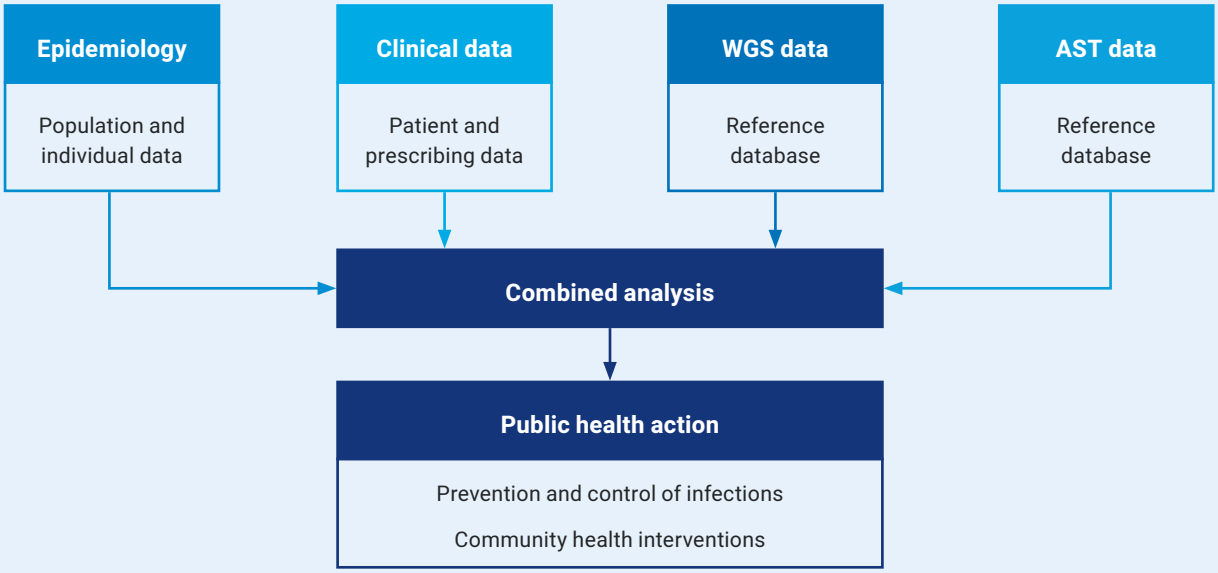
2.3 Potential benefits

2.3.1 Benefits for public health

- WGS can provide new insights into disease transmission and virulence and the dynamics of AMR when combined with epidemiological, clinical and phenotypic microbiological information. These insights provide useful information for risk assessment and for designing effective interventions.
- WGS can be used to identify and characterize pathogens rapidly and precisely. When combined with epidemiological information, it can facilitate linkages during the early detection phase of outbreaks, accurate tracing of transmission chains, precise delineation of the geographical spread of an outbreak and identification of sources of infection. Timely outbreak detection and removal of sources of infection can lead to substantial cost savings in public health.
- WGS allows genome-wide analysis and high-resolution subtyping of AMR pathogens, including characterization of AMR elements (e.g. subtyping of plasmids).
- WGS data are digital, and tests are done on computer systems. Consequently, use of WGS enables better standardization and reproducibility, providing greater inter-laboratory comparability than phenotypic testing. The fact that sequence data are digital and different algorithms can be used to analyse the same set of data also means that old data can be re-analysed, ensuring backward compatibility between new and old analyses.

WGS data can be used to verify the identity of AMR mechanisms in isolates with relevant phenotypic AMR or with discordant phenotypic AMR. It cannot, however, be used to quantify the level of phenotypic AMR, so that it is unsuitable for routine or predictive AST and therefore cannot replace phenotypic methods. Fig. 3 shows how WGS, AST and clinical and epidemiological data together can result in action for public health. Thus, epidemiological and clinical analysis provide individual and population data that can be used with the results of WGS and phenotypic AST in well-curated reference databases to design effective strategies for infection control and prevention and public health interventions. For example, AST data could indicate that a high proportion of strains of a bacterial pathogen in a region are resistant to several antibiotics, suggesting a new type of AMR. WGS and epidemiological data could be used to trace the transmission chain or locate AMR genes on mobile genetic elements, and this information could be used to infer what favoured the evolution of multidrug resistance in a certain pathogen.

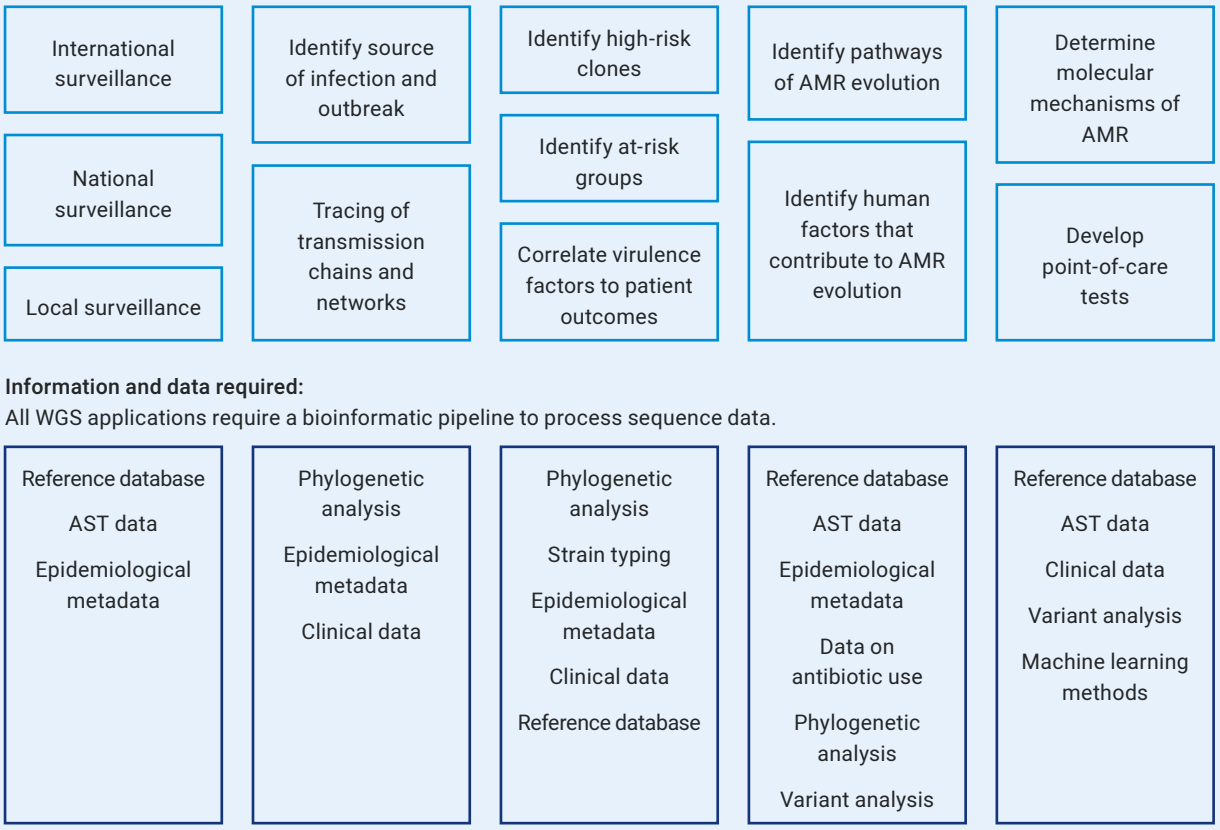
Fig. 3. Interlinkage of different data types for actionable public health results



The results of WGS alone cannot provide actionable public health results. Rather, the combined insights from all these data sources are required as a basis for public health guidelines to prevent the emergence and dissemination of new MDR pathogens.

Fig. 4 summarizes the general and pathogen-specific uses of WGS in AMR surveillance and the types of data and methods required. WGS can be applied for any pathogen in several ways. The information and methods required refer to broad categories of data and technology for WGS analysis in AMR surveillance.

Fig. 4. Applications of WGS in AMR surveillance



WGS can be applied to any pathogen in several ways. "Information/methods required" refers to broad categories of data and technology for WGS analysis for AMR surveillance. WGS can be used for international, national and local surveillance of AMR. Surveillance requires reference databases for genomic and AST data and can be combined with epidemiological metadata. Targeted surveillance can be used to identify the source of an outbreak of AMR infections and for tracing transmission chains or networks. This requires tools for phylogenetic analysis, epidemiological metadata and clinical data. WGS AMR surveillance data can be used to identify high-risk clones and groups at risk of infection and to correlate virulence factors with patient outcomes, which requires phylogenetic analysis and strain typing, in combination with epidemiological and clinical metadata and genomic reference databases. WGS AMR surveillance data can also be used to identify pathways of AMR evolution and human factors that contribute to the evolution of AMR. This requires genomic reference databases, AST data, epidemiological metadata, data on antibiotic use and tools for phylogenetic analysis and variant analysis. Moreover, WGS data can be used to identify the molecular mechanisms underlying AMR and to develop molecular point-of-care tests. This requires genomic reference databases, AST data, clinical data and tools for variant analysis and can be facilitated by machine learning methods.

Local uses of WGS for AMR surveillance include:

- detection of known AMR mechanisms;
- identification of novel AMR mechanisms, with phenotypic AST data and characterization as e.g. plasmid-mediated or clonal; and
- analysis of an outbreak at a single centre, such as a hospital.

Local, regional or national uses of WGS for AMR surveillance include:

- comparison of several genomes from different sites;
- analysis of local or regional transmission networks; and
- tracing sources of local or regional outbreaks.

International uses of WGS for AMR surveillance include:

- monitoring of pathogen populations;
- detection of high-risk and AMR clones; and
- assessment of the impact of interventions.

WGS analysis at local level is not necessarily easier or more feasible for low-resource settings than WGS analysis at regional or global level. Case study 1 (section 3.1.3) illustrates use of WGS for AMR surveillance in a local facility (a hospital). This type of analysis would, however, be difficult to conduct with no previous experience in WGS analysis. Section 3 gives examples of the use of WGS in AMR surveillance locally, nationally and internationally. Details of the infrastructure required are given in section 4.2.

2.3.2 Benefits for laboratories and data

- WGS is a streamlined technology, which can be applied to a wide array of microbial pathogens, once they are isolated *(14)*. This permits simultaneous monitoring and discrimination of co-infection pathogens, even if they have very similar phenotypic characteristics. The laboratory workflow can sometimes be automated and thus more efficient.
- WGS data are amenable to standardized analyses, and the raw data and results can easily be shared, facilitating rapid, equitable knowledge transfer. It can make global AMR surveillance more digital, with timely access to information on emerging AMR.
- The cost of testing for all known AMR mechanisms at once is only marginally higher than that of testing for only one mechanism, whereas, with phenotypic AST, testing all the necessary antimicrobials would require more reagents, materials and staff time.
- WGS can assist investigation of pathogen transmission *(15)*

2.3.3 Benefits for clinical practice

Although this document focuses on application of WGS in AMR surveillance, research on its application for clinical practice with regard to fast-growing bacteria is increasing rapidly.

- Comparison of the genomes of pathogens associated with recurrent disease can be used to differentiate between endogenous relapse (usually after failure of treatment or colonization) and exogenous re-infections, thus shaping public health responses.
- Global and regional knowledge gained from WGS of pathogens will support the development and interpretation of novel, targeted molecular diagnostics for AMR that can be used widely as rapid point-of-care tests.
- Predictions of AMR based on sequence data can be triangulated by comparison with archived sequences with AST results. If an isolate has the same or a similar sequence as that of a previously seen isolate and AST data indicate that the previous isolate was resistant, this provides additional confidence in the predicted AMR and builds a virtuous cycle of evidence that can ultimately be of direct clinical relevance.

Despite its benefits, WGS may not always be the most appropriate microbial typing method for surveillance purposes (see section 2.2 for more details). Purposes such as investigation of cross-transmission of pathogens or characterization of AMR mechanisms may require the use of several other molecular methods (see Annex 1). A major limitation of WGS as compared with other molecular methods is the high initial and recurrent investment and the complex technological and infrastructure requirements. For example, targeted sequencing or screening PCR, rather than WGS, can be effective for rapid sequencing of pathogen genes directly from clinical samples and when the presence of specific genes must be confirmed or excluded rapidly (Box 1). In “targeted” sequencing, instead of the entire genome, only the gene or genomic region of interest is sequenced, which is a rapid, cost-effective means of confirming specific AMR genes or mutations. Alternatively, simple molecular diagnostics, such as lateral flow immunochromatographic assays (based on the detection of proteins by antibodies), are frequently used to detect AMR.

BOX 1. MOLECULAR METHODS FOR AMR SURVEILLANCE

- **Targeted detection** for AMR surveillance *(1)* comprises detecting AMR genes or mutations. Sequencing of a pathogen's DNA is not always necessary for this purpose. For example, PCR can be used to detect AMR genes or mutations by selective amplification; microarrays can be used to detect AMR genes and mutations by the binding of pathogen DNA fragments to known complementary DNA sequences; and lateral flow assays detect the presence of AMR genes by the binding of antibodies to their protein products if the genes are expressed.
- **Targeted sequencing** involves use of PCR to amplify known AMR genes to confirm their presence or to identify resistance-conferring mutations.
- **Whole-genome sequencing** is used to determine the (nearly) complete DNA sequence of a pathogen, including parts of the genome that do not contain known AMR genes. Known resistance genes can be identified by comparison with reference DNA. The genome sequence can be used in combination with phenotypic AST data to identify novel AMR genes and mutations.

WGS is useful for analysing individual isolates with complex antimicrobial susceptibility profiles or multiple isolates with considerable genetic heterogeneity in the same AMR phenotype to determine the mechanism of AMR. In outbreaks, in which isolates with the same AMR phenotype have highly similar underlying genotypes and resistant isolates of the same strain, WGS can be used to infer the relatedness among isolates and hence transmission chains (Box 2). Other methods, such as pulsed-field gel electrophoresis, have been used for to assess clonal relatedness among pathogens in investigation of cross-transmission.

Considerations of the cost–effectiveness of WGS for AMR surveillance are outlined in section 2.5.

BOX 2. POTENTIAL BENEFITS OF WHOLE-GENOME SEQUENCING

- inference of phenotypic traits, e.g. AMR, and virulence factors;
- detection of novel molecular markers of drug resistance;
- rapid identification of AMR strains that cause infections and better public health decision-making because of its high sensitivity;
- highest possible resolution of molecular isolate identification and subtyping currently available, thus permitting high-resolution tracing of microbial transmission events and rapid, accurate identification of local, regional and global outbreaks;
- high-resolution tracing of mobile genetic determinants of AMR, such as plasmids and transposons, and accurate delineation of local, regional and global outbreaks of determinants of resistance;
- ready standardization and sharing of raw data for "big data" approaches;
- contribution to comprehensive genome databases linked to epidemiological and clinical metadata as invaluable resources for public health surveillance and clinical care;
- standardization of testing (digital epidemiology) and faster real-time surveillance; and
- potential facilitation of development of targeted diagnostics and novel antimicrobial treatments and vaccines.

2.4 Current limitations of use of whole-genome sequencing for surveillance of antimicrobial resistance

Despite the benefits of WGS for AMR surveillance, a number of limitations should be considered in making a balanced decision about its introduction for public health and clinical applications (Box 3).

BOX 3. CURRENT LIMITATIONS OF WGS FOR AMR SURVEILLANCE

- WGS technologies require substantial initial and sustained financial investments.
- Sequencing and bioinformatics are not part of the general knowledge or training of staff in laboratories in LMIC, and investment in training and continuous education of staff must be secured.
- Standard operating procedures, QA protocols and evidence-based guidelines should be developed for use of WGS in AMR surveillance.
- For most pathogens and antimicrobials, the predictive sensitivity and specificity of WGS for inferring AMR phenotypes are still too low for practical application.
- Data-sharing is not currently standard practice.

- **WGS technologies require substantial initial and sustained investments** in laboratory equipment, computing infrastructure and training. WGS may therefore not be cost–effective at small sites without the necessary infrastructure. When funds for the initial investment are not available, sending samples to a central sequencing laboratory (e.g. RRL, NRL) should be considered.
- **Internationally agreed standard operating procedures, quality assurance (QA) procedures and regulatory guidelines for WGS in AMR surveillance do not exist** and should be developed to ensure that WGS results from different laboratories are comparable and easy to collate and interpret. Protocols used in other surveillance programmes in which molecular methods and WGS are used (Global Microbial Identifier, US Centers for Disease Control and Prevention, US Food and Drug Administration, WHO surveillance of foodborne diseases, HIV drug resistance and MDR-TB) could be used as a starting point. International expert panels should agree on reference strains and AMR genes and mutations that can be used for quality control (QC).
- **WGS may result in false-positives or false-negatives if it is not complemented with phenotypic AST.** A result indicating a resistant phenotype and a positive molecular result implies expression of the AMR gene; however, genes or pseudo-genes may be present but not expressed, which can falsely predict AMR if a molecular test is used alone. Failure to detect the presence of a new unknown AMR gene can result in false prediction of the absence of AMR.

- **The sensitivity and specificity of genomic databases and algorithms for predicting phenotypic AMR vary by bacterial species, type of AMR and genomic database.** As the AMR mechanisms relevant to each organism or drug combination are better defined genotypically, WGS analyses should result in fewer false-negative predictions (a major error). With a few exceptions (e.g. *M. tuberculosis*, MRSA, foodborne pathogens *(16)*), genotypic prediction of AMR is not yet suitable for making decisions on the treatment of most bacterial pathogens *(16–18)*. The sensitivity and specificity of WGS-based AMR prediction will fluctuate over time, as more AMR mechanisms are discovered. Ongoing research and the expansion of surveillance databases with WGS and AST data will eventually increase the sensitivity of genotypic AMR predictions to the degree required for clinical and public health decision-making. Phenotypic data will still be required, however, to detect emerging mechanisms of AMR.
- **The development of QA protocols and guidelines for surveillance and clinical use must be coordinated.** GLASS, the central WHO programme for AMR surveillance, can provide the framework for such coordination, building on regional surveillance systems, such as those coordinated in Europe by the European Centre for Disease Prevention and Control (e.g. EURO-GASP, EURGen-Net) and experience in molecular case definitions, training, standardization and external QA in these networks. Widely applicable protocols will require sharing of pathogen sequence data, and therefore international agreements will have to be discussed.

2.5 Cost and economic advantages of whole-genome sequencing for surveillance of antimicrobial resistance

The economic advantages of WGS for AMR surveillance have not yet been proven unequivocally. The cost of establishing one or more WGS laboratories depends on existing laboratory facilities and the country. Not every country requires WGS facilities in order to participate in AMR surveillance that includes WGS. Regional "hub-and-spoke" surveillance systems in which participating countries with no WGS capacity send samples to an RRL with WGS capacity are an alternative. The cost of setting up a new WGS laboratory depends on the intended sequencing capacity, whether culture facilities are required, space requirements and available public infrastructure. In settings where additional investment in laboratory infrastructure is necessary, the introduction of WGS surveillance may not be a priority. For laboratories with molecular surveillance capacity (e.g. pulsed-field gel electrophoresis), the cost of introducing WGS has been estimated to be US$ 100 000–700 000 *(12)*, with additional costs for hiring and training an expert WGS surveillance team.

WGS of a single bacterial isolate costs US$ 35–300 in HIC, where the lower value is achievable only with very high throughput *(19)*, when the cost can be offset by consolidating the workflows for multiple isolates. WGS thus becomes more cost–efficient in higher-throughput laboratories. Consequently, when deciding whether to establish one large laboratory with greater capacity for WGS (e.g. an NRL or RRL) or several smaller ones, the first option is more cost–efficient.

Published estimates of the cost of WGS per isolate are similar to those of conventional molecular typing techniques: US$ 25–150 for pulsed-field gel electrophoresis and US$ 65–120 for multi-locus sequence typing *(19, 20)*. The estimated cost of repetitive element sequence-based PCR is markedly lower, at US$ 26 per isolate *(21)*. Thus, introducing a single WGS workflow may save costs, as it would obviate the need to replace or establish many conventional molecular techniques for pathogen typing.

As WGS becomes more widely used and more cost-efficient sequencing technologies are developed, its use in national surveillance systems will become more affordable by more countries. The greatest cost associated with microbial WGS is, however, that for the preparation of sequencing libraries, which has remained stable at US$ 60–74 per isolate since 2011 *(22)*. WGS is used in many different applications in medicine and biomedical research and is not restricted to AMR surveillance or pathogen genomics, unlike phenotypic AST, which is performed only in AMR surveillance and clinical microbiology. Consequently, the market forces that are driving down the cost of WGS technology should be much stronger than for phenotypic methods, and the prices are expected to decrease faster.

A major economic benefit of surveillance is preparedness, which is frequently overlooked because the cost of lack of preparedness is difficult to estimate. It has been estimated that a complete lack of preparedness for AMR between now and 2030 would increase extra health care expenditure to US$ 0.22 trillion annually, even in a setting with low AMR *(2)*. Good surveillance can help health systems to prevent major outbreaks of AMR pathogens and to be prepared if they do occur. As thorough understanding of the transmission dynamics of AMR organisms depends on WGS data, inclusion of WGS in AMR surveillance systems can have major economic benefits.

In terms of clinical usefulness, WGS may be marginally cheaper than routine diagnostic methods for some pathogens *(22, 23)*, and the cost of WGS in clinical applications should be compared with the potential savings made by shortening and controlling hospital outbreaks. A study on WGS for the analysis of a hospital outbreak of MRSA in the United Kingdom concluded that the health care costs associated with the outbreak were in excess of £ 10 000 (US$ 12 167), while the cost for WGS of one MRSA isolate (including sample preparation, library QC and sequencing) was £ 95 (US$ 116); i.e. sequencing all 26 isolates from the outbreak cost £ 2470 (US$ 3005) *(24)*. In this example, WGS analysis led directly to an intervention that stopped the outbreak. Similarly, modelling of the cost–effectiveness of WGS surveillance of MRSA suggested that WGS was more cost–effective than current surveillance *(23)*.

03 Examples of use of whole-genome sequencing in surveillance of antimicrobial resistance

3.1 Local case studies: analysis of single isolates or an outbreak at a single site

Individual isolates from a single site (such as a hospital, a health care centre or a city) can be analysed by WGS to identify AMR genes and determine whether there is transmission. The genome sequences of several isolates from the same clinical centre can be compared to confirm whether a local outbreak has occurred, allowing the design of strategies to manage it *(9)*.

For example, in a hospital in Beijing, China, WGS was used to show that an MDR strain of *Klebsiella pneumoniae* had been circulating in the hospital for a year before detection of the outbreak, mainly in intensive care units *(25)*. The genome sequence of an isolated pathogen and the location of AMR determinants on the core bacterial genome or on plasmids can indicate the risk of AMR spread, as bacteria of the same or different species commonly exchange DNA via plasmid conjugation. For example, WGS of multiple extended-spectrum β-lactamase (ESBL)-producing *K. pneumoniae* strains in a geriatric care ward in Australia indicated that the referring hospital was the source of AMR and where strain-to-strain transmission of a blaCTX-M-15 FIBK/FIIK plasmid occurred *(26)*. Rectal screening for ESBL organisms on admission to geriatric wards was recommended to guide patient management and infection control in such facilities.

3.1.1 Requirements

A limited number of isolates can be studied by WGS in laboratories established for local pathogen typing and surveillance. Laboratories should have experience in microbiological methods and the capacity to isolate bacteria from clinical specimens and isolate DNA. Personnel must be trained in WGS and bioinformatics. Contamination must be avoided to obtain meaningful results and should be assessed and verified when analysing the WGS data with software such as Kraken. Laboratories should have a computer that can store and process at least several gigabytes of sequence data or access to cloud servers. In addition, Internet access is required for many sequencers and reference databases. The latter may be downloaded, but this still requires temporary Internet access. Local laboratories may not require very high-throughput sequencing instruments such as HiSeq or NextSeq, but rapid results can be important, especially when the goal is to improve patient care or contain a local outbreak. These requirements may be too difficult to achieve in the vast majority of local and clinical laboratories, and samples can be sent to an NRL with bioinformatics and epidemiology expertise.

3.1.2 Reference data and tools

High-quality, curated, validated reference data are essential for accurate identification of pathogens and AMR genes. To ensure the comparability of results among sites, agreement should be reached on the AMR reference databases, analytical tools and workflow to be used. Table 4.2 in the WHO paper on WGS for foodborne disease surveillance *(12)* and Table 2 in Hendriksen et al. *(27)* list publicly accessible AMR and virulence databases for bacterial pathogens. Annex 3 includes a summary of these tables.

3.1.3 Case study 1: Use of whole-genome sequencing in resolving a local outbreak of methicillin-resistant *Staphylococcus aureus*

From reference 24.

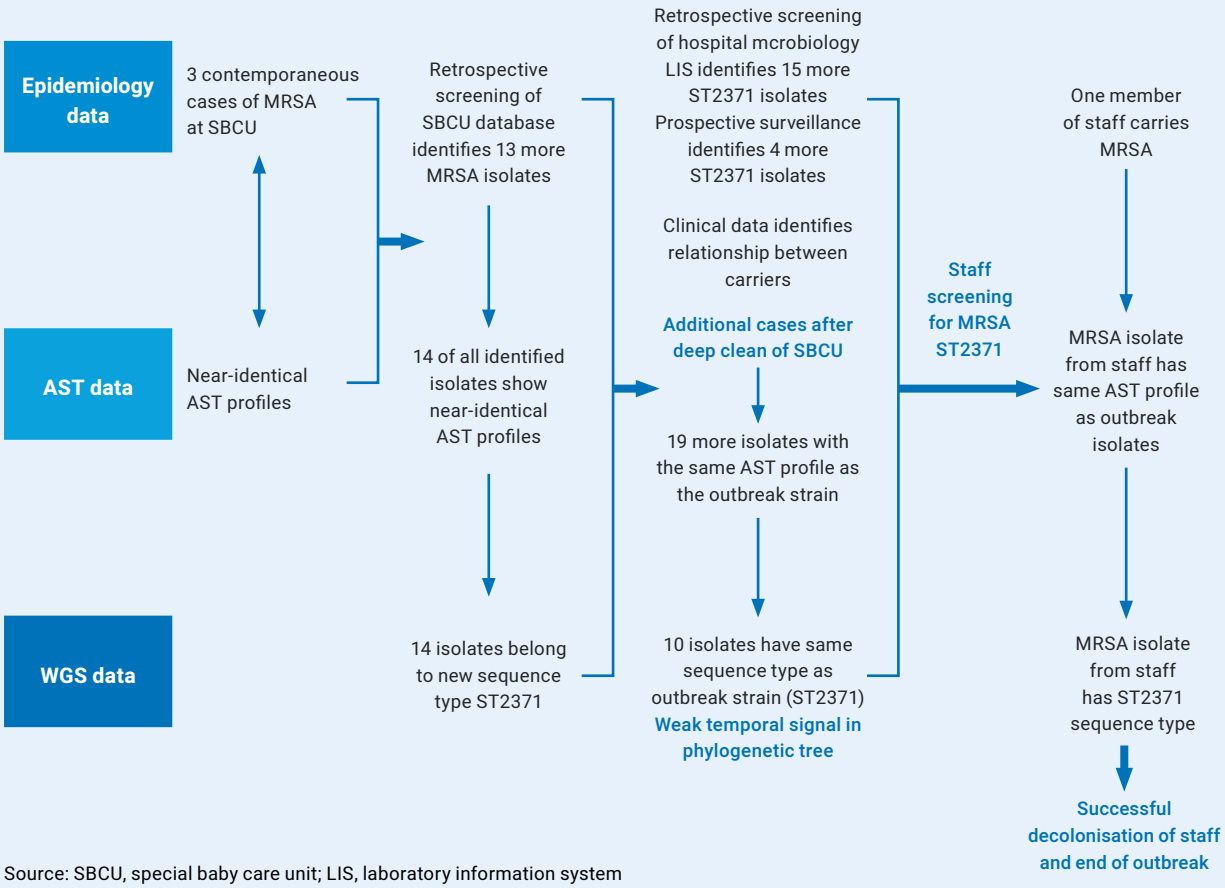
The infection control team at an infant care unit at Cambridge University Hospitals National Health Service Foundation Trust in the United Kingdom identified three contemporaneous cases of MRSA, which they suspected were linked on the basis of near-identical antimicrobial susceptibility (determined by disc diffusion). A review of MRSA isolates from the unit during the preceding 6 months identified 13 additional isolates that were putatively linked to the outbreak. As the cases occurred at three discrete times, the team could not confirm that all the isolates were from an extended outbreak. The unit was thoroughly cleaned, and surveillance was continued by weekly screening for MRSA, which nevertheless indicated a new confirmed infection with MRSA.

WGS showed that 14 of the 17 isolates belonged to a new MRSA sequence type (ST2371); three other isolates were the previously known sequence types ST1, ST8, or ST22. All the ST2371 isolates were found to be closely related by single-nucleotide polymorphism and phylogenetic analyses. WGS confirmed that the infection control team had correctly excluded isolates ST1, ST8 and ST22 from the outbreak according to their AST profiles but had incorrectly excluded two isolates because of erroneous initial AST results, which were later repeated and corrected.

To determine whether the outbreak strain was being transmitted in the community outside the special infant care unit, all MRSA isolates identified since the beginning of the outbreak with the same AST profile as the outbreak strain were selected from the Cambridge University Hospitals microbiology laboratory information system (Fig. 5). Ten more ST2371 isolates were identified. The sources of sample submission and the hospital information system were used to investigate an epidemiological link between the affected individuals and the infant care unit. A complex transmission network was reconstructed that included transmission pathways from infants to their mothers, from the mothers to other mothers in the postnatal ward and from mothers to their partners. All individuals were treated with MRSA decolonization therapy.

The re-emergence of ST2371 after thorough cleaning of the special infant care unit and the structure of the phylogenetic tree of the sequenced isolates, which did not show a strong temporal signal of patient-to-patient transmission, suggested that the source of infection was external to the unit. This conclusion could not have been reached on the basis of phenotypic AST data alone, which would not have allowed an assessment of relatedness among very similar isolates. All infant care unit staff were then screened for MRSA, and one staff member was identified as a carrier of ST2371. After successful decolonization of this person, the outbreak was stopped.

Fig. 5. Example of the use of WGS in resolving a local outbreak of MRSA



Source: SBCU, special baby care unit; LIS, laboratory information system

3.2 Subnational and national case studies: combination and comparison of multiple genomes from different sites

The whole-genome sequences of isolates can be combined at regional or national level to determine the geographical spread of a resistant strain and to trace transmission networks and identify at-risk groups. For example, WGS was used to trace the origin and AMR of a meropenem-resistant *Streptococcus pneumoniae* serotype after introduction of pneumococcal conjugate vaccines in Japan *(28)*. Another example is PulseNet, a programme in which WGS is used for molecular typing to identify outbreaks of enteric pathogens *(29)*. In Victoria, Australia, a combined genomic and epidemiological approach was used to identify different nosocomial transmission networks of *K. pneumoniae* carbapenemase-producing Enterobacterales (formerly known as Enterobacteriaceae), which led to infection control measures in real time and formulation of new surveillance and management guidelines *(30)*. A combined genomic and epidemiological approach was also used to identify a cross-border nosocomial transmission event in Europe among previously hospitalized travellers, which was traced to a local outbreak due to an OXA-48 carbapenemase-producing *K. pneumoniae* ST392 strain *(31)*.

3.2.1 Requirements

RRLs or NRLs should be established in order to include WGS in pathogen surveillance at regional or national level, with the capacity for high-throughput sequencing, computers with sufficient storage and processing power to analyse sequence data from isolates collected throughout a country and experienced personnel trained in WGS and bioinformatics. In addition, there should be a formal surveillance network to organize the collection of bacterial isolates and demographic, clinical and epidemiological metadata and to conduct sequencing, analysis and interpretation of results. NRLs should then develop the capacity for rapid response to outbreaks. The main challenge is in increasing data collection and storage capacity, from a low volume of data on susceptibility to data from WGS, which could be thousands of times larger.

3.2.2 Reference data and tools

RRLs and NRLs should maintain not only global reference databases for typing pathogens and identifying AMR genes but also their own databases of the epidemiology of AMR organisms in their area that link WGS data to relevant phenotypic AST data and epidemiological and clinical metadata. Data from WGS of isolates from different geographical sites can be used to infer mechanisms of AMR acquisition, clustering of different strains and transmission routes. For these purposes, laboratories should be equipped with tools for phylogenetics and data visualization, which should be designed for use in community health facilities and be flexible enough for use in different analytical systems. Open-access tools are available to more users, including those with insufficient funding for proprietary software. Open-source tools have a further advantage of being scrutinized by the user base, which can result in faster correction of software errors. Users with access to the source code can use it as a basis for developing novel software solutions to hitherto unresolved problems, which could in theory shorten the time to new scientific results. Ideally, epidemiologists external to reference institutions would be able to interpret high-level data and make appropriate deductions.

3.2.3 Case study 2: Integrating whole-genome sequencing into the national surveillance programme for antimicrobial resistance in the Philippines

Contributed by Dr Sonia Sia, Research Institute for Tropical Medicine, Philippines. Details provided by Agrimón et al. *(32)*.

The Philippines Department of Health established the national Antimicrobial Resistance Surveillance Program in 1988. The programme is based on phenotypic AST in sentinel site laboratories and now comprises 24 sentinel sites (22 Government hospitals and two private hospitals) and two gonorrhoea surveillance sites in the 17 regions of the country. Cases are identified in specimens of priority pathogens sent routinely to laboratories for clinical purposes. The programme issues annual reports on resistance rates and trends in bacteria of public health importance and has contributed data to GLASS since 2015. The frequency of AMR has increased during the past 10 years.

Local staff capacity to perform sequencing and bioinformatics analysis was developed by training and mentoring provided by the "See and Sequence" project, a collaborative research project with the Centre for Genomic Pathogen Surveillance in the United Kingdom *(33)*. Training began with DNA extraction, library preparation and hands-on sequencing with Illumina MiSeq devices, followed by establishment and local optimization of the assay. Bioinformatics training included assembly, annotation variant calling and phylogenetic analysis of bacterial genomes, followed by mentoring as analyses were conducted. Training in public health applications of WGS was complemented by the user-friendly web tools Microreact *(34, 35)* and Pathogenwatch *(36)*.

After staff training, local sequencing and bioinformatics facilities were made available through a successive collaborative research project, the Global Health Research Unit on Genomic Surveillance of AMR, funded by the National Institute for Health Research, in collaboration with the Centre for Genomic Pathogen Surveillance *(11)*.

The Antimicrobial Resistance Surveillance Program made the requisite adjustments to integrate capacity for WGS into national surveillance, which included allocating space for the sequencer and computer servers, revising staff work to include tasks related to sequencing (including time for local optimization of assays) and bioinformatics analysis, sourcing supplies for sequencing, which were limited on the local market, improving collection of metadata and developing competence in genomic epidemiology. In parallel, a large retrospective survey was conducted of eight bacterial pathogens of public health importance. Genomic data were collected on *K. pneumoniae, E. coli, Acinetobacter baumannii, Pseudomonas aeruginosa, Neisseria gonorrhoeae, Salmonella Typhi, non-typhoidal Salmonella,* and *S. aureus,* and isolates collected in 2013–2014 to the United Kingdom for sequencing. This activity provided a baseline genomic context of pathogens in the country for building local prospective surveillance. The sequence data were used for phylogenetic analysis, *in silico* genotyping, detection of genomic AMR determinants and characterization of plasmids carrying carbapenem-resistant genes. The metadata linked to each isolate were: location and date of specimen collection, type of specimen, from inpatients or outpatients, sex and age of patient and whether infections were acquired in the community or a hospital. The study also provided genomic data on a previously undetected outbreak of carbapenem-resistant *K. pneumoniae* in a tertiary hospital. Control measures implemented after identification of the outbreak included designation of a separate MDR organism room for active surveillance after identification of new carbapenem-resistant *K. pneumoniae* isolates from the neonatal intensive care unit and referral of new isolates to the Antimicrobial Resistance Surveillance Reference Laboratory for sequencing.

The laboratory will use WGS and bioinformatics to improve national AMR surveillance data by providing more granular information for investigations of the emergence and transmission of resistant pathogens in order to contribute pertinent information for action to mitigate the emergence and spread of resistant pathogens in the country.

3.2.4 Case study 3: Integrating whole-genome sequencing into laboratory-based surveillance in the National Reference Laboratory, Argentina

In the National Institute of Infectious Diseases "Dr Carlos G Malbrán", laboratory-based surveillance is used to study the characteristics, dynamics, transmission and diagnosis of human diseases. The institute is also the well-established NRL. WGS was introduced into the surveillance system, starting with pilot and seed projects *(37–41)*, first for foodborne pathogens *(39)* with Genome Trakr and then extending to others (see below). A national genomic platform was created to optimize use of WGS laboratory management resources, informatics infrastructure for genomics analysis, genomic epidemiology and capacity-building for genomic analysis. The advantage of this model is efficient use of expensive, short-shelf-life sequencing reagents, bioinformatics expertise, a coordinated approach to data analysis and efficient informatics infrastructure.

WGS was incorporated into laboratory-based surveillance for specific uses, to add value to the surveillance systems established for each disease:

- studies of outbreaks that could not be resolved with other laboratory techniques;
- emergence of pathogens for which no other technique is available; and
- studies of the dynamics and evolution of global clones that may affect public health surveillance and decision-making.

Case study 3A: Use of WGS to study the emergence of an *N. gonorrhoeae* azithromycin-resistant clone

The Gonococcal Antimicrobial Susceptibility Surveillance Programme was initiated in Argentina in the early 1990s by the NRL for sexually transmitted infections, to monitor the susceptibility of *N. gonorrhoeae* and for molecular characterization of isolates with emerging resistance to antibiotics. WGS was used to study an outbreak (2018–2019) of high- and low-level azithromycin-resistant *N. gonorrhoeae* in two provinces. WGS provided greater resolution and better identification of two clusters circulating in the two regions than usual subtyping methods such as multi-antigen and multi-locus sequence typing. Linkage of WGS and epidemiological data provided valuable information for establishing action plans and public health policies to mitigate the spread of resistant *N. gonorrhoeae*.

Case study 3B: Study of *mcr-1* plasmids in *E. coli* in a One Health approach

After the first description of *mcr-1* colistin-resistant *E. coli* in Latin America *(42)*, nine MDR *E. coli* isolates were recovered in hospitals in three cities between 2012 and 2016. Although they were not clonally related, WGS confirmed the presence of similar IncI2 *mcr-1*-harbouring plasmids *(43)*. At the same time, *mcr-1*-positive *E. coli* isolates were recovered from healthy chickens in commercial broiler farms in several provinces of Argentina. Comparative sequence analysis of the IncI2 plasmids indicated that a group of *mcr-1* plasmids with the same backbones were present in both poultry farm *E. coli* isolates and human clinical isolates *(19)*. The same IncI2 plasmid was also detected in *E. coli* isolates recovered from swine in three provinces. WGS helped to determine the role of this plasmid type in spreading the mechanism of resistance to polymyxins in the country.

Pilot and seed projects with various international partners were essential for capacity-building, implementation, trouble-shooting and generating proof of concept of use of WGS as evidence for stakeholders. More work is required to ensure the sustainability of the system (reagents, supply chain and bioinformatics analysis) to move from proof of concept to routine use in the surveillance system. Participation in validation and more capacity-building in the use and interpretation of genomic data are required, including for the NRL and epidemiologists. The NRL is establishing systematic data analysis, interpretation and reporting for surveillance purposes and pilot projects to evaluate the use of WGS for real-time surveillance of specific pathogens.

3.3 International case studies: monitoring of bacterial clones and populations

Combination of WGS data from several countries makes it possible to monitor pathogen populations and identify and track high-risk clones at the international level. For example, WGS of isolates from five continents showed that most of the MDR *Salmonella enterica* serotype Kentucky circulating globally today resulted from clonal expansion of a single lineage in one country that acquired chromosomal AMR genes several decades earlier *(20)*. WGS for *N. gonorrhoeae* has been included in European AMR surveillance *(44)*. Such surveys are necessary to provide the population context for more targeted investigations *(45)*. Data from large-scale international surveys can be used to design algorithms to predict novel AMR phenotypes from genotypes, as shown for *S. aureus (45)*. Similar multicentre cross-sectional genomic surveys have been conducted in countries in the European Union to determine the geographical distribution and epidemiological drivers of the dissemination of high-risk clones of MDR *K. pneumoniae* and *N. gonorrhoeae (44, 46)*.

3.3.1 Requirements

International use of WGS requires high-throughput sequencing instruments, advanced laboratory infrastructure and teams of highly trained personnel who can efficiently handle and process samples of various pathogens from various sites, which may require different methods of handling, sequencing and analysis. It also requires sufficient storage capacity and plans for data-sharing according to international protocols. Experience in conducting large-scale, structured epidemiological surveys will be necessary but can be supplied by external groups.

3.3.2 Reference data and tools

International WGS studies require extensive reference databases and computationally efficient analysis tools to process large quantities of data and to compare independently generated data with available datasets. Such studies will be possible only if WGS data for AMR surveillance are shared among the members of surveillance networks. Some types of analysis, such as phylogenetic analyses for reconstructing transmission chains and training algorithms based on machine learning techniques to predict AMR from sequence data, require data on raw sequences or at least individual isolates. If raw data cannot be shared, sharing of aggregated data should be considered.

3.3.3 Case study 4: Surveillance of carbapenem-resistant *Klebsiella pneumoniae* in Europe

From references *46* and *47*.

K. pneumoniae is one of the WHO GLASS priority pathogens, which are rapidly developing resistance to last-line treatments. Carbapenem-resistant *K. pneumoniae* is the fastest growing AMR threat in Europe in terms of morbidity and mortality. Biased, fragmented surveillance and lack of standardization in the characterization of isolates have made it difficult to identify reservoirs of AMR and to understand the transmission dynamics. The European Survey of Carbapenemase-producing Enterobacteriaceae was initiated to respond to the need for better understanding of the emergence and spread of carbapenem resistance in *K. pneumoniae* for the design of public health interventions. Between November 2013 and May 2014, laboratories in 244 hospitals in 32 European countries were asked to submit the first 10 isolates of *K. pneumoniae* and *E. coli* not susceptible to carbapenem that they received for diagnostic purposes, with 10 carbapenem-susceptible isolates for comparison for WGS. This approach generated an unbiased, continental, contemporaneous population group of 1717 representative clinical isolates.

A genome-wide search for β-lactam resistance determinants identified five resistome groups: group 1 contained isolates with one or more known carbapenemase genes, group 2 contained isolates without carbapenemase genes but with ESBL or AmpC genes plus porin defects, group 3 contained isolates with *ESBL* or *AmpC* genes but no porin defects, group 4 contained isolates without *ESBL* or *AmpC* genes but with porin defects, and group 5 contained isolates with none of the above determinants of AMR. Phenotypic AST with reference broth microdilution confirmed that group 1 isolates had the highest phenotypic AMR (median MIC, 32), followed by group 2 isolates (median MIC, 1). Concordance between AMR gene detection by WGS in the central sequencing laboratory and by PCR in national expert laboratories was 98.3–99.0%. Transmissibility within hospitals was shown to correlate with the extent of AMR. Phylogenetic analysis further subdivided group 1 isolates into four main lineages: ST11, ST15, ST101 and ST258/512. The majority of each of these lineages carried carbapenemase genes, except for ST11. A high average nucleotide identity of 99.9–100% among isolates within same major lineages indicated a lack of genetic diversity and recent common ancestry. Nevertheless, these lineages were widely distributed across Europe. For 52.6% of carbapenemase-positive isolates, the genetically nearest neighbour was from the same hospital, indicating that transmission occurred mainly within hospitals. Moreover, analysis of differences in pairwise single nucleotide polymorphism between isolate sequences revealed that inter-hospital spread was more frequent within than between countries. Cluster analysis of WGS data identified 21 single-nucleotide polymorphisms that were optimal in terms of sensitivity and specificity for discriminating between isolates from different hospital clusters. Given the strong relation observed between core genome diversity and geographical distance, plasmid spread of carbapenemase genes was unlikely to play a major role in this sample.

Phylogenetic comparison with isolate sequences from outside Europe confirmed that ST258/512 emerged in the USA. Furthermore, it showed that all isolates from Greece were in the same clade and that many isolates from other European countries cluster among the Greek isolates. This suggests that there was a single introduction event of the ST258/512 lineage from the USA to Greece, from where it spread to other European countries.

The conclusion of this study of WGS surveillance of carbapenemase-resistant *K. pneumoniae* in Europe was that carbapenemase resistance in *K. pneumoniae* spreads mainly through nosocomial transmission. While other routes of transmission cannot be excluded, public health interventions should focus on reinforcing the resilience of national hospital referral networks with a policy of no-tolerance infection control. Genomic pathogen surveillance should be maintained to intercept high-risk clones and their expansion before a major outbreak occurs. This structured survey can serve as a reference for assessing how effective public health control measures are in reducing the threat of carbapenemase-resistant *K. pneumoniae*.

Examples of uses of WGS in local, national and international studies of AMR are listed in Box 4.

BOX 4. EXAMPLES OF USE OF WGS IN LOCAL, NATIONAL AND INTERNATIONAL STUDIES OF AMR

Local use:

- Detect known AMR mechanisms and virulence factors.
- With AST data, identify novel mechanisms of resistance, e.g. with long-read sequencing to identify resistance-conferring plasmids.
- Investigate and control an outbreak at a single centre, e.g. rapid recognition of AMR clusters in health care facilities.

National use:

- Compare several genomes from various sites.
- In combination with AST data, identify novel mechanisms of resistance, e.g. with long-read sequencing to identify resistance-conferring plasmids.
- Analyse regional transmission networks, and monitor evolution of AMR.
- Trace sources of large, multi-jurisdictional outbreaks.
- Inform control strategies.

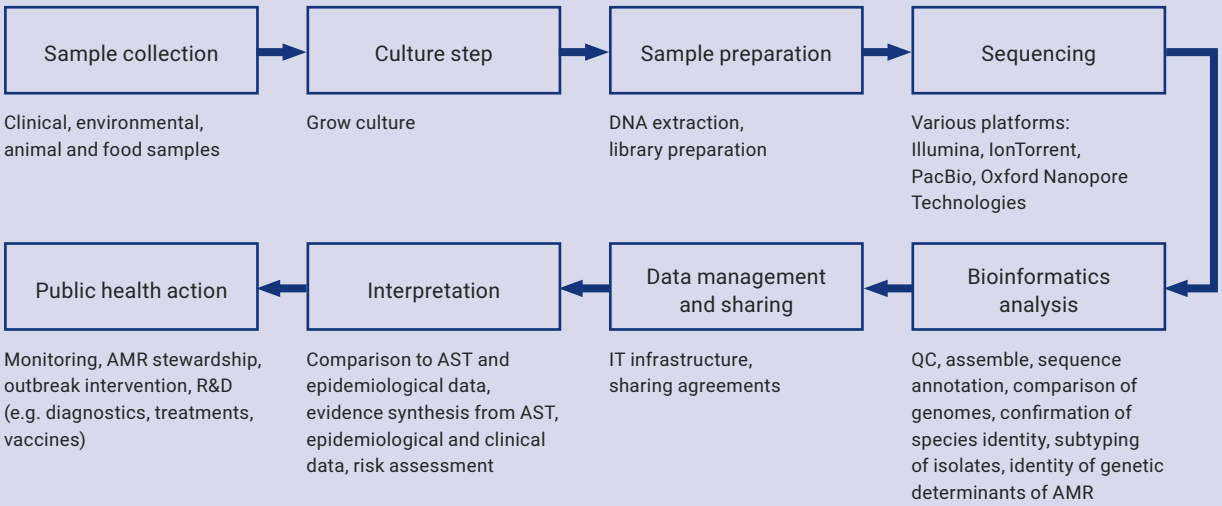
International use:

- Monitor pathogen populations and geographical and temporal genetic variation.
- Detect and track newly emerging high-risk clones, and trace their dissemination.
- Monitor the emergence or new resistance mechanisms to guide research and the development of new diagnostics and treatments.
- Monitor the spread of mobile genetic elements to new locations.
- Monitor the effects of global interventions.

04 Requirements for introducing whole-genome sequencing

This section outlines the possible workflow, infrastructure, platforms and bioinformatics required for laboratories to use WGS for AMR surveillance (Fig. 6). The choice of an appropriate sampling strategy and selection of relevant specimens are important before testing for AMR. All the necessary equipment and skills must be available for all steps in the workflow, including basic laboratory, sequencing and bioinformatics and data storage capacity, to generate and analyse WGS data, as each subsequent step depends on the output of the preceding steps, and actionable results are available only once the entire workflow has been completed. Having a sequencer is insufficient if the other elements are missing. The specific setup will depend on the intended public health use. All steps in the workflow require good QC and QA (see section 4.3 for details).

Fig. 6. Steps in WGS data generation and analysis



These steps are necessary for all analyses of WGS data. The type of bioinformatics analysis depends on the purpose of the analysis. Quality management is essential to guarantee reliable results (see section 4.3).

4.1 Timing of introduction

A decision to introduce WGS into an AMR surveillance system should be based on the defined objectives of surveillance and whether existing methods can meet those objectives (see section 2.2). At present, the basis for AMR surveillance is still phenotypic AST of clinical specimens, and many countries are still building and using their basic capacity for bacteriology and AST. Although these services can meet several of the objectives of AMR surveillance, WGS can add considerable value to AMR surveillance systems. Phenotypic and WGS surveillance systems can be built simultaneously, but only if sufficient resources and are available in the country. Generally, AMR surveillance based on phenotypic AST should be established first and then complemented with WGS. Independently of the microbiological methods used in AMR surveillance, countries should have the epidemiological capacity to design surveillance and to analyse and act on the results.

The key prerequisite for introducing WGS into a surveillance system is good microbiological and bioinformatics expertise in the public health system. In countries with no prior experience of molecular methods, WGS will usually first be implemented in the NRL. The NRL must therefore meet all the requirements for high-quality WGS analysis. Countries will also have an advantage if they have laboratories that can perform high-quality PCR and some bioinformatics capacity or computational infrastructure. WGS is a natural extension of PCR, involving many of the same methods (such as DNA extraction and storage and preparation of reagent mixes) and knowledge in genetics and molecular biology. If computational infrastructure remains to be developed, it may be preferable to rent cloud-based servers for data storage and processing to avoid the burden of high-capacity computer storage and maintenance. This, however, requires a reliable broadband Internet connection.

Countries may have heterogeneous AMR surveillance and laboratory and computational capacity. In such cases, WGS could be set up only in those regions in which it is feasible and introduced into other regions as the surveillance system develops. For example, WGS could be introduced in an NRL, with a network for specimen referral from local laboratories without WGS capacity. In countries that do not yet have the capacity for full phenotypic AMR surveillance coverage, a "hub-and-spoke" model could be used, whereby local sites send samples to the NRL (or the RRL in another country if there is no NRL with WGS capacity) for both phenotypic AST and molecular analysis, and the results are returned to the submitting laboratory. Such a model need not be restricted to AMR surveillance, and the same WGS hub could be for other surveillance programmes, such as for HIV, TB and influenza. Alternatively, WGS capacity in local or regional laboratories can be used to obtain molecular data from point prevalence surveys. A decentralized sequencing network is an option if a central sequencing laboratory becomes overloaded.

Staggered implementation of WGS surveillance in countries and regions should be coordinated to ensure comparable data, by setting international standards for validation and data-sharing early in the process. Standardized software for analysis and workflow management could be provided to members of a surveillance network. Countries that wish to strengthen both their phenotypic and their molecular surveillance capacity can save on overall costs by ensuring that new laboratories have the capacity for both phenotypic AST and molecular testing, including WGS, or by pooling resources with other WGS surveillance programmes. It is not necessary, however, that every laboratory in a surveillance system have both phenotypic and molecular testing capacity.

4.2 Infrastructure requirements for whole-genome sequencing

4.2.1 Laboratory

Laboratories planning to include WGS in their services should have adequate infrastructure for isolating organisms from clinical samples, culturing them and extracting DNA from cultured isolates. Nationally standardized routine sampling of bacterial isolates is a prerequisite for AMR surveillance, and capacity to culture microorganisms from samples and to isolate bacteria isolation is a key element, as the material for WGS of bacterial pathogens is usually purified DNA extracted from liquid culture (grown from a single colony) or from single colonies picked from primary culture plates *(48)*. WGS from direct clinical samples, while technically possible, is complex, with a high rate of failure, and currently requires additional time-consuming and costly steps.

Phenotypic AST data are essential for the detection of new AMR mechanisms (when a pathogen acquires resistance to a drug to which it was previously susceptible) by comparing the expression of phenotypic resistance with the presence of AMR genes or mutations. The infrastructure required for phenotypic testing is listed in the WHO GLASS guidance for NRLs *(49)*. The methods frequently used in phenotypic AST are culture in antimicrobial-containing liquid or solid media and gradient strip tests to determine MICs (broth microdilution, agar dilution), disc diffusion tests and automated methods. Ideally, WGS data are compared with those from broth microdilution AST, which is the "gold standard" for determining the MIC for fast-growing bacteria and is therefore expected to give the best correlation between the two methods. Continued understanding of the correlation between phenotypic and genotypic results will require guidelines for breakpoints and clearly defined clinical or epidemiological cut-off values.

Laboratories that are planning WGS workflows should have the capacity for high-quality PCR confirmed by internal and external QA. It is essential to ensure that the colony that is sequenced is the same as the one that is phenotyped. If the colony has been passaged, it could have acquired changes such as loss of plasmid or acquisition of new mutations, which reduces confidence in genotypic–phenotypic associations. If discrepancies are identified, the phenotypic and/or the molecular analysis will have to be repeated.

All of these issues must be considered in establishing WGS in a laboratory and steps put in place to ensure both the necessary workflow and internal and external QA standards to maintain quality. All the steps in the workflow must meet those standards to participate in global surveillance schemes and to upload data to databases. LMIC may be constrained in setting up a high-performing microbiology laboratory and upgrading it with WGS capacity by lack of resources, infrastructure, purchasing power and expertise. If all the requirements cannot be met, it might be preferable to send samples to a central sequencing laboratory (NRL or RRL) rather than upgrading local laboratories to full WGS capacity. Centralized sequencing can save significant costs if the challenges of sample transport can be overcome and the central laboratory has sufficient capacity to serve all submitting laboratories without a lengthy wait.

Current sequencing instruments require the following infrastructure:

- a reliable Internet connection;
- a continuous A/C electricity supply;
- vibration-free platforms;
- dust control for some equipment;
- molecular biology-quality water (which can be purchased in bottles or produced on site by treatment or filtration);
- temperature and humidity regulation within an adequate range for chemical reactions and temperature-sensitive equipment;
- cooling and storage for reagents and DNA, with periodic recording of storage temperatures;
- rooms that are air-tight and maintain a stable operating environment;
- enough space to set up a unidirectional workflow;
- a procurement system, including issuing tenders; and
- automated liquid handling systems or robots for high-throughput capacity.

Laboratories that do not meet these criteria might have to install generators, uninterrupted power supplies, air-conditioning and water purification systems in order to conduct WGS. Sufficient space for dedicated working surfaces and a unidirectional workflow are necessary to ensure the sterility of isolates and to avoid cross-contamination. There should be three physically separated areas for reagent handling, DNA template preparation and DNA amplification.

Not all laboratories in a surveillance network need have the capacity for both phenotypic and WGS surveillance. If the network is sufficiently large, some partners can perform only phenotyping or only sequencing to provide data on the geographical distribution of AMR, while those that conduct both AST and WGS can provide the database of AMR genes and mutations and genetic context linked to phenotypic AST data that are necessary to interpret an AMR profile.

4.2.2 Platforms

Various platforms are available for sequencing microbial genomes. The choice depends on the objectives of sequencing and the required depth of coverage, time to result, throughput and cost. Sequencing platforms can be classified as short-read and long-read, a “read” being an inferred part of a genome sequence that corresponds to a single DNA fragment. The most commonly used sequencing instruments for bacterial WGS are short-read platforms, which produce sequence fragments of fewer than 300 base pairs. The most frequently used machines for bacterial genome sequencing are Illumina platforms, which are based on an approach whereby DNA fragments are synthesized from fluorescently labelled nucleotides during sequencing. At each step, only one nucleotide is added to the growing DNA strand. The labels produce fluorescent light at different wavelengths and therefore indicate which nucleotide is added at each step. Because of the “synthesis” aspect of this platform, each nucleotide is sequenced hundreds or thousands of times (high “coverage”), which ensures high accuracy. Illumina offers two suites of instruments that are suitable for pathogen sequencing: HiSeq, NextSeq and NovaSeq instruments offer very high-throughput sequencing for large sequencing centres, whereas the iSeq, MiSeq and MiniSeq instruments are more suitable for regional or local and new laboratories; however, the cost per isolate sequenced on these smaller platforms may be significantly higher if batching cannot be optimized.

An alternative approach to short-read sequencing is semiconductor sequencing, as with Ion Torrent devices (ThermoFisher). These sequencers identify nucleotides in a DNA strand from pH changes during synthesis. Fewer bioinformatics tools are built for Ion Torrent-generated data than for Illumina-generated data.

Although short-read platforms are highly accurate, they may still leave gaps in sequenced genomes, so that it is difficult to assemble small fragments into complex genome regions, such as tandem repeats and GC-biased regions or regions that contain several copies of the same small mobile genetic element.

Long-read sequencing platforms that generate DNA sequences of more than 10 000 base pairs can overcome these shortcomings, as they sequence very large regions of the genome and are therefore most useful for constructing complete whole genomes and plasmids *(50)*. The main long-read platforms are currently Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) instruments. The PacBio platform is based on a variation of the sequencing-by-synthesis approach, called single-molecule real-time sequencing, which allows very short run times *(9)*. These sequencers are, however, unsuitable for new or small WGS laboratories owing to the high cost of instruments and reagents, their vulnerability to DNA fragmentation, their large size and a higher error rate.

ONT has developed a novel sequencing technology in which individual DNA or RNA molecules pass through engineered protein nanopores, and the electric current across each pore is measured and can be converted into DNA sequence information. Nanopore sequencers (Flongle, MinION, GridION, PromethION) generate long reads and can be analysed while sequencing is still in progress. The MinION and Flongle sequencers are small and portable and have been used at field sites, such as during the outbreak of Ebola virus disease in West Africa in 2013–2016. ONT sequencers require significantly more target genomic DNA than short-read sequencers and have a relatively high base call error rate *(50)*. Although the reads are long and can complement output from short-read platforms, the coverage is lower than that of short-read platforms, and, consequently, the overall error rate is higher. While this still poses a problem, as ONT devices mature, they may become reasonable technology for newly established sequencing laboratories, as they do not require a huge initial investment or service contract and can be used in laboratories that have intermittent power outages.

Nevertheless, analysis and assembly of long-read sequences require access to adequate high-performance computing environments. High error rates can be overcome by performing short-read sequencing in parallel, ideally from the same DNA extract used in ONT long-read sequencing. A recently developed method of hybrid assembly of both short and long reads has been extremely useful for resolving complex bacterial genomes and mega-plasmids with high accuracy. Consideration should be given to the extra cost of conducting both long-read and short-read sequencing and whether the question at hand requires the extra information provided.

A table listing the technical details of different sequencing instruments is provided in the WHO paper on WGS for foodborne disease surveillance *(12)*, and Annex 3 provides an updated version. Table 1 summarizes the suitability of different WGS sequencing platforms for different laboratory types. When choosing a sequencer, laboratories should weigh the initial investment cost against the cost of reagents and of support contracts, which differ by manufacturer.

Table 1. Suitable WGS platforms by laboratory type

LABORATORY TYPE	PURPOSE	SUITABLE PLATFORMS
New laboratory or local laboratory	Sequencing of a limited number of samples Fast detection of specific AMR mutations for outbreak management	Illumina MiniSeq Illumina iSeq
Subnational regional laboratory	Sequencing of a potentially large number of samples for outbreak detection and AMR typing	Illumina MiSeq Ion Torrent devices Illumina NextSeq (if covering large region) Illumina NovaSeq (if covering large region) GridION (ONT)
National reference laboratory	Sequencing of a large number of samples High-throughput sequencing	Illumina MiSeq Illumina NextSeq Illumina NovaSeq Ion Torrent devices PacBio devices GridION (ONT) PromethION (ONT)

4.2.3 Bioinformatics analysis software

WGS bioinformatics combines biology, computer science, mathematics and statistics and is used to analyse and interpret WGS data. Bioinformatics pipelines are a series of algorithms for each step in the workflow, including QC checks and downstream analysis, such as sequence alignment and sequence comparison. Bioinformatics is a rapidly evolving field, and there are currently no universally accepted "gold-standard" tools for bioinformatics analysis of AMR. Laboratories with WGS expertise prefer different algorithms and can differentially tailor them to their needs, while laboratories with no expertise will require support in deciding what analyses they need and the necessary hardware, software and information technology infrastructure. The optimal tool for a task depends on the setting, the sequencer used, the genetic characteristics of the sequenced organism and the purpose of the analysis; however, use of algorithms customized for specific situations can limit the functionality of a bioinformatics workflow, and the results of different laboratories will not be comparable. Building standard bioinformatics capacity in many countries simultaneously will be a unique opportunity to harmonize methods and QC and QA mechanisms to ensure the comparability of future results (see section 4.3). It will also provide an opportunity to share computational infrastructure or modes of access (e.g. via a commercial cloud provider), whereby many users can access the same software for their analyses. Harmonization does not require all laboratories to use exactly the same tools; they can use any tool that is validated for the intended use.

Each step in bacterial genome analysis requires specialized bioinformatics tools, of which there are many, with different strengths and weaknesses (see Fig. 6 for a generic bioinformatics pipeline). The tools must be integrated into a coherent workflow, and thresholds and filter settings should be transparent and explicit to minimize errors and biases and maximize comparability. Ideally, it should be possible to link the output of bioinformatics pipelines to phenotypic testing results and epidemiological and clinical metadata to create a knowledge base for surveillance, without compromising patients' protected health information.

Highly trained, experienced personnel are necessary to decide what tools to use and when. Equally, the interpretation of results is not straightforward, and they may be difficult to translate into action by non-specialists. Consequently, countries that wish to include WGS in their AMR surveillance system must build bioinformatics capacity, and standardized, end-to-end bioinformatics solutions should be made publicly available for use by non-experts. In the context of deployment of WGS for national surveillance of AMR pathogens by convergent bioinformatics approaches in Europe *(51)*, a user survey identified the main hurdles as lack of expertise in integrative analysis and interpretation of epidemiological and sequence data, lack of interoperability of surveillance systems and lack of access to user-friendly international nomenclature.

The genome sequences of pathogens can be compared with reference genomes to identify species and to type strains. Web-accessible or command-line in silico bacterial typing methods (e.g. single-nucleotide polymorphism typing, cg multi-locus sequence typing) are available for WGS for more finely grained subtyping and more sensitive AMR gene detection than microbiological or other molecular methods such as pulsed-field gel electrophoresis and multi-locus sequence typing. WGS data on pathogens for which the genome is well characterized and represented in publicly accessible databases can be used to predict important phenotypic features such as AMR and virulence. Although this is still not possible for all bacteria, the method has been used successfully in profiling HIV resistance, for directing antiretroviral therapy and for detecting clinically relevant AMR in Mycobacterium tuberculosis *(52, 53).* Application of next-generation sequencing technology for detecting mutations associated with AMR in M. tuberculosis was described in extensive technical guide published by WHO in 2018 *(54)*.

The available databases and online tools perform differently in detection of AMR determinants. A single, public database of all known AMR genes and mutations could be established by extending and improving existing databases. This database could be organized by bacterial species and include data on the AMR phenotypes of sequenced isolates. Smaller, targeted databases would also be acceptable for specific uses, including detection of AMR. Careful curation and updating of reference databases are important because bacteria evolve rapidly. In particular, different bacterial strains or species can acquire new AMR and virulence genes by lateral gene transfer. If databases are not well maintained and do not contain enough sufficiently diverse, quality-controlled sequences, use of these databases for genotypic prediction of AMR may give erroneous results. Some widely used databases are listed in Annex 3 (Table A3.4). Meta-platforms are available for different pathogens that contain data from separate genome databases to facilitate analyses (e.g. nextstrain, iDseq). The WHO paper on WGS for foodborne disease surveillance lists some bioinformatics tools for WGS analysis of bacterial pathogens *(12),* and a recently updated list of tools for identifying AMR genes by Hendriksen et al. *(27)* is reproduced in Annex 3 (Table A3.5).

Open-access, open-source and proprietary bioinformatics software tools are available, as are open-access, public-access and closed-access sequence databases. The definitions used in this document are as follows:

- Open-access software or tool: a tool that can be accessed by anybody.
- Open-source software or tool: a software code that is publicly available.
- Proprietary software or tool: a tool that is owned by a company or institution, with restricted access. Frequently, access can be bought for a fee.
- Open-access database: data for which the providers do not retain rights.
- Public-access database: a tool or database that may be used for free but only by those people who require access (e.g. public health officials and epidemiologists). Data providers seek information and control of the downloading and use of sequence data, most notably acknowledgement for collaboration if used in publications and/or public communications based on analyses of sequences they have provided, and assurances that products, such as diagnostics, therapeutics and preventive agents, developed with sequence data are accessible to the countries in which the disease burden is greatest and from which the sequences originated. Access may be approved upon registration.
- Closed-access database: a database that may be accessed only by individuals who have been granted access. The sequence data providers require that only non-publicly accessible databases be used, and members of a network may collaborate and share information, but sequences are not accessible to the general public. There is no open registration.

Advantages and disadvantages of open-access versus proprietary software

Table 2 indicates that proprietary software tends to have more user-friendly interfaces and requires less training. Some proprietary software also offers support packages, which may be useful for inexperienced users. These advantages, however, usually come at a cost, with loss of the ability to customize analyses. In contrast, open-source software is typically free to use and can frequently be customized for the user's needs, if the user has sufficient training in use of the software. As the code is accessible, the method is more transparent and it is easier to investigate the technical weaknesses of the method and why some analyses fail.

Table 2. Advantages and disadvantages of open-access versus proprietary analytical tools for WGS

	ADVANTAGES	DISADVANTAGES
Proprietary	• May have more user-friendly interface • May offer tailored user support • Usually does not require much training • High product stability and standardization	• Cost may be high • Methods not transparent • Limited control of the design, coding standards and flexibility for making modifications for new types of analysis • Time lag in obtaining support from providers based in other time zones • Prevents data-sharing
Open-access	• No cost to user • Community or developer may offer support • Transparent • May be customized by user • "Bugs" and security issues may be recognized and resolved faster because the code is open and more people look at it.	• User interface may not be intuitive • Less tailored user support • Typically requires more training in bioinformatics and programming • Product may be unstable, with changing versions on open-source platforms • Complex certification for data security and laboratory accreditation

Access models for bioinformatics analysis software

WGS data and, especially, the associated metadata may include sensitive information; therefore, open-access databases may not be the most suitable. Closed-access databases, as used, for example, in the WHO HIV Drug Resistance Surveillance Programme and the WHO Tuberculosis Surveillance Programme, guarantee that sensitive information is shared only with trusted partners. Because access to data is essential to improving public health, efforts are needed to develop alternative public-access options to allow proprietary databases towards public-access databases while preserving their public health purposes and confidentiality requirements

A good example of a public-access database is that of the Global Initiative on Sharing All Influenza Data. Access to the EpiFlu database is free of charge to all individuals who identify themselves and agree to uphold the Initiative's data-sharing agreement. The terms of use include agreement that data are not to be copied from the database or distributed to third parties, a requirement to acknowledge the source of data in publications (originating laboratory and submitting laboratory) and agreement to collaborate with representatives of the originating laboratory as appropriate *(55)*. The choice of databases and software should take into consideration privacy, intellectual property and publication rights. Data-sharing models are further discussed in section 4.6.

To maximize their usefulness for public health, databases of pathogen genome sequences, including an agreed level of metadata, should preferably be publicly accessible, with clear rules about who is granted access and how data and results generated from the data can be used and publicized. Public-access databases must nevertheless be maintained and curated, entailing costs. While public-access databases and open-source tools do not entail additional costs to users in LMIC, the computational infrastructure and expert personnel required to run analyses will incur costs for the global community. Lack of computational infrastructure is often cited as a barrier to use of WGS for AMR surveillance in LMIC *(56).* The cost of setting up computing infrastructure for WGS analysis is typically US$ tens of thousands. In addition, there are maintenance costs for continuous power, cooling and replacement of components. A laboratory that is acquiring bioinformatics infrastructure de novo requires robust storage. Questions to be considered before setting up storage facilities are: should storage be delocalized? Can redundant storage be implemented? How frequently will data be backed up? What data will be stored – raw data or only final results? Especially for reference laboratories, secondary, off-site storage is necessary in case something happens to the primary site. Skills development for systems administration and engineering for bioinformatics computational systems are also required.

There are, however, cloud-based alternatives to local computing infrastructure, onto which data can be uploaded to computing resources hosted elsewhere. This may be a useful option, especially for small local and newly established laboratories, because a lack of bioinformatics capacity is frequently a barrier to the use of WGS in public health, even if laboratory capacity and sequencing capacity are available. Surveillance models in which bioinformatics capacity is centralized in an NRL or RRL and in which local centres send samples or generated sequence data to the central hub are a potential solution. For web-based analyses and for uploading data to a server in a reference centre, a reliable broadband Internet connection is nevertheless required. Initiation of WGS in AMR surveillance can be linked to other pathogen surveillance activities in order to share bioinformatics capacity and find joint solutions to common infrastructure barriers. If everything is ready for introducing WGS in a country but the country's bioinformatics capacity is still underdeveloped, this service can be outsourced temporarily to a company or institution.

4.3 Quality assurance, quality control and international standardization

There are no internationally recognized standards for QC of WGS data for pathogen typing and AMR gene detection. Some national and pathogen-specific surveillance networks (e.g. PulseNet, Gen-FS), public institutions (e.g. US Centers for Disease Control and Prevention, US Food and Drug Administration) and the Global Microbial Identifier) have agreed standards that can be extended or adapted *(57–59)*. A QC programme for WGS-based AMR surveillance could be part of a general accreditation programme for bacteriology and for future AMR surveillance. The most frequently used metrics for assessing the quality of WGS data were summarized by Ellington et al. *(17)*. If WGS is to be included in an international AMR surveillance system, common QC metrics must be agreed upon to ensure that results are interpretable and comparable. The minimum metrics required may depend partly on the sequencing technology used and the organism sequenced *(17)*. Once a bacterial genome has been sequenced and before any downstream in silico analysis, the sequenced fragments must be cleaned and run through QC. Certain analyses may require a genome assembly, the quality of which depends on the assembler algorithm, the features of the genome and the quality of the raw data, although it can be difficult to distinguish assembly errors from biologically relevant genetic changes *(60)*. An evidence-based decision must be made on the minimum gene coverage (the number of unique reads that include a given nucleotide in the reconstructed sequence) necessary for detecting genes and variants, especially if AMR phenotypes are to be predicted from sequence data. If this is not possible, a system of "present", "absent" or "indeterminate", based on statistical methods, should be developed for WGS surveillance that mirrors the phenotypic susceptible/increased exposure (European Committee on Antimicrobial Susceptibility Testing) or intermediate (Clinical and Laboratory Standards Institute)/ resistant systems. Such standardization has been done for TB and foodborne infections *(61, 62)*. (Note that genes and mutations are always either absent or present; however, the genomic sequence may not indicate the presence or absence of phenotypic expression of AMR. Consequently, "indeterminate" in WGS analysis is not the same as "intermediate" in phenotypic analysis, and the organism could be either susceptible or resistant.)

Laboratories should participate not only in internal QA of the WGS procedure but also in external QA of their WGS activities. External QA for microbiology laboratories is outlined in the WHO GLASS guidance for NRLs *(49)*. External QA for WGS laboratories should be based on well-characterized, externally confirmed isolates with known phenotypic characteristics and mutations of particular interest. This reference panel of isolates should be sent to laboratories to be sequenced and analysed. The results of phenotypic AST and WGS-inferred AMR should be concordant. External QA tests should be flexible enough for different laboratory protocols but stringent enough to ensure that all laboratories correctly detect AMR from their WGS analysis. In-silico proficiency testing of distributed raw sequence reads can be used as a complementary approach to testing multi-centre reproducibility of bioinformatics assembly and analysis pipelines with classical in-vitro external QA based on distribution of coded isolates or biological samples to measure inter-laboratory reproducibility of both "wet lab" and "dry lab" procedures for WGS data production and analysis.

Several attempts have been made to create global external QA for WGS; however, their development has suffered mainly from lack of resources *(55)*. The European Commission and the Fleming Fund have mandated the European Union Reference Laboratory on Antimicrobial Resistance to develop external QA by 2020. Initiatives such as the Global Microbial Identifier offer ring trials that are extremely useful for testing the capacity of WGS laboratories *(57)*.

With the use of validation strains and their own protocols, laboratories should be able to identify a bacterial strain and determine the presence or absence of AMR genes and mutations with a specified accuracy. This should be judged for key pathogens and key AMR genes. For example, systems that cannot detect the specified threshold value of New Delhi metallo-β-lactamase producers or MRSA after analysis of good-quality sequence data may not be acceptable. A document validating such capacity should be the first step in providing WGS data to a biobank. A "gold-standard" panel of pathogen strains and AMR genes and mutations could be provided by a databank such as the American Type Culture Collection *(63)*. The US Centers for Disease Control and Prevention and the US Food and Drug Administration have a freely available biobank of resistant isolates, all of which have been sequenced. The Global Microbial Identifier has established a proficiency testing framework for laboratories that conduct WGS-based AMR surveillance and tested it with foodborne pathogens *(57, 64)*. These databases and protocols could form the basis of an internationally agreed external QA standard. Standards must be strict enough to ensure that surveillance data are of good quality but should not be so restrictive that newly established laboratories, many of which will be in LMIC, cannot participate. For some pathogens, such as *N. gonorrhoeae*, WHO has developed international reference strains, including reference genomes, for global QA of AMR testing and genomic investigations *(65)*.

To ensure that only high-quality data are uploaded to databases, every sequence or result derived from sequence data must meet minimum quality standards that are relatively easy to check before uploading. Only WGS data that are linked with a minimum standard of metadata should be used for analysis. These standards should be defined and agreed upon by all countries participating in a global surveillance network. Once accurate inference of AMR is consistently achieved, laboratories should be able to determine whether isolates are related on the basis of sequence similarity.

4.4 Procurement

An important aspect of the affordability and cost-effectiveness of WGS for AMR surveillance is capital procurement of equipment. The cost of a sequencing instrument in HIC ranges from about US$ 20 000 for an Illumina iSeq, US$ 50 000 for an Illumina MiniSeq or an Ion PMG to US$ 695 000 for a PacBio RSII or US$ 1 000 000 for an Illumina HiSeq X machine (2018 prices) *(12)*. ONT MinION devices are significantly cheaper, at US$ 1000 *(12)*; however, they can currently be used only in combination with short-read sequencing data. The cost per gigabyte (Gb) DNA sequenced falls as the throughput of a sequencing instrument rises. Thus, the cost per Gb is about US$ 7–10 for the Illumina HiSeq X and US$ 200–400 for the Illumina MiniSeq *(12)*. The cost for sequencing 1 Gb DNA with an ONT MinION device is about US$ 100–400 *(64)*. In relation to throughput, the cost per Gb sequenced is slightly lower for Illumina than for Ion Torrent or PacBio devices. The significant recurrent costs for reagents, maintenance and service contracts should also be considered.

Laboratories should choose a sequencing instrument that meets their current and expected requirements for throughput, which has implications for batching (sequencing many isolates at the same time) and, in turn, affects cost and turn-around time. Another important consideration is whether short-read and/or long-read sequencing will be required, because none of the current technologies can deliver both. The choice of platform is also limited by the available bioinformatics. Moreover, it is worth considering whether it would be more suitable for the setting to establish an NRL or an RRL with high-throughput capacity for centralized WGS analysis rather than establishing several smaller laboratories with lower throughput. New shipping protocols and best-practice guidelines might have to be developed, depending on the type and number of samples to be shipped *(49)*. The protocols could be drawn up by local laboratories or laboratory networks and published for use by other laboratories. Purchasing and support consortia could help establish and support local suppliers for reagents and services in LMIC, which would be an important step towards sustainable laboratory supply chains and infrastructure.

The costs of device and reagents and of maintenance and support contracts may be higher in LMIC than in HIC because of supplier-based pricing models in which product prices are weighed against demand. Annex 4 gives a breakdown of the costs for initial purchase and installation of an Illumina MiSeq instrument and recurrent costs for sequencing in four LMIC in four WHO regions and a comparison with the same costs in the United Kingdom (Table A4.1). Such bias could be removed if several laboratories or countries formed a purchasing consortium to negotiate lower prices with suppliers. Procurement costs can also be reduced by purchasing through WHO regional offices and, if possible, by using WHO lists of approved suppliers. Shipping and delivery of reagents or samples can be problematic in LMIC if the reagents have a short expiry period, such as for ONT devices, if deliveries are delayed, if they are held in customs or if the delivery systems cannot handle dry ice.

Additional costs of equipment and infrastructure can double the cost of setting up a WGS laboratory over that of the sequencer *(66)*. Some laboratories may already have the necessary infrastructure but might have to reorganize or enlarge the workspace. Extension of a laboratory building to accommodate a medium- to high-throughput sequencer will require considerable extra expenditure, as will instruments for automated DNA extraction, robots for high-throughput library preparation, shearing instruments and other equipment. For example, Public Health Wales has three MiSeq instruments that together cost £ 200 000–300 000 (US$ 260 000–390 500), robotics for library preparation for an additional £ 160 000 (US$ 208 000) and computational hardware for data storage and processing that cost £ 150 000 (US$ 195 000) (12). These costs are comparable to those for building a new laboratory with capacity for high-throughput phenotypic AST; however, WGS can replace molecular methods still in use in clinical and reference laboratories, such as pulsed-field gel electrophoresis and multi-locus sequence typing, thus saving overall costs.

Downstream bioinformatics analysis also includes costs for skilled personnel and computing resources. Servers must be purchased, installed and serviced. If an adequate server cannot be bought in the country, it might have to be shipped from abroad, incurring additional costs for transport and customs. Alternatively, cloud-based servers can be rented; however, they require reliable broadband Internet access. Table A4.2 in Annex 4 lists the cost of an Internet connection relative to the cost of living in four LMIC, the United Kingdom and the USA. Although the relative cost of an Internet connection is lower in India than in the USA, it is 1.7–4.5 times higher in other LMIC *(44)*.

4.5 Training

Expertise in WGS technology can be obtained in many ways, including online courses, university courses (such as a specialized master's degree), courses offered by sequencing centres and joint research initiatives between HIC and LMIC for skills transfer and learning. These formats may provide tuition in one specialized area, e.g. laboratory techniques or bioinformatics analyses; however, the most useful training programmes provide an integrated approach that covers all the disciplines relevant to WGS for AMR surveillance and train teams rather than individuals. Training programmes should include hands-on laboratory work, bioinformatics skills, QA, biosafety, procurement and use of laboratory information management systems. Optimally, training is provided at the site in which WGS is to be introduced, as transferring newly acquired skills from one laboratory to another is not straightforward. Schemes whereby newly trained WGS scientists can ask questions and seek advice from experienced mentors can consolidate and advance skills. Expertise in procurement and supply chains is also valuable. As WGS programmes often rely on large grants, training in responsible grant management may be necessary.

Most training initiatives are based in HICs, far from trainees' home bases, and forms such as degree programmes may be expensive. It may be complicated and expensive to obtain a visa. Local or regional training programmes and expertise should therefore be made available, as has been done in the WHO Region of the Americas; and India and South Africa have significant expertise in WGS that could be shared regionally. Such initiatives could start with "train the trainer" programmes to train experts in teaching skills. Training courses could be set up in WHO RRLs to train representatives from NRLs, and regular webinars could be organized for continuing education.

4.6 Data collection, sharing and storage

WGS data can be collected for AMR surveillance continuously (by constant submission of newly sequenced isolates to a database) or in structured surveys (in which randomly selected isolates that have been sequenced during a specified period are submitted to a database). The latter option is usually easier to implement. Data collected during surveys provide the necessary context for interpreting isolates with novel AMR mechanisms and isolates sequenced during an outbreak investigation. The selection of isolates for background surveillance should be unbiased. In contrast, surveillance data collected during an outbreak are necessarily biased as they include isolates suspected of being part of the outbreak. The types of data collected should be defined according to the objectives of surveillance or the outbreak investigation.

For WGS to be useful in national and international AMR surveillance, the data generated by individual laboratories must be shared. Standardized databases are available for the storage of genomic and AMR data, but due consideration must be given to the type of database to be used (open, public or closed access, see section 4.2.3) according to the purposes and requirements of the surveillance system. Prediction of phenotypic AMR from sequence data is based on comparisons with the sequences of known AMR phenotypes. Identification of high-risk clones is based on phylogenetic analyses, and identification of new high-risk lineages is based on the genetics of the local or regional pathogen population. Databases are therefore required to compare sequences.

Although the benefits of data-sharing in infectious disease surveillance are generally acknowledged, concern about the economic, political or legal implications of results derived from shared data for countries, organizations or individuals, data ownership and privacy may elicit reluctance to share data *(67)*. Alternative platforms for sharing data on pathogen sequences should be available to meet needs and preferences in different situations. Data providers may seek certain control of or reassurance about the downloading and use of sequence data and particularly acknowledgement of collaboration when results from shared data are used in publications or public communications. Use of clinical data for diagnostics development may require additional approval from institutional review boards and patients who provide their data (see section 5). There should also be assurance that products, such as diagnostics, therapeutics and preventive agents that are developed with sequence data are accessible in the countries in which the disease burden is focused and from which the sequences originated. Public databases must address these concerns.

Sharing of sequence data and the associated metadata also poses regulatory, logistic and technological issues. Agreements on what data to share and with whom must be negotiated, and local or regional requirements on data protection must be considered. An example of regional regulation for data protection is the European Union General Data Protection Regulation *(68)*.

AMR surveillance systems should therefore define their data-sharing mechanisms and also define what data and with whom data should be shared to meet public health needs. The first question is which data should be shared to meet the agreed objectives of AMR surveillance and to answer epidemiological questions: complete sequence data or files on variants (i.e. only those positions in a genome that differ among isolates), data on individual isolates or nationally aggregated data? The type of metadata to be shared will depend partly on the pathogen, but some common metadata fields could include the time of isolation of the bacterial strain, place of origin of the isolate, place of sequencing and sample type. As epidemiological metadata are particularly sensitive, access will probably be restricted. WGS data on isolates, without metadata, can be made publicly available, while detailed laboratory results might not be shared openly.

The second question is who should have access to the data: public health authorities responsible for the surveillance of AMR, members of AMR surveillance networks, collaborators or researchers on request? Or should all data be publicly accessible? Another central question is where the data will be stored and where the servers with the data will be located: locally or in a central repository? The answers to questions on most appropriate ways of sharing sequence data should be addressed by AMR surveillance systems and networks at all levels.

4.6.1 Metadata

Metadata consist of a synthesis of genomic data with clinical, laboratory and epidemiological data. Data on isolate sequences must be combined with phenotypic, epidemiological and clinical metadata for actionable results. Hence, to maximize the use of WGS data, countries should have enough epidemiologically trained staff who work in the field to collect such data. The type of metadata required will depend on the infective agent and the intended use of the data and should conform to minimum requirements for the collection of phenotypic AST data in GLASS. The Global Microbial Identifier consortium has defined a minimum set of contextual data, the Minimal Data for Matching, that is now used by the European Nucleotide Archive and the US National Center for Biotechnology Information *(62)*. At a minimum, metadata for surveillance should include anonymized information on who was infected (age, gender, site of isolation, but no patient-identifiable information), where (geographical location) and when (year). Additional useful information includes the source of the isolate (e.g. anatomical site, tissue, stool), indication for testing

(i.e. clinical infection versus surveillance swabs) and AST data. For foodborne or zoonotic infections, the foods or animals that were the source of infection should be included, if known. For sexually transmitted infections, the sexual orientation of patients is valuable information, as different strains may circulate in different networks of sexual partners. For environmental samples, a description of the physical material constituting the sample and the location of collection (longitude, latitude) could be recorded.

Frequently, WGS data are generated more quickly and shared more rapidly than the associated metadata because of the high potential automation and standardization of WGS and the absence or underutilization of laboratory information management systems in many LMIC; however, WGS data are far less useful for actionable insights in the absence of metadata. When sequences are uploaded onto an online database, they are assigned an accession number, which, when rapid action is required, such as during an outbreak of a highly transmissible pathogen, allow rapid linkage of metadata to WGS data through defined, standardized formats and input fields. Epidemiologists, clinicians and public health workers should be made aware of the metadata to be collected for AMR surveillance and how to share them. A common concern in providing and analysing metadata is privacy, and they should be provided in such a way that individuals or, sometimes, institutions are nearly impossible to identify, such as with unique subject identifiers. A local laboratory should keep patient data, which may be shared with RRLs or the NRL. NRLs store data on the laboratories in their networks but may aggregate data at national level before sharing them internationally.

MICs or, when disc diffusion methods are used, the diameters of zones of growth inhibition should be collected, with data on internal QC. Clinical breakpoints and epidemiological cut-off values may change over time, and it can therefore be difficult to interpret past data if results are reported only as the discrete categories susceptible/increased exposure (European Committee on Antimicrobial Susceptibility Testing) or intermediate (Clinical and Laboratory Standards Institute)/resistant. Moreover, epidemiologists working on AMR should declare which standardized method, guideline, ingredients and components used and submit their internal QC data. The standards of either the European Committee on Antimicrobial Susceptibility Testing or the Clinical and Laboratory Standards Institute may be used if raw data are stored.

A laboratory information management system is key to standardization of metadata collection and integration of metadata with WGS data. Standardization of metadata will be most effective if entry forms contain fields that must be completed before uploading. In clinical and microbiological research, information management systems are routinely used to record data on patients and isolates (See WHO GLASS guidance for NRLs *(49)*); however, the type of system used varies within and between countries and settings, ranging from written notes to sophisticated software packages. For WGS in AMR surveillance, the system should be digital (at least in NRLs), based on a documented data standard and provide a variable and an interface to link relevant metadata to WGS data. Laboratory information management systems used to generate surveillance data should include a data life-cycle management system. The interoperability of the information management systems used by different laboratories or for different workflows must be ensured to achieve standardization.

Synthesis of metadata with genomic phylogenetic analyses offers important insights for AMR surveillance, including elucidation of the mechanisms of cross-species transmission, potential modes of pathogen transmission and who in a population contributes most to transmission. WGS data can be sufficiently discriminatory to target cases linked to community or hospital contacts and hence prevent further spread and to investigate genetically related cases with no clear epidemiological link. Continuous genomic surveillance allows identification of determinants of transmission, monitoring of pathogen evolution and adaptation, accurate, timely diagnosis of infections with epidemic potential and refining of strategies for their control *(69).*

4.6.2 National and international collation of data for the Global Antimicrobial Resistance and Use Surveillance System

WHO GLASS collects, manages and analyses AST data from microbiology laboratories. The countries enrolled in GLASS can use any software for data entry; however, WHONET software *(70)* is free of charge and has been adapted for the purposes of GLASS for countries that have no specified software for AMR surveillance. WHONET can convert laboratory codes and formats into its own structure and could be adapted to include WGS results. WHONET is, however, currently based on simple text format, whereas a WGS database is preferably based on a more flexible format that allows complex queries, such as the SQL format.

Other WHO surveillance programmes in which WGS or molecular data are used in their reporting systems include the HIV Drug Resistance *(52)* and the Multidrug-resistant Tuberculosis *(71)* programmes. In the HIV Drug Resistance programme, a free, publicly available genotypic AMR database that can be accessed via html or an automated web service has been integrated into the workflow of the network *(72)*. The database includes an interpretation system for predicting the AMR phenotype of a given HIV strain from its genome sequence, which can be used by clinicians to decide on treatment *(72)*. In the MDR-TB programme, molecular assays (line probe assays and nucleic acid amplification tests) and targeted next-generation sequencing directly from clinical specimens are used to accelerate diagnosis and for surveillance *(63, 71, 73, 74)*.

With WGS, GLASS can use existing microbial sequence databases in AMR surveillance. To facilitate such use, a unified nomenclature should be developed for data types and data fields, with a single interface for accessing all the databases. The functionality of the interface should be flexible, as different countries are at different stages of implementation of AMR surveillance *(75)*. The data submitted to databases should meet internal QC parameters, and the laboratory should demonstrate adequate performance in external QA.

05

Use of whole-genome sequence data for developing in-vitro diagnostics and new treatments and vaccines for infections caused by pathogens resistant to antimicrobial agents

Molecular in-vitro diagnostic tests for AMR are used to detect the acquired genes or mutations in the genomes of bacterial pathogens that make them resistant to one or more antimicrobials. Some AMR diagnostic tests are based on detection of the proteins encoded by acquired AMR genes, for example, lateral flow immunochromatographic tests to detect clinically relevant carbapenemases (OXA-48-like, *K. pneumoniae* carbapenemase, New Delhi metallo-β-lactamases, imipenemase, Verona integron-encoded metallo-β-lactamase, MCR-1, ESBL and PBP2a). Population-level data on pathogen genomes and genes and mutations that mediate AMR can be used in both the development of novel molecular diagnostics and the interpretation of existing tests for detecting antimicrobial-resistant pathogens. As AMR in bacterial pathogens evolves rapidly, continued molecular and phenotypic AMR surveillance will be necessary to maintain the sensitivity and specificity of molecular diagnostic tests *(76).* Targeted molecular diagnostics are cheaper than WGS and may also be suitable for use in settings where full implementation of WGS is not yet possible and where sequencing from direct clinical samples is critical for turn-around time. Moreover, WGS surveillance data that are representative for a geographical region can be used to improve region-specific interpretation of molecular diagnostic tests. Novel molecular diagnostics should be validated against well-characterized strains with relevant AMR genes and mutations. Several molecular diagnostic tests for AMR have been developed and approved for clinical use. Details of different molecular diagnostics and the requirements for their use in AMR surveillance are described elsewhere *(8).*

WGS surveillance data can also be used in the development of vaccines and novel antimicrobial treatments. For example, knowledge of the molecular mechanisms behind an AMR phenotype can be used in selecting candidate drugs that evade those AMR mechanisms and reduce the risk of cross-resistance. A biobank of resistant strains and their genome data could be used during preclinical testing of novel antimicrobials to ensure that they are active not only against clinically relevant strains on one continent but against strains with diverse genetic backgrounds *(77).*

06 Conclusions

WGS can greatly improve local, regional and global AMR surveillance. The barriers to translating WGS AMR surveillance into public health actions and to introducing WGS into LMIC include high initial and recurrent investment, the cost of laboratory and computing infrastructure and expertise, high maintenance costs and training requirements; however, there are solutions to most of these problems, so that countries can participate in WGS AMR surveillance networks. For example, regional purchasing consortia can negotiate lower prices for sequencing instruments and kits, and all bioinformatics analysis can be run on cloud-based servers so that local laboratories do not require expensive on-site computing infrastructure.

The costs of WGS are expected to fall further as it becomes more widely used, not only for pathogen sequencing but also for sequencing human genomes for medical applications. AMR surveillance systems that cannot meet the requirements for setting up a WGS laboratory can participate in "hub-and-spoke" models, in which pathogen isolates from local centres are sent to a reference laboratory for sequencing.

Application of WGS for global AMR surveillance will require standardization of WGS methods and validation to guarantee high-quality, comparable data from different laboratories. The fact that many countries are planning to or introducing WGS into their AMR surveillance system offers an opportunity for a high level of standardization. An initial step could be WGS of a few selected GLASS priority pathogens, building on national or regional experience. This would demonstrate the added value of WGS for AMR surveillance, and the pathogens and AMR mechanisms to be surveyed should be chosen with this purpose in mind. Such initiatives could have a worldwide impact, as many AMR pathogens in LMIC harbour new AMR mechanisms.

07 References

1. Sustainable Development Goals. New York City (NY): United Nations; 2015 (https://www.un.org/sustainabledevelopment/sustainable-development-goals/, accessed April 2020).
2. Jonas O, Irwin A, Berthe F, Le Gall F, Marquez P. Drug-resistant infections : a threat to our economic future. Washington DC: World Bank; 2017.
3. Jasovský D, Littmann J, Zorzet A, Cars O. Antimicrobial resistance-a threat to the world's sustainable development. Uppsala J Med Sci. 2016;121(3):159–64.
4. Gerner-Smidt P, Besser J, Concepción-Acevedo J, Folster JP, Huffman J, Joseph LA, et al. Whole genome sequencing: bridging one-health surveillance of foodborne diseases. Front Public Health. 2019;7:172.
5. One Health Global. One Health Global Network webportal 2015 (http://www.onehealthglobal.net/what-is-one-health/, accessed April 2020).
6. Global antimicrobial resistance surveillance system: manual for early implementation. Geneva: World Health Organization; 2015.
7. Emerging antimicrobial resistance reporting framework. Geneva: World Health Organization; 2018.
8. Molecular methods for antimicrobial resistance (AMR) diagnostics to enhance the Global Antimicrobial Resistance Surveillance System. Geneva: World Health Organization; 2019.
9. Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van Schaik W, et al. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev. 2017;30(4):1015–63.
10. Hendriksen RS, Lukjancenko O, Munk P, Hjelmso MH, Verani JR, Ng'eno E, et al. Pathogen surveillance in the informal settlement, Kibera, Kenya, using a metagenomics approach. PLoS One. 2019;14(10):e0222531.
11. Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, Lukjancenko O, et al. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nature Commun. 2019;10(1):1124.
12. Whole genome sequencing for foodborne disease surveillance: landscape paper. Geneva: World Health Organization; 2018.
13. Khan ZA, Siddiqui MF, Park S. Current and emerging methods of antibiotic susceptibility testing. Diagnostics (Basel). 2019;9(2):49.
14. Genomics for infectious disease management. Cambridge: PHG Foundation; 2016 (http://www.phgfoundation.org/research/genomics-for-infectious-disease-management, accessed April 2020).
15. Eyre DW, Shaw R, Adams H, Cooper T, Crook DW, Griffin RM, et al. WGS to determine the extent of *Clostridioides difficile* transmission in a high incidence setting in North Wales in 2015. J Antimicrob Chemother. 2019;74(4):1092–100.
16. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, 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):e00483-19.

17. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect. 2017;23(1):2–22.

18. Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, Pelegrin AC, et al. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. bioRxiv. 2019:793885.

19. Dominguez JE, Faccone D, Tijet N, Gomez S, Corso A, Fernández-Miyakawa ME, et al. Characterization of *Escherichia coli* carrying mcr-1 plasmids recovered from food animals from Argentina. Front Cell Infect Microbiol. 2019;9:41.

20. Hawkey J, Le Hello S, Doublet B, Granier SA, Hendriksen RS, Fricke WF, et al. Global phylogenomics of multidrug-resistant Salmonella enterica serotype Kentucky ST198. Microb Genomics. 2019;5(7):000269.

21. Fitzpatrick MA, Ozer EA, Hauser AR. Utility of whole-genome sequencing in characterizing *Acinetobacter* epidemiology and analyzing hospital outbreaks. J Clin Microbiol. 2016;54(3):593–612.

22. Balloux F, Brønstad Brynildsrud O, van Dorp L, Shaw LP, Chen H, Harris KA, et al. From theory to practice: translating whole-genome sequencing (WGS) into the clinic. Trends Microbiol. 2018;26(12):1035–48.

23. Dymond A, Davies H, Mealing S, Pollit V, Coll F, Brown NM, et al. Genomic surveillance of methicillin-resistant *Staphylococcus aureus:* a mathematical early modelling study of cost effectiveness. Clin Infect Dis. 2020;70(8):1613–9.

24. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus:* a descriptive study. Lancet Infect Dis. 2013;13(2):130–6.

25. van Dorp L, Wang Q, Shaw LP, Acman M, Brynildsrud OB, Eldholm V, et al. Rapid phenotypic evolution in multidrug-resistant *Klebsiella pneumoniae* hospital outbreak strains. Microb Genomics. 2019;5(4):10.1099/mgen.0.000263.

26. Gorrie CL, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, et al. Antimicrobial-resistant *Klebsiella pneumoniae* carriage and infection in specialized geriatric care wards linked to acquisition in the referring hospital. Clin Infect Dis. 2018;67(2):161–70.

27. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using genomics to track global antimicrobial resistance. Front Public Health. 2019;7:242.

28. Nakano S, Fujisawa T, Ito Y, Chang B, Matsumura Y, Yamamoto M, et al. Spread of meropenem-resistant *Streptococcus pneumoniae* serotype 15A-ST63 clone in Japan, 2012–2014. Emerg Infect Dis. 2018;24(2):275–83.

29. PulseNet. Atlanta (GA): Centers for Disease Control and Prevention; 2019 (https://www.cdc.gov/pulsenet/index.html, accessed April 2020).

30. Kwong JC, Lane CR, Romanes F, Goncalves da Silva A, Easton M, Cronin K, et al. Translating genomics into practice for real-time surveillance and response to carbapenemase-producing Enterobacteriaceae: evidence from a complex multi-institutional KPC outbreak. PeerJ. 2018;6:e4210.

31. Carbapenemase-producing (OXA-48) *Klebsiella pneumoniae* ST392 in travellers previously hospitalised in Gran Canaria, Spain. Stockholm: European Centre for Disease Prevention and Control; 2018.

32. Argimón S, Masim MAL, Gayeta JM, Lagrada ML, Macaranas PKV, Cohen V, et al. See and sequence: integrating whole-genome sequencing within the National Antimicrobial Resistance Surveillance Program in the Philippines. bioRxiv. 2019:808378.

33. Pathogen Surveillance Net. Hinxton (Cambs): Centre for Genomic Pathogen Surveillance; 2019 (https://www.pathogensurveillance.net/, accessed April 2020).

34. Argimon S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genomics. 2016;2(11):e000093.

35. Microreact. Hinxton (Cambs): Centre for Genomic Pathogen Surveillance; 2018 (https://microreact.org/showcase, accessed April 2020).

36. Pathogenwatch. Hinxton (Cambs): Centre for Genomic Pathogen Surveillance; 2018 (https://pathogen.watch/, accessed April 2020).

37. Baker KS, Campos J, Pichel M, Della Gaspera A, Duarte-Martínez F, Campos-Chacón E, et al. Whole genome sequencing of *Shigella sonnei* through PulseNet Latin America and Caribbean: advancing global surveillance of foodborne illnesses. Clin Microbiol Infect. 2017;23(11:845–53.

38. Nadon C, Van Walle, Campos J, Chinen I, Concepcion-Acevedo J, et al. PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. Eur Surveill. 2017;22(23):pii 30544.

39. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, et al. Distinct *Salmonella enteritidis* lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat Genet. 2016;48(10):1211–7.

40. Wong VK, Baker S, Pickard DJ, Parkhill J, Page AJ, Feasey NA, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella Typhi* identifies inter- and intracontinental transmission events. Nat Genet. 2015;47(6):632–9.

41. Franz E, Rotariu O, Lopes BS, MacRae M, Bono JL, Laing C, et al. Phylogeographic analysis reveals multiple international transmission events have driven the global emergence of *Escherichia coli* O157:H7. Clin Infect Dis. 2019;69(3):428–37.

42. Rapoport M, Faccone D, Pasteran F, Ceriana P, Albornoz E, Petroni A, et al. First description of mcr-1-mediated colistin resistance in human infections caused by *Escherichia coli* in Latin America. Antimicrob Agents Chemother. 2016;60(7):4412–3.

43. Tijet N, Faccone D, Rapoport M, Seah C, Pasterán F, Ceriana P, et al. Molecular characteristics of mcr-1-carrying plasmids and new mcr-1 variant recovered from polyclonal clinical Escherichia coli from Argentina and Canada. PLoS One. 2017;12(7):e0180347.

44. Bridging genomics, microbiology and big data to tackle global antimicrobial resistance (AMR). Bethesda (MD): Global Health Research Unit. National Institute for Health Research; 2019 (https://ghru.pathogensurveillance.net/, accessed April 2020).

45. Aanensen DM, Feil EJ, Holden MT, Dordel J, Yeats CA, Fedosejev A, et al. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. MBio. 2016;7(3):e00444-16.

46. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. Nature Microbiol. 2019;4(11):1919–29.

47. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasevic AT, et al. Occurrence of carbapenemase-producing *Klebsiella pneumoniae* and *Escherichia coli* in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. Lancet Infect Dis. 2017;17(2):153–63.

48. Koser CU, Fraser LJ, Ioannou A, Becq J, Ellington MJ, Holden MT, et al. Rapid single-colony whole-genome sequencing of bacterial pathogens. J Antimicrob Chemother. 2014;69(5):1275–81.

49. WHO GLASS guidance for national reference laboratories. Geneva: World Health Organization; 2020.

50. Gwinn M, MacCannell D, Armstrong GL. Next-generation sequencing of infectious pathogens. JAMA. 2019;321(9):893¬–4.

51. Revez J, Espinosa L, Albiger B, Leitmeyer KC, Struelens MJ, ECDC National Microbiology Focal Points and Experts Group. Survey on the use of whole-genome sequencing for infectious diseases surveillance: rapid expansion of European national capacities, 2015–2016. Front Public Health. 2017;5:347.

52. HIV drug resistance. Geneva: World Health Organization; 2019 (https://www.who.int/hiv/topics/drugresistance/en/, accessed April 2020).

53. Allix-Beguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. New Engl J Med. 2018;379(15):1403–15.

54. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in *Mycobacterium tuberculosis* complex: technical guide. Geneva: World Health Organization; 2018.

55. GISAID EpiFlu Database access agreement. Munich: Freunde von GISAID; 2011 (https://www.gisaid.org/registration/terms-of-use/, accessed April 2020).

56. Vong S, Anciaux A, Hulth A, Stelling J, Thamlikitkul V, Gupta S, et al. Using information technology to improve surveillance of antimicrobial resistance in South East Asia. BMJ Clin Res. 2017;358:j3781.

57. Moran-Gilad J, Sintchenko V, Pedersen SK, Wolfgang WJ, Pettengill J, Strain E, et al. Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. BMC Infect Dis. 2015;15:174.

58. Tolar B, Joseph LA, Schroeder MN, Stroika S, Ribot EM, Hise KB, et al. An overview of PulseNet USA databases. Foodborne Pathog Dis. 2019;16(7):457–62.

59. Timme RE, Rand H, Shumway M, Trees EK, Simmons M, Agarwala R, et al. Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. PeerJ. 2017;5:e3893-e.

60. Lüth S, Kleta S, Al Dahouk S. Whole genome sequencing as a typing tool for foodborne pathogens like *Listeria monocytogenes* – The way towards global harmonisation and data exchange. Trends Food Sci Technol. 2018;73:67–75.

61. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of *Mycobacterium tuberculosis:* current standards and open issues. Nat Rev Microbiol. 2019;17(9):533–45.

62. Taboada EN, Graham MR, Carriço JA, Van Domselaar G. Food safety in the age of next generation sequencing, bioinformatics, and open data access. Front Microbiol. 2017;8:909.

63. ATCC 2019. Manassas (VA): American Type Culture Collection; 2019 (https://www.atcc.org/, accessed April 2020).

64. Hendriksen RS, Karlsmose Pederson S, Larsen MV, Neubert Pedersen J, Lukjancenko O, Kaas RS, et al. The proficiency test (pilot) report of the global microbial identifier (GMI) initiative, year 2014. Lingby: National Food Institute, Technical University of Denmark; 2016.

65. Unemo M, Golparian D, Sánchez-Busó L, Grad Y, Jacobsson S, Ohnishi M, et al. The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. J Antimicrob Chemother. 2016;71(11):3096–108.

66. Abdurrahman ST, Emenyonu N, Obasanya OJ, Lawson L, Dacombe R, Muhammad M, et al. The hidden costs of installing Xpert machines in a tuberculosis high-burden country: experiences from Nigeria. Pan Afr Med J. 2014;18:277.

67. Johnson SB, Parker M. The ethics of sequencing infectious disease pathogens for clinical and public health. Nat Rev Genet. 2019;20(6):313–5.

68. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Off J Eur Union; 2016;L 119/1.

69. Sintchenko V, Holmes EC. The role of pathogen genomics in assessing disease transmission. BMJ. 2015;350:h1314.

70. O'Brien TF, Stelling JM. WHONET: an information system for monitoring antimicrobial resistance. Emerg Infect Dis. 1995;1(2):66.

71. Guidelines for surveillance of drug resistance in tuberculosis. Geneva: World Health Organization; 2015.

72. Shafer RW. Rationale and uses of a public HIV drug-resistance database. J Infect Dis. 2006;194(Suppl 1):S51–8.

73. WHO treatment guidelines for drug-resistant tuberculosis (2016 update). Geneva: World Health Organization; 2016.

74. Global tuberculosis report 2018. Geneva: World Health Organization; 2018.

75. Seale AC, Gordon NC, Islam J, Peacock SJ, Scott JAG. AMR surveillance in low and middle-income settings – A roadmap for participation in the Global Antimicrobial Surveillance System (GLASS). Wellcome Open Res. 2017;2:92.

76. Hicks AL, Kissler SM, Lipsitch M, Grad YH. Surveillance to maintain the sensitivity of genotype-based antibiotic resistance diagnostics. PLoS Biol. 2019;17(11):e3000547.

77. Köser CU, Ellington MJ, Peacock SJ. Whole-genome sequencing to control antimicrobial resistance. Trends Genet. 2014;30(9):401–7.

08 Annexes

Annex 1. Mechanisms of antimicrobial resistance

Antimicrobial resistance mechanisms fall into four main categories: (i) limiting uptake of a drug; (ii) modifying a drug target; (iii) inactivating a drug; and (iv) active drug efflux. Intrinsic resistance may make use of limiting uptake, drug inactivation, and drug efflux; acquired resistance mechanisms used may be drug target modification, drug inactivation, and drug efflux. Because of differences in structure, there is variation in the types of mechanisms used by gram negative bacteria versus gram positive bacteria (1).

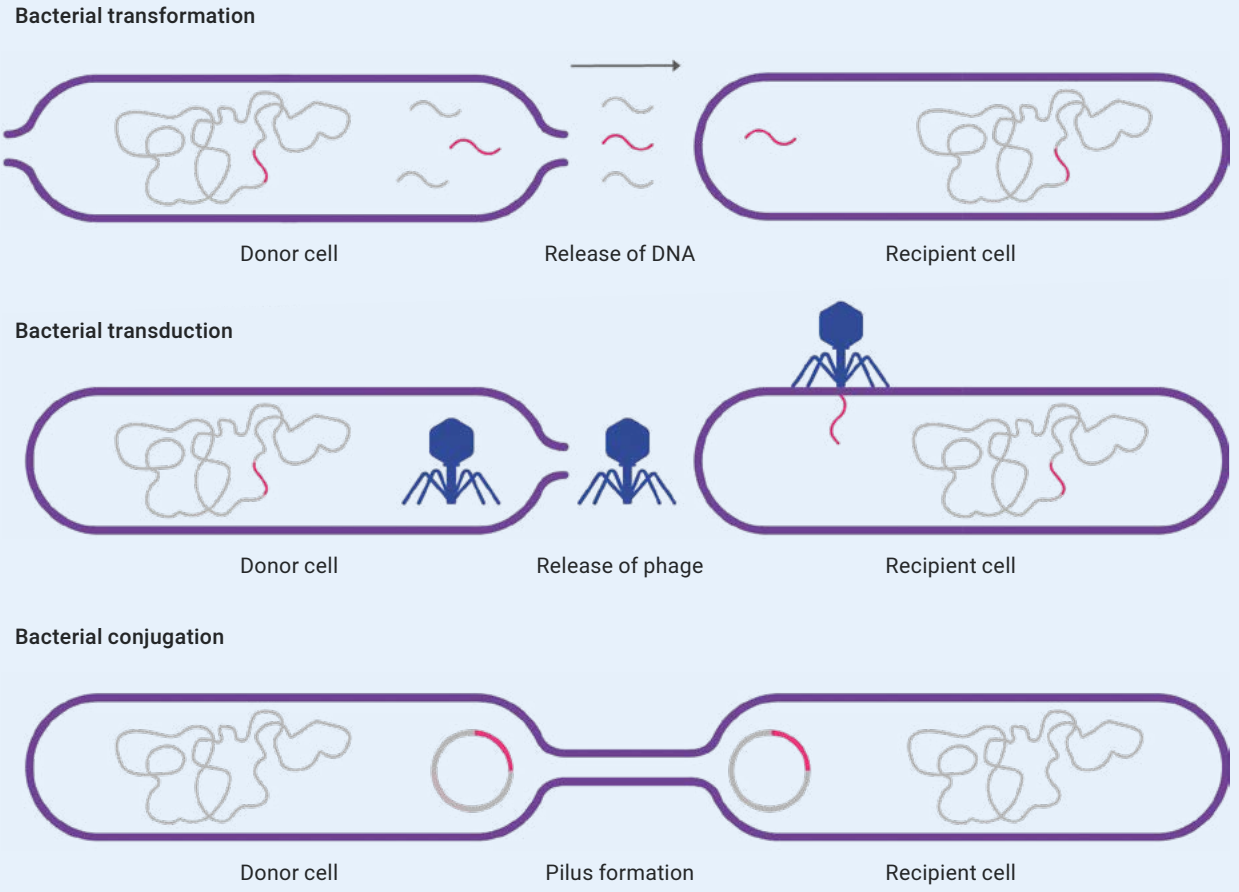
Microbial pathogens acquire resistance to antimicrobials through a variety of different mechanisms that can lead to differential expression of AMR genes. WGS can help to elucidate both the novel molecular mechanisms underlying resistance to specific antimicrobials and the pathways of AMR acquisition and can provide the evidence base of associations between genotype and phenotype, identify variants, sequence types and plasmid types, which are critical for interpreting WGS data for public health purposes. Better monitoring of the mechanisms and sources of resistance can help to better understand emergence and spread, and thus inform efforts fro prevention. As more WGS data are generated through AMR surveillance systems, the database for AMR related analyses will also improve. Extending and improving existing databases of AMR genes and mutations and algorithms to analyse them will represent a valuable global public health good to help inform and target strategies for the prevention and control AMR.

Some AMR mechanisms are intrinsic characteristics to certain pathogens. For example, *Pseudomonas aeruginosa* has a high level of intrinsic resistance to many antimicrobials because of restricted outer membrane permeability, efflux systems that pump antimicrobials out of the cell and the production of antimicrobial-inactivating enzymes such as β-lactamases *(2).* Resistance to other classes of antimicrobials is acquired by selection of mutations in core genes. For example, point mutations at amino acid positions 91 and 95 of the *gyrA* gene and 86 through 88 and 91 of the *parC* gene can confer quinolone resistance to quinolone in *N. gonorrhoeae (3).* It has also been shown that re-arrangement of non-coding DNA elements can lead to the expression of novel peptides that confer resistance to aminoglycosides in *E. coli (4).*

The commonest way of acquiring AMR by bacterial pathogens however, is by lateral gene transfer among different strains of the same bacterial species and even among different bacterial species or genera. For example, in a hospital environment, various bacterial species may have the same resistance determinants that are spread by mobile genetic elements. Lateral gene transfer comprises four general mechanisms (see Fig. A.1):

- Two different bacterial strains can exchange plasmids that may harbour AMR genes, a process called conjugation. For instance, previously susceptible *A. baumannii* strains can acquire AMR genes to carbapenems by conjugation *(5).* Conjugative plasmids act as vehicles for mobile elements, including transposons and integrons, to disseminate AMR determinants rapidly in bacteria.
- Another mechanism by which AMR genes can be acquired is by transformation, in which some bacteria take up DNA from the environment (originating, for example, from lysed bacterial cells) and integrate it into their own genomes, is another mechanism by which AMR genes can be acquired. Transformation plays a prominent role in AMR acquisition in *Neisseria gonorrhoeae,* which is naturally competent for transformation throughout its entire life cycle *(6).*
- A third way of acquiring AMR is transduction, i.e. the transfer of DNA through bacteriophages (viruses that infect bacterial cells). For example, it was shown that different *Salmonella* strains and penicillin non-susceptible *S. pneumoniae* strains acquired MDR during transduction events via bacteriophages *(7, 8).*
- Finally, AMR genes can, also, be disseminated through vesicles in the outer membrane *(9).*

Fig. A1.1. Mechanisms of lateral gene transfer in bacteria that can lead to the spread of AMR genes and mutations



In transformation, some bacteria take up DNA from the environment (originating, for example, from lysed bacterial cells) and integrate it into their own genomes, which is a mechanism by which AMR genes can be acquired. In transduction, DNA is transferred from one bacterial cell to another by bacteriophages (viruses that infect bacterial cells). In conjugation, plasmids (circular extrachromosomal bacterial DNA) act as vehicles for mobile elements, including transposons and integrons, to disseminate AMR determinants in bacteria. In addition, plasmids can be passed from one bacterial cell to another via outer membrane vesicles, for example in *Acinetobacter* spp.

In some pathogens, the mechanisms of resistance to a specific antimicrobial are highly targeted and are well understood. This is the case, for example, for methicillin resistance in *S. aureus* and for fluoroquinolone resistance in *N. gonorrhoeae (10, 11).* In other pathogens, the various molecular mechanisms that contribute to AMR are less well characterized and difficult to predict by genotyping alone, e.g. as in the case of cefixime and ceftriaxone resistance in *N. gonorrhoeae (11–13).* Moreover, the expressed phenotype may differ according to the genetic epistatic background of the pathogen or strain, making prediction of AMR from mutations more complex.

A sustained, global effort will be required as microbial pathogens continue to evolve and acquire resistance determinants. Work might start with a few selected pathogens, to optimize the impact of WGS as soon as possible. National AMR surveillance systems in many countries are currently focusing on GLASS priority pathogens. Therefore, a possible strategy for introducing WGS into AMR surveillance could be to begin with GLASS priority pathogens or the list of critical resistance phenotypes of the reporting

framework for emerging antimicrobial resistance (GLASS-EAR) *(14).* Alternatively, a selected subset of pathogens for which the molecular AMR mechanisms are well understood could be selected *(15),* to prioritize actionable data over novel discovery. An advantage of this approach would be the availability of a core set of software and standard operating procedures, derived from laboratories that are already providing such services in this area, to accelerate development of new services and foster the development of standards and best practice. Table A2.1 in Annex 2 lists current GLASS priority pathogens and critical AMR phenotypes.

References

1. Reygaert WC. An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiol. 2018;4(3):482–501.
2. Sintchenko V, Holmes EC. The role of pathogen genomics in assessing disease transmission. BMJ. 2015;350:h1314.
3. Moran-Gilad J, Sintchenko V, Pedersen SK, Wolfgang WJ, Pettengill J, Strain E, et al. Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. BMC Infect Dis. 2015;15:174.
4. Tolar B, Joseph LA, Schroeder MN, Stroika S, Ribot EM, Hise KB, et al. An overview of PulseNet USA databases. Foodborne Pathog Dis. 2019;16(7):457–62.
5. Timme RE, Rand H, Shumway M, Trees EK, Simmons M, Agarwala R, et al. Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. PeerJ. 2017;5:e3893-e.
6. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect. 2017;23(1):2–22.
7. Lüth S, Kleta S, Al Dahouk S. Whole genome sequencing as a typing tool for foodborne pathogens like *Listeria monocytogenes* – The way towards global harmonisation and data exchange. Trends Food Sci Technol. 2018;73:67–75.
8. Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, et al. Whole genome sequencing of *Mycobacterium tuberculosis:* current standards and open issues. Nat Rev Microbiol. 2019;17(9):533–45.
9. ATCC 2019. Manassas (VA): American Type Culture Collection; 2019 (https://www.atcc.org/, accessed April 2020).
10. Hendriksen RS, Karlsmose Pederson S, Larsen MV, Neubert Pedersen J, Lukjancenko O, Kaas RS, et al. The proficiency test (pilot) report of the global microbial identifier (GMI) initiative, year 2014. Lingby: National Food Institute, Technical University of Denmark; 2016.
11. Unemo M, Golparian D, Sánchez-Busó L, Grad Y, Jacobsson S, Ohnishi M, et al. The novel 2016 WHO *Neisseria gonorrhoeae* reference strains for global quality assurance of laboratory investigations: phenotypic, genetic and reference genome characterization. J Antimicrob Chemother. 2016;71(11):3096–108.
12. Abdurrahman ST, Emenyonu N, Obasanya OJ, Lawson L, Dacombe R, Muhammad M, et al. The hidden costs of installing Xpert machines in a tuberculosis high-burden country: experiences from Nigeria. Pan Afr Med J. 2014;18:277.
13. Emerging antimicrobial resistance reporting framework. Geneva: World Health Organization; 2018.
14. Bridging genomics, microbiology and big data to tackle global antimicrobial resistance (AMR). Bethesda (MD): Global Health Research Unit. National Institute for Health Research; 2019 (https://ghru.pathogensurveillance.net/, accessed April 2020).
15. Eyre DW, Shaw R, Adams H, Cooper T, Crook DW, Griffin RM, et al. WGS to determine the extent of *Clostridioides difficile* transmission in a high incidence setting in North Wales in 2015. J Antimicrob Chemother. 2019;74(4):1092–100.

Annex 2. Global Antimicrobial Resistance and Use Surveillance System priority pathogens and associated antimicrobial resistance

PATHOGEN	ANTIMICROBIAL CLASS
Escherichia coli	Sulfonamides and trimethoprim Fluoroquinolones Extended-spectrum cephalosporins Carbapenems Polymyxins Penicillins
Klebsiella pneumoniae	Sulfonamides and trimethoprim Fluoroquinolones Extended-spectrum cephalosporins Carbapenems Polymyxins Penicillins
Acinetobacter spp.	Tetracyclines Aminoglycosides Carbapenems Polymyxins
Staphylococcus aureus	Penicillinase-stable β-lactams
Streptococcus pneumoniae	Penicillins Sulfonamides and trimethoprim Extended-spectrum cephalosporins
Salmonella spp.	Fluoroquinolones Extended-spectrum cephalosporins Carbapenems
Shigella spp.	Fluoroquinolones Extended-spectrum cephalosporins Macrolides
Neisseria gonorrhoeae	Extended-spectrum cephalosporins Macrolides Aminocyclitols Fluoroquinolones Aminoglycosides
*Pseudomonas aeruginosa**	Extended-spectrum cephalosporins Aminoglycosides Quinolones Polymyxins
*Candida spp.**	Triazole antifungals Amphotericin B Echinocandin antifungals

* Not currently in GLASS but will be included in the near future.

Annex 3. Sequencing devices and bioinformatics tools

Table A3.1. Technical specifications and cost of available sequencing devices

PLATFORM	READ LENGTH (BP)	YIELD (GB)	RUN TIME	INSTRUMENT COST (US$)	ANNUAL CONTRACT COST (US$)	COST PER GB (US$)	DISADVANTAGES	ADVANTAGES
Illumina MiniSeq	50–150	1.6–7.5	7–25 h	50 000	5 000	200–400	High cost per Gb	Low instrument cost, established technology, low error rate
Illumina MiSeq	75–300	0.5–15	4–56 h	99 000	14 000	250–2 000	High cost per Gb	Low instrument cost, established technology, low error rate
Illumina NextSeq	75–150	16–120	15–29 h	250 000	32 000	33–34	High instrument cost	Low cost per Gb, established technology, low error rate
Illumina NovaSeq 5000 and 6000	50–150	960–2064	2–6 days	850 000–985 000	60 875–152 225	10–18	High instrument cost, only suitable for the highest-throughput laboratories	Low cost per Gb, established technology, low error rate
Ion PGM	200–400	0.03–2	3.7–23 h	49 000	5 000–10 000	400–2 000	Incapable of paired-end sequencing, poor homopolymer performance, high cost per Gb	Rapid sequencing run
Ion Proton	≤ 200	≤ 10	2–4 h	224 000	20 000–30 000	80	Incapable of paired-end sequencing, poor homopolymer performance	Low cost per Gb, rapid sequencing run
Ion S5	200–400	0.6–8	2.5–6 h	65 000	9 000–18 000	80–500	Incapable of paired-end sequencing, high cost per Gb	Rapid sequencing run
Pacific BioSciences RSII	20 000	1	4 h	695 000	84 000	1 000	13% single pass error rate, high cost per Gb, high instrument cost	Very long read lengths, can sacrifice length for accuracy, rapid run time
Pacific BioSciences Sequel	20 000	5	4 h	350 00	20 000	1 000	13% single pass error rate, high cost per Gb, high instrument cost	Very long read lengths, can sacrifice length for accuracy, rapid run time
Oxford Nanopore MK 1 MinION	≤ 200 000	< 10	< 48 h	1 000	0	100–400	10% single pass error rate, increased indel errors in repeat regions, high cost per Gb	Very low instrument cost, portable

Source: reference *1*

Gb, gigabits. Information on Illumina HiSeq instruments is not given in the table, as they are to be discontinued; information on the novel Illumina NovaSeq instruments is included.

Table A3.2. Assembly tools and compatible platforms

ASSEMBLY TOOL	WORKS WITH PLATFORMS	REFERENCE
Canu	PacBio, ONT	*(3)*
IDBA-IU	Illumina	*(4)*
Minimap/miniasm	PacBio, ONT	*(5)*
MIRA	Illumina, Ion Torrent, can perform hybrid assembly	*(6)*
RAY	Illumina, Ion Torrent, can perform hybrid assembly	*(7)*
Shovill	Illumina	Unpublished
SKESA	Illumina	*(8)*
SPAdes	Illumina, Ion Torrent, PacBio, ONT, can perform hybrid assembly	*(9)*
Unicycler	Illumina, Ion Torrent, PacBio, ONT, can perform hybrid assembly	*(10)*
Velvet	Illumina	*(11)*

Source: modified from references *1, 2.* All listed assembly tools are free software.

Table A3.3. Metrics commonly used for quality control of WGS data

METRIC	EXPLANATION
Absorbance ratio	QC of DNA extraction, tests for contamination of DNA with protein or RNA, measured at 260–289 nm, a ratio of 1.8 indicates no relevant contamination.
Identity check	QC of DNA extraction. Highly polymorphic microsatellite markers are compared with other potential sources of DNA to rule out contamination.
QC of reads	Software FastQC checks of raw sequence data from high-throughput sequencing
Taxonomic testing of sequenced isolates	Software for microbiological identification of the organism from which the reads were obtained (e.g. Kraken or other software based on average nucleotide identity). This check ensures that the sequenced organism is of the expected genus and species.
Number of reads	Sequence yield or amount of sequenced DNA
Average read length	Average length of all reads measured in base pairs
Depth of coverage	Number of reads times the average read length divided by the length of the genome to estimate the average number of times each nucleotide was sequenced; known as “coverage”
Size of assembled genome	Contamination is present if the ratio of the size of the assembled genome to the size of the expected genome is that expected
Total number of contigs	Total number of contigs assembled; < 250–300 generally indicates good quality for short reads.
N50	The minimum contig length required to cover half the genome. A “contig” is a set of overlapping DNA fragments that represent an inferred region of the genome. N50 > 15 000 usually indicates good quality, but a minimum of 30 000 base pairs is often preferred.

Source: reference *12*

Table A3.4. Publicly available, open-access AMR gene databases

DATABASE	LINK	REFERENCE
ABRES Finder	http://scbt.sastra.edu/ABRES/index.php	Unpublished
Abricate	https://github.com/tseemann/abricate	Unpublished
ARDB	https://ardb.cbcb.umd.edu/	*(14)*
ARG-ANNOT	https://www.mediterranee-infection.com/arg-annot/	*(15)*
BacMet	http://bacmet.biomedicine.gu.se/	*(16)*
β-Lactamases database	http://ifr48.timone.univ-mrs.fr/betalactamase/public/	Unpublished
BLAD	http://www.blad.co.in/	Unpublished
BLDB	http://bldb.eu/	*(17)*
CARD	https://card.mcmaster.ca/	*(18)*
CBMAR	http://proteininformatics.org/mkumar/lactamasedb/	*(19)*
DeepARG-DB	https://bench.cs.vt.edu/deeparg	*(20)*
FARMEDB	http://staff.washington.edu/jwallace/farme/index.html	Unpublished
Galileo AMR	https://galileoamr.arcbio.com/mara/	*(21)*
INTEGRALL	http://integrall.bio.ua.pt/?	*(22)*
ISfinder	https://www-is.biotoul.fr/index.php	*(23)*
LacED	http://www.laced.unistuttgart.de/	*(24)*
LREfinder	https://cge.cbs.dtu.dk/services/LRE-finder/	*(25)*
MEGARes	https://megares.meglab.org/	*(26)*
MUBII-TB-DB	https://umr5558-bibiserv.univ-lyon1.fr/mubii/mubii-select.cgi	*(27)*
Mykrobe	http://www.mykrobe.com/	*(28)*
NDARO	https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/	*(29)*
PATRIC	https://www.patricbrc.org/	*(30)*
PointFinder	https://cge.cbs.dtu.dk/services/ResFinder/	*(31)*
RAC	http://rac.aihi.mq.edu.au/rac/	*(32)*
RED-DB	http://www.fibim.unisi.it/REDDB/	Unpublished
ResFinder	https://cge.cbs.dtu.dk/services/data.php	*(33)*
SARG	https://galaxyproject.org/use/args-oap/	*(34, 35)*
SCCmec Finder	https://cge.cbs.dtu.dk/services/SCCmecFinder/	*(36)*
TBDReaM	https://tbdreamdb.ki.se/Info/	*(37)*
Tetracycline MLS	https://faculty.washington.edu/marilynr/	Unpublished
U-CARE	http://www.ebioinformatics.net/ucare/	*(38)*

Sources: references *1, 13.*

Table A3.4. Publicly available, open-access AMR gene databases

TOOL	TARGET AMR	LINK	REFERENCE
ABRES Finder	General	http://scbt.sastra.edu/ABRES/index.php	Unpublished
ABRICATE	General	https://github.com/tseemann/abricate	Unpublished
AMRtime	AMR genes in metagenomic data	https://github.com/beiko-lab/AMRtime	*(40)*
ARDB	General	https://ardb.cbcb.umd.edu/	*(14)*
ARG-ANNOT	General	https://www.mediterranee-infection.com/arg-annot/	*(15)*
ARGDIT	Toolkit for validation and integration of AMR gene database	https://github.com/phglab/ARGDIT	*(41)*
ARG-miner	Robust, comprehensive curation of AMR databases	https://bench.cs.vt.edu/argminer/#/home	*(19)*
ARIBA	General (single isolate sequences)	https://github.com/sanger-pathogens/ariba	*(42)*
BacMet	Biocide and metal resistance	http://bacmet.biomedicine.gu.se/	*(16)*
BLAD	β-Lactamases	http://www.blad.co.in/	Unpublished
BLDB	β-Lactamases	http://bldb.eu/	*(17)*
CARD	General AMR	https://card.mcmaster.ca/home	*(18)*
CBMAR	β-Lactamases	http://proteininformatics.org/mkumar/lactamasedb/	*(19)*
DeepARG	AMR genes in metagenomic data	https://bench.cs.vt.edu/deeparg	*(20)*
FARMEDB	AMR genes discovered by functional metagenomics	http://staff.washington.edu/jwallace/farme/index.html	Unpublished
Galileo AMR (MARA, RAC)	AMR genes in Gram-negative bacteria	https://galileoamr.arcbio.com/mara/	*(21)*
GROOT	AMR genes in metagenomic data	https://github.com/will-rowe/groot	*(43)*
INTEGRALL	AMR genes and associated integrons	http://integrall.bio.ua.pt/?	*(22)*
IRIDA plugin AMR detection	General	https://github.com/phac-nml/irida-plugin-amr-detection	Unpublished
Kmer resistance	General	https://cge.cbs.dtu.dk/services/KmerResistance-2.2/	*(44)*
LacED	β-Lactamases	http://www.laced.uni-stuttgart.de/	*(23)*
LREfinder	Linezolid resistance in enterococci	https://cge.cbs.dtu.dk/services/LRE-finder/	*(24)*
MEGARes (AMRplusplus)	General	https://megares.meglab.org/	*(25)*
MUBII-TB-DB	AMR in *M. tuberculosis*	https://umr5558-bibiserv.univ-lyon1.fr/mubii/mubii-select.cgi	*(27)*
Mustard	AMR determinants in human gut microbiota	http://mgps.eu/Mustard/	*(45)*
Mykrobe	AMR in M. tuberculosis and *S. aureus*	https://www.mykrobe.com/	*(27)*

TOOL	TARGET AMR	LINK	REFERENCE
NCBI AMRFinder	General	https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/AMRFinder/	*(46)*
Noradab	General	http://noradab.bi.up.ac.za/	*(47)*
PathogenWatch	Species-specific, focused on WHO GLASS priority pathogens	http://pathogen.watch	Unpublished
PATRIC	General AMR	https://www.patricbrc.org/	*(29)*
PointFinder	Selected mutations in chromosomal genes of *E. coli, Salmonella spp., Campylobacter spp., S. aureus, Enterococcus spp., M. tuberculosis, N. gonorrhoeae*	https://cge.cbs.dtu.dk/services/ResFinder/	*(30)*
RED-DB	β-Lactam, glycopeptide, aminoglycoside, tetracycline, sulfonamide, macrolide, lincosamide, streptogramin B, oxazolidinone and quinolone resistance	http://www.fibim.unisi.it/REDDB/	Unpublished
ResCap	Antimicrobial, biocide and metal resistance	https://github.com/valflanza/ResCap	*(48)*
ResFams	AMR genes discovered by functional metagenomics	http://www.dantaslab.org/resfams	*(49)*
ResFinder	General	https://cge.cbs.dtu.dk/services/ResFinder/	*(33)*
ResFinderFG	AMR genes discovered by functional metagenomics	https://cge.cbs.dtu.dk/services/ResFinderFG-1.0/	Unpublished
SARG (ARGs-OAP, ARGpore)	AMR genes in metagenomic data	https://smile.hku.hk/SARGs	*(34)*
SCCmec Finder	SCCmec elements in *S. aureus*	https://cge.cbs.dtu.dk/services/SCCmecFinder/	*(50)*
ShortBRED	AMR genes in metagenomic data	http://huttenhower.sph.harvard.edu/shortbred	*(51)*
SRST2	General	https://github.com/katholt/srst2	*(52)*
SSTAR	General	https://github.com/tomdeman-bio/Sequence-Search-Tool-for-Antimicrobial-Resistance-SSTAR-	*(53)*
TBDReaM	AMR in *M. tuberculosis*	https://tbdreamdb.ki.se/Info/	*(37)*
Tetracycline MLS nomenclature	Macrolide, lincosamide, streptogramin and tetracycline resistance	https://faculty.washington.edu/marilynr/	Unpublished
U-CARE	AMR in *E. coli*	http://www.e-bioinformatics.net/ucare/	*(38)*
β-lactamases Database	β-Lactamases	https://ifr48.timone.univ-mrs.fr/beta-lactamase/public/	Unpublished

Modified from reference *39.*

Some of these tools have their own AMR gene sequence databases (compare Table A2.2). Target AMR indicates the type of AMR genes detected with the tool.

References

1. The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in *Mycobacterium tuberculosis* complex: technical guide. Geneva: World Health Organization; 2018.
2. Revez J, Espinosa L, Albiger B, Leitmeyer KC, Struelens MJ, ECDC National Microbiology Focal Points and Experts Group. Survey on the use of whole-genome sequencing for infectious diseases surveillance: rapid expansion of European national capacities, 2015–2016. Front Public Health. 2017;5:347.
3. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.
4. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28(11):1420–8.
5. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016;32(14):2103–10.
6. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. Comput Sci Biol. 1999;99:45–56.
7. Boisvert S, Laviolette F, Corbeil J. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol. 2010;17(11):1519–33.
8. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 2018;19(1):153.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
10. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
11. Zerbino DR. Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics. 2010;Chapter 11:Unit 11.5
12. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, et al. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus:* a descriptive study. Lancet Infect Dis. 2013;13(2):130–6.
13. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, 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):e00483-19.
14. Liu B, Pop M. ARDB – Antibiotic Resistance Genes Database. Nucleic Acids Res. 2009;37(Database issue):D443–7.
15. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother. 2014;58(1):212–20.
16. Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DGJ. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res. 2014;42(Database issue):D737–43.
17. Naas T, Oueslati S, Bonnin RA, Dabos ML, Zavala A, Dortet L, et al. Beta-lactamase database (BLDB) – structure and function. J Enzyme Inhib Med Chem. 2017;32(1):917–9.
18. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017;45(D1):D566–73.
19. Srivastava A, Singhal N, Goel M, Virdi JS, Kumar M. CBMAR: a comprehensive beta-lactamase molecular annotation resource. Database (Oxford). 2014;2014:bau111.
20. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. 2018;6(1):23.
21. Partridge SR, Tsafnat G. Automated annotation of mobile antibiotic resistance in Gram-negative bacteria: the Multiple Antibiotic Resistance Annotator (MARA) and database. J Antimicrob Chemother. 2018;73(4):883–90.
22. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics. 2009;25(8):1096–8.
23. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32–6.
24. Thai QK, Bös F, Pleiss J. The Lactamase Engineering Database: a critical survey of TEM sequences in public databases. BMC Genomics. 2009;10(1):390.
25. Hasman H, Clausen P, Kaya H, Hansen F, Knudsen JD, Wang M, et al. LRE-Finder, a web tool for detection of the 23S rRNA mutations and the *optrA, cfr, cfr(B)* and *poxtA* genes encoding linezolid resistance in enterococci from whole-genome sequences. J Antimicrob Chemother. 2019;74(6):1473–6.
26. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, et al. MEGARes: an antimicrobial resistance database for high throughput sequencing. Nucleic Acids Res. 2017;45(D1):D574–80.

27. Flandrois JP, Lina G, Dumitrescu O. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis.* BMC Bioinformatics. 2014;15(1):107.
28. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis.* Nature Commun. 2015;6(1):10063.
29. National Database of Antibiotic Resistant Organisms (NDARO). Rockville (MD): National Center for Biotechnology Information; 2019 (https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/, accessed April 2020).
30. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res. 2017;45(D1):D535–42.
31. Zankari E, Allesoe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother. 2017;72(10):2764–8.
32. Tsafnat G, Copty J, Partridge SR. RAC: Repository of antibiotic resistance cassettes. Database (Oxford). 2011;2011:bar054-bar.
33. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640–4.
34. Yang Y, Jiang X, Chai B, Ma L, Li B, Zhang A, et al. ARGs-OAP: online analysis pipeline for antibiotic resistance genes detection from metagenomic data using an integrated structured ARG-database. Bioinformatics. 2016;32(15):2346–51.
35. Yin X, Jiang XT, Chai B, Li L, Yang Y, Cole JR, et al. ARGs-OAP v2.0 with an expanded SARG database and hidden Markov models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. Bioinformatics. 2018;34(13):2263–70.
36. Classification of Staphylococcal Cassette Chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrob Agents Chemother. 2009;53(12):4961–7.
37. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. PLoS Med. 2009;6(2):e1000002.
38. Saha SB, Uttam V, Verma V. u-CARE: user-friendly comprehensive antibiotic resistance repository of *Escherichia coli*. J Clin Pathol. 2015;68(8):648–51.
39. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using genomics to track global antimicrobial resistance. Front Public Health. 2019;7:242.
40. Maguire F, Alcock B, Brinkman F, McArthur A, Beiko R, editors. Rapid accurate Identification of antimicrobial resistance determinants from metagenomic data. In: Third American Society for Microbiology Meeting on Rapid Applied Microbial Next-generation Sequencing and Bioinformatics Pipelines; Washington DC; 2018.
41. Chiu JKH, Ong RT. ARGDIT: a validation and integration toolkit for antimicrobial resistance gene databases. Bioinformatics. 2019;35(14):2466–74.
42. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb Genomics. 2017;3(10):e000131.
43. Rowe WPM, Winn MD. Indexed variation graphs for efficient and accurate resistome profiling. Bioinformatics. 2018;34(21):3601–8.
44. Clausen PT, Zankari E, Aarestrup FM, Lund O. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. J Antimicrob Chemother. 2016;71(9):2484–8.
45. Ruppé E, Ghozlane A, Tap J, Pons N, Alvarez AS, Maziers N, et al. Prediction of the intestinal resistome by a three-dimensional structure-based method. Nature Microbiol. 2019;4(1):112–23.
46. Van Goethem MW, Pierneef R, Bezuidt OKI, Van De Peer Y, Cowan DA, Makhalanyane TP. A reservoir of "historical" antibiotic resistance genes in remote pristine Antarctic soils. Microbiome. 2018;6(1):40.
47. Lanza VF, Baquero F, Martínez JL, Ramos-Ruíz R, González-Zorn B, Andremont A, et al. In-depth resistome analysis by targeted metagenomics. Microbiome. 2018;6(1):11.
48. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 2015;9(1):207–16.
49. Kaya H, Hasman H, Larsen J, Stegger M, Johannesen TB, Allesøe RL, et al. SCCmecFinder, a web-based tool for typing of Staphylococcal cassette chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. mSphere. 2018;3(1):e00612-17.
50. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C. High-specificity targeted functional profiling in microbial communities with ShortBRED. PLoS Comput Biol. 2015;11(12):e1004557.
51. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):90.
52. de Man TJB, Limbago BM. SSTAR, a stand-alone easy-to-use antimicrobial resistance gene predictor. mSphere. 2016;1(1):e00050-15.

Annex 4. Cost of sequencing devices and broadband connection in four low- and middle-income countries

Table A4.1. Costs for installing and running Illumina MiSeq instruments in The Philippines, Colombia, Nigeria, India and the United Kingdom

COUNTRY	INITIAL INVESTMENT (US$)		RECURRING COSTS (US$)		
	Instrument (corrected for PPP)	Training and installation[a]	Accessories[b]	MiSeq v2 300-cycle kit[c] (corrected for PPP)	Annual maintenance contract[d]
Colombia	391 636.80	Included in purchasing cost	1987	4058.40	22 657
India	488 846.40	Included in purchasing cost	1546	4979.30	26 088
Nigeria	434 584.00	Included in purchasing cost (shipping only to nearest port)	Unknown	4119.30	15 985–30 724
Philippines	610 185.80	Included in purchasing cost	Included in purchasing cost	8743.20	Not yet sold in the Philippines; under warranty until 2022
United Kingdom	138 444.30	Included in purchasing cost	0	1645.70	13 227

Data collected by the Global Health Research Unit on Genomic Surveillance of Antimicrobial Resistance *(1)*, funded by the United Kingdom National Institute for Health Research.

PPP, purchasing power parity, 2018 values: USA, 1; Philippines, 18.30; Colombia, 1310.70; Nigeria, 110.20; India, 18.10; United Kingdom, 0.70 *(2)*.

[a] Includes import and shipping of the instrument (in most cases from a local distributor on the same continent or region), hardware and software installations, running one test kit cartridge (v3 - 600 cycle) and the standard warranty.

[b] Include special vibration-free benching, air-conditioning units, uninterrupted power supply units. In some cases, these are included in the sale price of the instruments. In other cases (i.e. Nigeria), a concrete platform was built instead of a table (external contractors).

[c] Includes the reagent cartridge, a flowcell and other buffers.

[d] Calculated per year beyond the warranty period.

Table A4.2. Economic indicators by country and relative cost of Internet connection

COUNTRY	PPP	EXCHANGE RATE (JUNE 2017)	RELATIVE COST OF LIVING	GDP PER CAPITA (US$)	INTERNET COST PER MONTH (US$)	RELATIVE INTERNET COST PER MONTH	INTERNET PACKAGE
Colombia	1306.94	2972.44	0.44	14 437	137.7	1.72	60 Mbps Unlimited
India	17.74	64.45	0.28	7 194	59.7	0.75	75 Mbps 280 Gb
Nigeria	102.46	322.49	0.32	5 941	292.8	3.66	1 Mbps 60 Gb
Philippines	18.06	49.92	0.36	8 360	359.9	4.5	100 Mbps Unlimited
United Kingdom	0.71	0.78	0.91	44 292	60.6	0.76	70 Mbps Unlimited
USA	1	1	1	59 791	80	1	75 Mbps Unlimited

Data collected by the Global Health Research Unit on Genomic Surveillance of Antimicrobial Resistance *(1),* funded by the United Kingdom National Institute for Health Research.

PP, purchasing power parity; GDP, gross domestic product; Mbps, megabits per second; Gb, gigabits

PPP values from reference 3, exchange rates from reference 4, GDP per capita from reference 5. Relative cost of living was calculated as the ratio of PPP and the exchange rate. Values for Internet cost per month and Internet package information were contributed by the participating centres. The relative Internet cost per month was calculated in US$ and divided by the Internet cost per month in the USA as a reference value.

References

1. Bridging genomics, microbiology and big data to tackle global antimicrobial resistance (AMR). Rockville (MD): Genomic Surveillance of Antimicrobial Resistance, Global Health Research Unit, National Institute for Health Research; 2019 (https://ghru.pathogensurveillance.net/, accessed April 2020).
2. Knoema. New York City (NY); undated (https://knoema.com/atlas/United-States-of-America/topics/Economy/Inflation-and-Prices/Purchasing-power-parity?compareTo=GB,CO,NG,IN,PH, accessed April 2010).
3. United States of America purchasing power parity, 1960–2018. Knoema. New York City (NY); undated.
4. United States of America Exchange rate, 2010–2019. Knoema. New York City (NY); undated.
5. United States of America GDP per capita based on PPP, 1980–2018. Knoema. New York City (NY); undated

ISBN 978-92-4-001100-7

World Health Organization
20 avenue Appia
1211 Geneva 72 - Switzerland
https//www.who.int/health-topics/antimicrobial-resistance