Service Contract for the provision of EU networking and support for public health reference laboratory functions for antimicrobial resistance in priority healthcare associated infections

SC 2019 74 01

EURGen-RefLabCap agreed common WGS-based genome analysis methods and standard protocols for national surveillance and integrated outbreak investigations of carbapenem- and/or colistin-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii*

Version n°: 1.0
Date: 14-04-2023

This report was produced under the EU Third Health Programme 2014-2020 under a service contract with the Consumers, Health, Agriculture and Food Executive Agency (Chafea) acting under the mandate from the European Commission. From 1 April 2021, a new executive Agency with name HaDEA (Health and Digital Executive Agency) is taking over all contractual obligations from Chafea. The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the Commission/Executive Agency. The Commission/Executive Agency do not guarantee the accuracy of the data included in this study. Neither the Commission/Executive Agency nor any person acting on the Commission's/Executive Agency's behalf may be held responsible for the use which may be made of the information contained therein.

Authors: Valeria Bortolaia, Ana Rita Rebelo, Camilla Wiuff Coia, Rene S Hendriksen, Birgitte Helwigh, Anders Rhod Larsen

Citation: Technical University of Denmark and Statens Serum Institute, Denmark (2023). EURGen-RefLabCap agreed common WGS-based genome analysis methods and standard protocols for national surveillance and integrated outbreak investigations of carbapenem- and/or colistin-resistant *Pseudomonas aeruginosa* and *Acinetobacter baumannii*. Available from https://www.eurgen-reflabcap.eu

ISBN: 978-87-7586-020-3

EUROPEAN COMMISSION
Directorate-General Health and Food Safety (DG SANTE)
Directorate B — Public health, Cancer and Health security
Unit B2 — Health security
*L-2*920 Luxembourg
Email : SANTE-CONSULT-B2@ec.europa.eu

Health and Digital Executive Agency (HaDEA)
HaDEA COV2
Place Rogier, 16
B-1049 BRUXELLES
Belgium
Email : HaDEA-HP-TENDER@ec.europa.eu

TABLE OF CONTENTS

1. INTRODUCTION 4
2. GUIDANCE DOCUMENT 5
3. SURVEILLANCE OF CARBAPENEM- AND/OR COLISTIN-RESISTANT *P. AERUGINOSA* AND *A. BAUMANNII* AND OUTBREAK INVESTIGATION 19

1. INTRODUCTION

The EURGen-RefLabCap project is complementary to the European Centre of Disease Prevention and Control (ECDC) European Antimicrobial Resistance Genes Surveillance Network (EURGen-Net). The project aims at improving capacities of National Reference Laboratories (NRLs) in European countries for identification and for phenotypic and genotypic characterization of carbapenem- and/or colistin-resistant *Enterobacterales* (CCRE), *Pseudomonas aeruginosa* and *Acinetobacter baumannii*. Furthermore, the project aims at strengthening capacities for national surveillance and outbreak investigation of those pathogens and improve the availability and quality of European-level molecular surveillance data. One of the main goals of the EURGen-RefLabCap project is to support modernisation of diagnostic and molecular typing tests using whole-genome sequencing (WGS) analytical methods to achieve those respective aims.

This guidance document provides a framework to perform WGS directed towards short-read paired-end massive parallel synthesis sequencing, specifically using Illumina platforms (Illumina, Inc., San Diego, CA, USA) such as MiSeq and NextSeq. In addition, it presents the framework for bioinformatic analysis of *P. aeruginosa* and *A. baumannii* using three pipelines to detect antimicrobial resistance determinants – particularly, acquired resistance to colistin and carbapenems. The protocol covers the steps of obtaining high-quality DNA, performing library preparation and sequencing of the DNA, performing bioinformatics analysis (taxonomic analysis, bacterial typing, detection of genetic determinants of antimicrobial resistance, cluster analysis) and adopting best practices for data management. Furthermore, this protocol defines specific quality control (QC) strategies, QC parameters and respective thresholds. Using other WGS platforms might yield results of equally good quality, but the bioinformatics tools and QC thresholds should be adapted accordingly.

Note: In most cases, WGS-based outbreak analysis cannot stand alone in outbreak investigations, but it is a powerful tool to guide epidemiological investigations.

2. GUIDANCE DOCUMENT

This guidance document describes the steps and key parameters necessary to generate and analyse WGS data of *P. aeruginosa* and *A. baumannii*. For each step, different methods, kits, and tools exist. This guidance document mentions only some of the available methods and tools, as possibilities are almost endless and continuously updated. In the case of bioinformatics analysis, this guidance document mainly refers to open-source, curated bioinformatics tools and databases.

Each laboratory should carefully consider and take into account the existing and available consumables, kits, tools and equipment that can be applied to the WGS procedure. Users might opt to employ different approaches as long as these are properly validated for the purpose. Of note, the EURGen-RefLabCap neither endorses nor is endorsed by any of the companies, brands or products referred in this document.

It cannot be stressed enough that quality control (QC) at specific checkpoints (as described further in this document) is critical to ensure confidence in the correctness of obtained results. Therefore, users should always perform a thorough evaluation of the quality of raw reads and assemblies before any further analysis.

Compared to other bacterial species frequently involved in clinical infections, *A. baumannii* and *P. aeruginosa* present some challenges when being analysed through WGS.

For *A. baumannii*, determining the exact species within the *A. baumannii* complex can be challenging even for laboratories equipped with MALDI-TOF and 16S rRNA gene PCR/sequencing technologies. WGS of a non-*baumannii* species in the *A. baumannii* complex, which was mistakenly identified as *A. baumannii*, would result in inability to map an adequate number of cgMLST targets, which will trigger a failure of the cgMLST default target identification (Fida et al., 2022). This is something to consider for possible troubleshooting of cluster analysis results.

For both *A. baumannii* and *P. aeruginosa*, WGS-based detection of AMR genes presents challenges due to the fact that, in these species, multiple chromosomal mechanisms (active efflux, porin alteration or deficiencies), which would often not be detectable by WGS, may mediate AMR.

Carbapenemase testing before WGS would be clinically useful particularly for *P. aeruginosa* as carbapenem resistance in this species may often be mediated by mechanisms other than carbapenemases. The EUCAST guidelines for detection of resistance mechanisms and specific resistances of clinical and/or epidemiological importance describe methods for phenotypic carbapenemase detection in *A. baumannii* and *P. aeruginosa*, which might be conducted before selecting isolates for WGS.

Finally, users of this guidance document are encouraged to use the EURGen-RefLabCap network to ask for support and/or share expertise.

Procedure	Theory/ Comments
DNA extraction and QC	
1. From a primary culture, select one single isolated colony to prepare a subculture.	Streaking out a fresh culture from a single colony should be implemented as a routine.
2. Inspect the subculture carefully to ensure purity. If the culture is not pure, prepare a new subculture.	Do not extract DNA from cultures that are not pure.
3. Extract bacterial DNA using in-house protocols or commercial kits.	Examples of commercial kits are, among others, ThermoFisher Easy-DNA gDNA Purification Kit and Qiagen DNeasy Blood & Tissue Kit. A range of instruments exists for more automated high-throughput DNA extraction, two examples being the MagNa Pure 96 and the KingFisher instruments. Be aware that DNA prepared by boiling lysis is not suitable for WGS. Some laboratories prepare boiling lysates prior to DNA extraction by use of commercial kits for safety reasons. DNA should be eluted in double distilled water or Tris-HCl. EDTA-containing elution buffers should be avoided as they can interfere with subsequent processes. Extraction methods based on salt and ethanol precipitation can result in poor plasmid extraction, which can be problematic for detection of acquired antimicrobial resistance (AMR) genes, that often reside on plasmids.
4. Measure UV 260/280 and 260/230 absorbance ratio values of the DNA samples to confirm that they are in the interval 1.8–2.0 and 2.0-2.2, respectively. If absorbance ratio values are outside the interval, the DNA should be re-extracted.	This can be done by using, for example, Nanodrop. Absorbance ratio values inside these ranges are obtained when DNA is contaminant-free and with high-molecular weight, which is crucial for WGS. Alternatively, size and quality of DNA fragments can be assessed using, for example, Bioanalyzer. *This step is crucial during development and implementation stages, but may be considered optional once the whole laboratory pipeline is fully validated and running in routine.*
DNA Concentration and Dilution	

5. Measure the concentration of the undiluted DNA samples. If below the necessary concentration, re-extract the DNA. If above the necessary concentration, dilute the DNA with the adequate buffer to achieve a final concentration in accordance with the library preparation protocol.	The use of a specific concentration of DNA is crucial for genomic library preparation. Many laboratories quantify DNA by using Qubit fluorometer and kits such as dsDNA reagent kits (ThermoFisher) or Quant-iT™ 1X dsDNA Assay Kit (Invitrogen), among others. Alternatively, Nanodrop spectrophotometer may be used to quantify DNA, using only 1-2 µl of sample volume. However, Nanodrop can overestimate the dsDNA present in the sample and the approach using Qubit is preferred. *Example: a final concentration of 2 ng/µl is needed if using the Nextera XT Library Preparation Reference Guide, with input of 5 µl of each library.* This infographic nicely summarises principles, advantages and disadvantages of the most commonly used methods for DNA quantification.
6. Confirm the DNA concentration of the diluted samples.	This may be done using, for example, the Qubit fluorometer and the Qubit™ dsDNA High Sensitivity Assay Kit *This step is crucial during development and implementation stages, but may be considered optional once the whole laboratory pipeline is fully validated and running in routine.*
7. The DNA dilution and confirmation of the DNA concentration should be repeated until the desired concentration is achieved.	In case the initial DNA concentration is too low it will be necessary to re-extract DNA from the sample or concentrate the DNA solution. Depending on the library preparation kit, this step can be avoided. For example, there are some Illumina Library preparation kits (e.g. Illumina DNA prep) that allow a broad range of dsDNA starting concentrations.
Library preparation and DNA sequencing	
8. Perform library preparation. Various methods for library preparation exist and they depend on the chosen sequencing platform.	Currently, Illumina is the most widely used sequencing platform, and protocols with preparation guidelines for specific library kits and guidelines for sequencing on the specific machinery are frequently updated and available on the Illumina website. Examples are the Illumina Nextera XT Reference guide and Illumina DNA Prep Reference Guide.

Pool libraries and load the sequencer following the manufacturer's instructions.	An interesting modification of the Illumina DNA Prep protocol, which however is recommendable only after having gained experience with the standard protocol, is the Hackflex protocol. Examples of protocols are MiSeq System Denature and Dilute Libraries Guide (15039740) and NextSeq System Denature and Dilute Libraries Guide (15048776). Other library preparation kits and protocols can be used. According to the choice, other reference guides and accessory documents might be needed.
Raw reads extraction, quality control and filtering	
9. Extract the raw reads and store them locally. The raw reads might be located on the Illumina sequencer or in a cloud solution such as Illumina sequence hub.	The raw reads are in the FASTQ file format, which also includes quality metrics (Phred scores). The cloud solution also offers a range of visual QC parameters to evaluate the sequencing run.
10. Perform QC of the sequence data. QC metrics should be determined including, as a minimum, the *average read length*, *coverage,* and *number of reads*.	It is crucial to assess the quality of the sequence data, as poor quality data lead to erroneous genomic analysis results. FastQC and Raspberry are examples of tools that can be used for this purpose. Average read length should be equal to the expected read length from the sequencing platform. Depth of coverage, defined as the number of times the sequencing machine sequences the genome, should be as high as possible. No harmonised cut-off exists, but a coverage of at least 50X should be the target in public health settings. Lower coverage values may interfere with downstream analysis and prevent comparison of inter-laboratory data. Thus, these should not be implemented routinely, even if they might be accepted for specific internal analysis. Number of reads should be sufficient to ensure a coverage of at least 50X, using the formula: "Coverage = Number of reads x (Read length / Genome size)". *Of note, this formula calculates the theoretical coverage. However, reads are not distributed evenly over an entire genome, whereby many bases will be covered by fewer reads than the average*

	coverage, while other bases will be covered by much more reads than average. The real coverage can be quantified by mapping reads to a reference genome.
Raw reads should be examined for potential contaminations.	KRAKEN is an example of a software that can be used to quantify the number of reads assigned to species other than the target species. The percentage of reads assigned to other species should be residual (for example less than 5%). Contamination checks can also be facilitated by tools such as KmerFinder or rMLST. An informative review and benchmarking study of different software to assess genome contamination can be found here.
Raw reads should be trimmed for adaptors and low-quality regions.	Using tools such as Bbtools or Trimmomatic.
If QC thresholds are not achieved, the DNA should be re-sequenced or even re-extracted.	
Genome assembly and QC	
11. Assemble the reads (FASTQ files) into contigs (FASTA files).	Genome assembly can be done by two methods: reference-based assembly by mapping, and *de novo* assembly. Important considerations should be made when choosing which method to use. This article provides valuable information for reference-based assembly, whereas this article examines *de novo* assembly methods. There are several software to produce genome assemblies. This article provides an overview of the most common workflows for producing bacterial assemblies, according to 2020 data. This study shows that the most used assembly software in NCBI RefSeq (the NCBI Reference Sequence Database) is SPAdes. *It is crucial that the assembly pipeline includes post-assembly corrections by mapping of the reads to the assembly and updating the consensus sequence. This can be performed e.g. by using the careful option with SPAdes.*

	Benchmarking of widely used assembly software can be found here. Most assembly programs can be installed locally, and many institutions performing WGS routinely have this step incorporated into their analysis pipeline.
12. Perform QC of the assembly.	A tool that can be used for this purpose is QUAST. Other available public QC and assembly pipelines, such as BIFROST, exist on Github or other repositories.
QC metrics should include, as a minimum: *number of contigs, N50, coverage* and *genome size*.	Most assembly QC parameters are dependent on the sequencing platform and bacterial species. Based on empirical data, if using Illumina platforms to analyse *P. aeruginosa* and *A. baumannii*:
The proposed QC thresholds should, in principle, guarantee that results obtained with FASTA files are comparable with results obtained with FASTQ files. Furthermore, using benchmarking datasets ensures that the selected assembly tool and QC thresholds yield accurate results.	Number of contigs should be less than 500. A higher number may point to poor sequence quality or to contamination (also with isolates belonging to the same species, which is not always detectable with species identification tools or through analysis of raw reads). N50 should be high, and larger than 15,000. Depth of genome coverage should be at least 50X. Genome size should be within 10% of deviation of the expected genome size. A larger genome size can indicate that the sample was contaminated (including with isolates belonging to the same species), while a smaller genome size can be due to poor DNA extraction or insufficient amount of sequenced data. *A. baumannii* genome size is generally approximately 4 million bp, while *P. aeruginosa* genome size can vary greatly, ranging from 5.5 to 7 million bps.
If QC thresholds are not achieved, the DNA should be re-sequenced or even re-extracted.	
Bacterial species identification and QC	
13. Use a curated bioinformatics tool to perform species identification and ensure that QC thresholds specific for the selected tool are fulfilled.	QC statistics from cgMLST analysis (see details at point #20) can be sufficient to confirm the species, with <90% core loci present indicating either that the isolate is not belonging to the expected species, or that the assembly is of low quality. To be able to differentiate between these possibilities, it is necessary to use other species identification algorithms/tools. Examples of commonly used species identification algorithms/tools are ANI

If QC thresholds of the primary and secondary tools are not fulfilled, the DNA should be re-extracted.	methods, KmerFinder, MASH, PROKKA, rMLST, Centrifuge, 16S rRNA, srst2, among others. Independent of the tool used, it is of critical importance to fulfil the QC thresholds specific for the selected tool. For example: If using ANI, the QC threshold should be confirmed as follows: - ANI > 95% between the genome under investigation and the reference genome of the type strain of the expected species (see for example BacDive to identify type strains of bacterial species). If using KmerFinder, the QC thresholds should be confirmed as follows: - at least 90% of template and of query coverage when summing up the several hits from the same species; - low number of individual hits; - high score (naturally occurring when both previous thresholds are fulfilled); - absence (or very low percentage) of hits belonging to different species. If using rMLST, the QC thresholds should be confirmed as follows: - at least 96% of support; - absence of hits belonging to different species. If the QC thresholds of your chosen tool are not fulfilled, species can be determined with a second tool if QC thresholds for this second tool are fulfilled.
Bacterial isolate typing	
14. Use a species-specific MLST typing scheme such as PubMLST.	If a sequence type (ST) is not assigned, a different scheme may be used, if available. If no schemes successfully assign a ST, the target isolate might also represent a new ST. However, the ST assignment could be affected by i) bad quality raw reads or bad quality assembly in a gene sequence belonging to the MLST scheme; and/or ii) contamination with isolates belonging to the

	same species. These eventualities should be considered when troubleshooting.
Detection of antimicrobial resistance genes (ARGs) and chromosomal point mutations (PMs) mediating antimicrobial resistance in *P. aeruginosa* and *A. baumannii*	
15. Use a curated bioinformatics tool to perform detection of genes (ARGs) and chromosomal point mutations (PMs) mediating AMR. If using ResFinder, the default analysis thresholds of minimum 90% of identity and minimum 60% of length are recommended. If using AMRFinderPlus, the default analysis thresholds of minimum 90% of identity and minimum 50% of length are recommended. If using CARD-RGI, the analysis parameters of "perfect and strict hits only" and "include nudge [nudge ≥95% identity Loose hits to strict]" are recommended. It is also possible to combine more than one tool and/or database for detection of ARGs and PMs, which requires careful evaluation of the results obtained.	Currently there are three main tools with associated databases to detect ARGs in WGS data, independent of the bacterial species: ResFinder, AMRFinderPlus and CARD-RGI. Sensitivity and specificity of the tools for detection of ARGs and PMs may be modified by adjusting the thresholds and/or parameters used for the analysis. The recommended thresholds do not allow differentiation between different types of beta-lactamases belonging to the same family. For beta-lactamases, it is a major surveillance task to have a complete DNA sequence and enzyme identification, including any allelic variation. Therefore, an additional revision of beta-lactamase results obtained with the recommended thresholds is needed, and the variants with the highest identity should be selected (see also step 16). Detection of PMs in WGS data requires a species-specific database. In the case of *A. baumannii* and *P. aeruginosa*, PMs databases are currently only available in AMRFinderPlus. It is important to note that there are tools that use the AMRFinderPlus, CARD and/or ResFinder databases with their own algorithms. Users should always remember to verify the database versions to ensure that the updates regularly done in the "original" databases have been captured. Furthermore, users of these tools should be aware that using the same database with different algorithms can lead to different results.
16. Evaluate the results (also called "hits") obtained with the chosen tool.	The output of the bioinformatics tools for detection of AMR genes must be carefully interpreted. Both *A. baumannii* and *P. aeruginosa* harbour intrinsic AMR genes, which however may confer clinically relevant resistance only in presence of strong promoters due to insertion sequences or PMs. Thus, detection of an AMR gene does not indicate phenotypic resistance by default. Furthermore, absence of AMR genes does not indicate

	phenotypic susceptibility by default, since complex mechanisms such as increased efflux and diminished permeability may often play a role in resistance in these species, and the genetic basis of these resistance mechanisms are often undetectable by WGS. Be aware of which ARG and PMs are included in your chosen database: lack of hits might be due to real absence of the genes or mutations in the query genome but might also be due to absence of those in the database. For ARGs: Length and identity of the gene(s) in the query genome (i.e. the genome you sequenced) should be equal to 100% of the gene(s) in the database used by the tool. If length < 100% and identity ≤ 100%, it should be verified if the gene is artificially truncated due to being positioned at the beginning or end of a contig or if it is truly a partial gene. If identity is < 100% and length ≤ 100%, it should be confirmed by searching other databases or literature if that variant has been described; if not, the impact of the nucleotide mutation(s) on the amino-acid sequence may be assessed: • Silent mutation: this scenario is consistent with a predicted phenotype of microbiological resistance to the relevant antimicrobial(s) (with few exceptions). • Other type of mutation: it is recommended not to predict an AMR phenotype but to report the detected gene variant and its attributes. The presence of multiple genes from the same gene family should be carefully evaluated to determine if it is an artefact of the tool/database used (which is revealed by observing if the genes are placed at the same positions in the same contig) or if it is a true occurrence. Generally, this scenario is consistent with a predicted phenotype of microbiological resistance to the relevant antimicrobial(s).

	For chromosomal PMs: Specific PMs or combinations of PMs in selected genes and bacterial species are known to mediate resistance to specific antimicrobials. If these "known" mutations are detected, the isolate likely exhibits resistance to the specific antimicrobial(s). If detecting "unknown" mutations (mutations for which a role in AMR has not been elucidated yet), results should be reported but the phenotype cannot be predicted. Bear in mind that PMs mediating AMR are generally species-specific. If performing direct analysis (for example through BLAST, as opposed to using a curated bioinformatics tool) the assessment of PMs should be done against the species-specific wild-type sequence of the target gene.
Cluster analysis and quality thresholds	
17. Design a general approach regarding frequency of cluster analysis.	For outbreak investigation purposes, cluster analysis can be initiated as soon as there are suspicions of an outbreak, and repeated as often as needed once new isolates are collected. Warning signs that might suggest that cluster analysis should be conducted are, for example, increase in incidence of a certain species or a certain sero- or sequence type, or observing unexpected AMR profiles. For routine surveillance purposes it may be decided to perform the analysis every last Friday of each month, as an example.
18. Choose a general approach regarding isolates to include in the cluster analysis.	Inclusion criteria may be "all isolates from the species", "all isolates belonging to the same MLST", "all isolates collected in the last three months", etc. These choices depend on the local and national epidemiological distribution and attributes of the species of interest. Historical data can aid in designing adequate inclusion criteria. Consider that epidemiological information is necessary for understanding the significance of cluster analysis results.
19. Perform single nucleotide polymorphisms (SNP)-based phylogenetic analysis. SNP analysis is performed with FASTQ files, and an adequate reference should be selected (i.e., an isolate with predicted high	Example of tools for SNP analysis are CSIphylogeny, FastTree and Snippy, among others. The results should be interpreted.

genetic relatedness, which can be one of the isolates being investigated).	In the absence of validated thresholds for genetic relatedness, information retrieved from specific studies is reported below. Importantly, such thresholds should be understood in the context of the specific objectives and methods of those studies. - For *A. baumannii*, a genetic relatedness threshold of ≤ 2-3 SNPs has been suggested to distinguish non-outbreak from outbreak strains (Fitzpatrick et al., 2016) - For *P. aeruginosa*, a genetic relatedness threshold of ≤ 5 SNPs has been suggested to distinguish non-outbreak from outbreak strains (Pelegrin et al., 2019). Thresholds higher than those mentioned above (e.g. up to 25 SNPs) should not be ignored because "significance" of the difference between isolates should be judged based on a comprehensive understanding of the genetics and epidemiology of the pathogen and the setting in which the issue is being observed (Hwang et al., 2021). To ensure quality, at least 90% of each query genome should have been included in the alignment to create the distance matrix; lower percentages of alignment directly suggest limited relatedness of the isolates or a non-optimal choice of reference genome used for mapping.
20. Additionally, or instead, choose a different clustering approach such as species-specific core-genome MLST (cgMLST) schemes.	For *A. baumannii*, there are currently two cgMLST schemes ("Oxford scheme": Bartual et al., 2005 and "Pasteur scheme": Diancourt et al., 2010). Freely-available, online, user-friendly interfaces for clustering approaches for *A. baumannii* are cgMLSTFinder and PathogenWatch, among others. When comparing results, it is important to be aware of which cgMLST schemes(s) are used by the chosen tool(s). For *P. aeruginosa*, although at least three cgMLST schemes have been proposed (de Sales et al., 2020; Tönnies et al., 2021; Cunningham et al., 2022), there is no freely-available online tool for cgMLST typing, that can be operated without knowledge of command line. The cgMLST proposed by de Sales et al. is publicly

	available but requires command line expertise.
	
	cgMLST approaches may provide lower resolution than SNP-based analysis. However, a well-designed and thoroughly validated cgMLST scheme may produce more robust comparisons than SNP analysis, especially for bacterial species which undergo rapid recombination events. cgMLST is also suitable for long-term surveillance as computations generally scale better with dataset size.
	
	The results should be interpreted. In the absence of validated thresholds for genetic relatedness, information retrieved from specific studies is reported below. Importantly, such thresholds should be understood in the context of the specific objectives and methods of those studies. In the future, the data generated by EURGen-Net could serve to validate or adjust these thresholds.
	
	Thresholds currently described are:
	
	- For *A. baumannii*, laboratory-validated allelic thresholds of relatedness using the "Pasteur scheme" (Fida et al., 2022) are:
	  ○ ≤ 9 allelic differences: related;
	  ○ 10 to 200 allelic differences: possibly related;
	  ○ > 200 allelic differences: unrelated.
	
	- For *P. aeruginosa*, laboratory-validated allelic thresholds of relatedness using the Mayo cgMLST scheme (Cunningham et al., 2022) are:
	  ○ ≤ 6 allelic differences: related;
	  ○ 7 to 100 allelic differences: possibly related;
	  ○ > 101 allelic differences: unrelated.
	This is similar to what was reported in another study, which also reported information on epidemiological links in addition to genetic relatedness. In this study, an allelic threshold of ≤ 12 allelic differences using a cgMLST locally developed using BioNumerics could identify epidemiologically linked isolates (Blanc et al., 2020).

	To ensure quality, at least 90% of the cgMLST loci present in the scheme must be assigned to each isolate being compared. *For outbreaks, closer similarity than the cgMLST thresholds suggested here is likely to be observed. Furthermore, outbreaks may be polyclonal, related to environmental contamination, which is a complicating factor in the interpretation of WGS data.* *It is also important to note that thresholds for cgMLST allele differences and SNP differences are not interchangeable!*
Data and metadata storage	
21. Store raw reads perpetually, either in private or public databases.	Raw reads must be accompanied by minimum metadata parameters. Examples of minimum fields are: - metadata of the isolate: bacterial species, sample collection date, type of clinical specimen, antimicrobial susceptibility test results, storage location. Patient data, epidemiological and clinical data and hospital data would ideally be linked to isolate data. An example of data useful from a European surveillance perspective can be found in the Protocol-genomic-surveillance-resistant-Enterobacteriaceae; - details on DNA extraction: date of extraction, kit used, DNA concentration, storage location; - details on library preparation protocol: date of preparation, kit used, DNA concentration of each input library, layout of the microtiter plate, normalization and dilution approaches; - sequencing platform and sequencing run: platform name, sequencing run number, sequencing start date, sequencing end date, sequencing yield; - raw reads QC: average read length, coverage, number of reads.

22. Store trimmed and assembled data likewise.	If storing assembled data, information on the assembly approach and respective QC should be included.
23. Store bioinformatic results, if feasible.	If storing bioinformatics results, at least the following details should be stored: information on the workflow, QC results, date of the analysis and/or version of the bioinformatics tools and databases used, and interpretation guidelines that were used.

3. SURVEILLANCE OF CARBAPENEM- AND/OR COLISTIN-RESISTANT *P. AERUGINOSA* AND *A. BAUMANNII* AND OUTBREAK INVESTIGATION

Analysis of WGS data for *P. aeruginosa* and/or *A. baumannii*, together with epidemiological data, is vital for detecting the emergence of high-risk clones/plasmids, monitoring of time and spatial trends, detection, and investigation of outbreaks in both community and healthcare settings and for the identification of high-risk populations, sources of transmission and prevention and control measures.

WGS-based routine and/or sentinel genomic surveillance of healthcare priority pathogens provide a cornerstone in both local, regional and national epidemic preparedness. As a first step, laboratories should implement a local sampling strategy, laboratory and clinical case definitions aligned with EUCAST guidance and EU case definitions for communicable diseases, and selection criteria for performing WGS.

WGS-based surveillance includes steps for detection of genetic determinants of AMR. Investigation mainly focuses on acquired ARGs and chromosomal PMs in specific target genes. Either of these mechanisms can lead to decreased susceptibility towards antimicrobials of clinical relevance.

It is important to note that one isolate harbouring ARGs or PMs that mediate resistance towards a class of antimicrobials can express different phenotypes to the individual agents included in that antimicrobial class. Also, different gene variants within the same gene family can lead to different phenotypes. Finally, there can be situations where the presence of an ARG will not lead to phenotypic resistance, due to variation in gene expression, possible simultaneous changes in expression of efflux pumps, and potential porin loss. Similarly, not all PMs in target genes will lead to phenotypic resistance. However, due to incomplete knowledge regarding the effects of all possible mutations in target genes, and the possibility that these PMs have a cumulative effect in the expression of resistance phenotypes, these should also be kept under surveillance. It should be reiterated that, for both *A. baumannii* and *P. aeruginosa*, WGS-based prediction of AMR is even more challenging due to the fact that multiple chromosomal mechanisms (active efflux, porin alteration or deficiencies), which would often not be detectable by WGS, may mediate AMR.

In addition to the investigation of ARGs and PMs, selected isolates from a defined site (such as a hospital or healthcare facility, the community, a region or country) can be further analysed by WGS to determine the genetic relatedness between isolates. This requires the use of a suite of genomic typing tools, including but not limited to MLST, cgMLST, and SNP-based analysis. Furthermore, plasmid content and presence of genes encoding virulence factors may also be determined using WGS data. These bacterial typing and cluster analysis strategies are able to support epidemiological analysis aimed at monitoring the introduction and expansion of high-risk multidrug-resistant clones, transmission events and detection of clusters and outbreaks.

The analytical WGS pipeline should be designed to meet the identified characterization and cluster analysis needs, by using sequencing and bioinformatics approaches that produce standardized results. Thus, to ensure comparability of WGS results among sites, agreement should be reached on the minimum quality control parameters and respective thresholds. These threshold parameters should be established with caution and always be used in combination with clinical epidemiological data, population and species characteristics.

Finally, by uploading raw reads with associated metadata to international databases, such as the European Nucleotide Archive and the National Center for Biotechnology Information, and by actively engaging in participation in the upcoming ECDC portal EpiPulse, investigations can be extended to assess cross-border transmission.

Other supporting documentation

Please refer to the EURGen-RefLabCap website https://www.eurgen-reflabcap.eu/.

[Catalogue number]

Publications Office

doi:[number]