

PROTOCOL FOR WHOLE GENOME SEQUENCING AND BIOINFORMATIC ANALYSIS OF BACTERIAL ISOLATES RELATED TO THE EU MONITORING OF ANTIMICROBIAL RESISTANCE

AUTHORED BY THE EURL-AR

1ST VERSION - JANUARY 2020

HISTORY OF CHANGES				
Version	Sections changed	Description of change	Date	Approval
1	All through the document	Minor modifications based on comments from the EURL Working group of NGS	10 Jan 2020	The EURL Working Group of NGS
Draft	New draft document	-	06 Dec 2019	Authors

Background

Application of whole genome sequencing (WGS) offers the possibility to rapidly transform and improve our understanding of the epidemiology of antimicrobial resistance (AMR) determinants. It can also enhance the prediction of antimicrobial susceptibility testing (AST), in view of the extensive information obtained with a single assay.

At present, antimicrobial susceptibility testing is performed by well-defined laboratory methods. In these, either the minimum antimicrobial concentration at which bacterial growth is inhibited (MIC, minimum inhibitory concentration), or the diameter of the zone of inhibition of the bacterial growth around an antimicrobial-containing disc are measured. The numerical values obtained (from the MIC or inhibition zone diameters) are then interpreted according to set criteria to classify isolates as wildtype/non-wildtype or susceptible/resistant, if the outcome of interest is the occurrence of acquired resistance determinants or prediction of clinical outcome, respectively.

Although supported by long-term practice and by availability of comprehensive guidelines for performance and interpretation of AST (EUCAST, CLSI), this approach still suffers from serious drawbacks; *e.g.* when defining MICs/inhibition zone diameters for bacteriostatic antimicrobials. Furthermore, each bacterial isolate can only be tested for susceptibility to a limited number of antimicrobials to contain costs. In addition, an error rate is inherent to the methodology that causes problems in result interpretations. This is especially ambiguous when MICs/inhibition zone diameters are close to breakpoint values for the definition of wildtype/non-wildtype or susceptibility/resistance.

Sequence-based technologies produce their outputs from the DNA fingerprint of the input material. This makes them prone to errors due to unforeseen contaminations. The use of WGS for prediction of resistance genes is not a trivial protocol, as it covers a range of different laboratory processes and bioinformatics applications, which can be performed in many different ways.

The whole genome sequencing (WGS) processes divides into three overall processes:

- Bacterial isolation, DNA preparation and DNA quality and quantity assessment
- Library preparation, library quality and quantity assessment and sequencing
- Bioinformatics analysis

Suggestions for commonly used technologies and methods within these areas are covered in this protocol.



Contents

Table of Contents

Important notes.....	4
Protocol	5
Bacterial isolation	5
DNA preparation and quality assessment.....	5
Library preparation.....	6
Sequencing	6
Assessment of the genome quality	7
AMR gene prediction	7
Additional analysis and sub-typing.....	7
Proficiency test	7
Online training	7
References	8
Abbreviations and acronyms.....	9
Links	10

Important notes

I) Sequencing technologies and preparation protocol:

The sequencing technology used for generating DNA sequences is dependent on the equipment available by the NRL. The selected methodologies are of less importance, as long as the generated DNA sequences are of sufficient quality for the downstream sequence analysis. Likewise, the choice of methods/commercial kits for DNA extraction and library preparation, the type of flow cell used during the many steps in the sequencing process is the choice of the NRL or as per recommended to use for the specific sequencing equipment.

II) Continuous improvements and updates are ongoing:

There is a continuous development in all parts of the WGS processes, and protocols and products are updated at regular intervals. Therefore, protocols for the specific processes are included as links to the manufacturer's homepage, and laboratories should always check if an updated protocol is available online. When using Illumina technologies, it is always recommended to sign up for the Illumina newsletter, which contains information on new products, protocols, product recalls and other relevant information. This protocol will contain examples of widely applied protocols, and is by no means a list of all available products and protocols. A list of abbreviations is included in Table 1 and all links to protocols, resources and tools are collected in Table 2.

III) Most commonly used platforms:

The majority of laboratories performing sequencing are currently employing Illumina technologies (Illumina, Inc., San Diego, CA; <https://www.illumina.com/>), which, dependent on the required throughput, could be Illumina MiniSeq, Illumina MiSeq, Illumina NextSeq, or Illumina HiSeq, and thereby this protocol will be aimed at these technologies.

Examples of other widely used sequencing platform technologies available are Oxford Nanopore (**Link 1**; Table 1) and ThermoFisher Ion Torrent (**Link 2**) and performance indicators are furthermore described in the WHO landscape paper (WHO, 2018).

Protocol

Bacterial isolation

The bacterial isolates should be obtained through the laboratory's standard procedure for isolation of the organisms. The EURL-AR recommended protocols for isolation of ESBL, ampC and carbapenemase-producing *E. coli* from fresh meat samples and from caecal samples are available at the EURL homepage (**Link 3**).

For *Salmonella* spp., refer to the routine procedures used for isolation of *Salmonella* spp. for the national control programme. Procedures for preparation of samples are described in the EN-ISO 6887 series and for some specific products at the EURL-*Salmonella* website (**Link 4**). The international accepted method for the detection of *Salmonella* is described in EN-ISO 6579-1:2017, and for serotyping of *Salmonella* in CEN-ISO/TR 6579-3:2014. Alternative methods for detection as well as for serotyping of *Salmonella* can be used, provided they are validated in accordance with the relevant procedure described in one of the parts of EN-ISO 16140.

The purity of the bacterial isolates is crucial for obtaining DNA of the requested isolates only.

DNA preparation and quality assessment

DNA should be extracted from bacterial colonies without introducing contaminants or inhibitors, such as foreign isolates or chemicals like EDTA. The recommended extraction kit can be dependent on the preferred library preparation kit, but the laboratory's routine method for DNA extraction will often be applicable for WGS. A few issues can be considered; boiling lysates for extraction of DNA are generally not recommended to be used for WGS, and further; DNA extraction protocols based on salt and ethanol precipitation can cause the lack of plasmid extraction, which could be relevant for determining the presence of AMR genes.

One example of a widely used extraction kit is Invitrogen Easy-DNA™ Kit (Invitrogen, Carlsbad, CA, USA) (**Link 5**).

DNA can also be extracted partly or fully automated using high-throughput robotic workstations. One example of this type of equipment is a MagNA Pure 96 Instrument. A wide range of instruments are also available (**Link 6**).

DNA quantity and quality (including purity) should be assessed before proceeding to library preparation. The quality can be assessed using spectrophotometric methods, the DNA quantity can be assessed by using fluorometric methods to measure the DNA concentrations. The required DNA quantity is dependent on the library kit specifications.

Many laboratories use the Qubit dsDNA assay kit (Invitrogen, Carlsbad, CA, USA) for quantification. Hereby an overview of applications (**Link 7**) and an example of a current protocol using Qubit 4 Fluorometer (**Link 8**).

Library preparation

The methods used for library preparation are dependent on the intended sequencing platform. For library preparation from bacterial DNA, the kits recommended by Illumina, and most commonly used, are the Nextera DNA Flex and Nextera XT Library Preparation Kits. The specific protocols for sample preparations can be found using the Illumina NexteraXT™ Guide 15031942 (**Link 9**) or Illumina Nextera™ DNA Flex Library Prep Reference Guide (1000000025416) (**Link 10**). These protocols are regularly updated on the Illumina website.

Accurate quantification and proper quality check of bacterial DNA libraries are crucial for a successful sequencing run. Different methods for quantification and quality control are recommended depending on the sequencing library kit being used (**Link 11**).

For DNA quantification, fluorometric methods like Qubit or bead-based normalization are recommended, but qPCR is also an option. For quality control, Illumina recommend Bioanalyzer or FragmentAnalyzer. The quality control is optional according to Illumina when using bead-based normalization, however this can be problematic if a number of samples have low DNA concentration after library preparation.

The choice of flow cell for the sequencing run is dependent on the laboratory's need for capacity regarding both the amount of isolates sequenced, the amount of data coverage generated per sequencing run and the desired coverage for each bacterial genome. Flow cells are available for both single isolates or for several hundred isolates per sequencing run. A full list of the current reagents and flow cells available for the different sequencing platforms is regularly updated by Illumina (**Link 12**).

Sequencing

Examples of current protocols of how NexteraXT libraries are loaded onto an Illumina MiSeq (Guide 15039740; **Link 13**) or NextSeq (Guide 15048776; **Link 14**) reagent cartridge using Reagent Kit v2 or v3 are available. Protocols are depending on the sequencing equipment, and are regularly updated on the Illumina website. Protocols for other Illumina sequencing platform are also available on Illumina website.

When acquiring Illumina sequencing platforms, Illumina offers free guidance and training for their equipment and protocols. Illumina software offers a range of QC parameters for each sequencing run. These are explained in the experiment manager software guide (Guide 15031335; **Link 15**). Similar training services and QC parameters are also offered by the providers of other sequencing platforms (e.g. Thermo Fisher Ion Torrent).

Assessment of the genome quality

Quality control (QC) is essential to guarantee accuracy and precision of any laboratory test results, including WGS. Sequences (raw reads or contigs) of poor quality can lead to major errors in prediction of AST by failing to reveal AMR genes or mutations. Other sources of error may derive from contamination of the DNA or from erroneous data handling. At present, different QC parameters are available to control and standardize WGS procedures (Ellington et al. 2017). Only datasets that pass agreed QC metrics should be used in AST predictions. A set of QC parameters for draft genome assembly and their explanation has been listed by the EUCAST committee (Ellington et al. 2017). Some of the more commonly used are; number of reads, average read length, depth of coverage, size of assembled genome, total number of contigs and N50.

Several tools are available for sequence quality control. One tool is made available by DTU Food (**Link 16**). Several other QC pipelines or web-based tools are also available. More information on available tools can be found on the ENGAGE website, including a guide for genome quality assessment with a description of important criteria (**Link 17**).

AMR gene prediction

Genomes assessed to pass the QC demands are further analysed for the predicted presence of acquired AMR genes and chromosomal point mutations using available web-based tools. A recent review that describes with examples available tools and databases for antimicrobial resistance detection has been published (Hendriksen et al. 2019). The EURL-AR recommends using the ResFinder tool, which is available from the CGE website (Developed, owned and curated by DTU; **Link 18**).

Additional analysis and sub-typing

Further bioinformatics analyses, such as species determination, sub-typing of a range of different bacteria and cluster analysis, can be conducted based on other EURL-AR recommended pipelines and are available from the CGE website (**Link 19**). These include determination of MLST, cgMLST, serotyping, identification of virulence factors and plasmids, and phylogenetic SNP analysis.

Proficiency test

It is advisable for laboratories to participate in proficiency tests or ring trials regarding WGS and identification of AMR genes, to be able to evaluate the laboratory performance. One example of this is the DTU Genomic Proficiency Test 2020, an inter-laboratory performance test, which is organised to run in 2020 (**Link 20**).

Online training

DTU provides free online courses in both WGS and AMR. Visit the course home pages at Coursera for more details: Whole genome sequencing of bacterial genomes - tools and applications (**Link 21**) and Antimicrobial resistance - theory and methods (**Link 22**).

References

CEN-ISO/TR 6579-3:2014. Microbiology of the food chain – Horizontal method for the detection, enumeration and serotyping of *Salmonella* - Part 3: Guidelines for serotyping of *Salmonella* spp.

Ellington, MJ et al. (2017) The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee Clinical Microbiology and Infection, Volume 23, Issue 1, 2 – 22
DOI: <https://doi.org/10.1016/j.cmi.2016.11.012>

EN-ISO 6579-1:2017. Microbiology of the food chain – Horizontal method for the detection, enumeration and serotyping of *Salmonella* - Part 1: Detection of *Salmonella* spp.

EN-ISO 6887:2010-2017. Microbiology of the food chain - Preparation of test samples, initial suspension and decimal dilutions for microbiological examination - all parts

EN-ISO 16140: 2016-2020. Microbiology of the food chain – Method validation – all parts

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

WHO (2018) Whole genome sequencing for foodborne disease surveillance: landscape paper. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO

Abbreviations and acronyms

Table 1: Commonly used abbreviations and definitions

Abbreviation	Explanation
AMR	Antimicrobial resistance
AST	Antimicrobial susceptibility testing
CGE	Centre for Genomic Epidemiology
cgMLST	Core genome multi locus sequence type
DTU	Technical University of Denmark
ENGAGE	Project: Establishing Next Generation sequencing Ability for Genomic analysis in Europe
ESBL	Extended Spectrum Beta-Lactamase
EUCAST	The European Committee on Antimicrobial Susceptibility Testing
EURL-AR	European Union Reference Laboratory for antimicrobial resistance
CLSI	The Clinical and Laboratory Standards Institute
MIC	Minimal inhibitory concentration
MLST	Multi locus sequence type
N50	Quality control parameter; N50 is defined as the sequence length of the shortest contig at 50% of the total genome length
NGS	Next –generation sequencing
NRL	National reference laboratory
QC	Quality control
qPCR	Quantitative PCR - a method of quantifying DNA based on PCR
SNP	Single nucleotide polymorphism
WGS	Whole genome sequencing
WHO	World Health Organisation

Links

Table 2: Collection of links referred to in the protocol, including last date of accession

Link#	Method or content	Last accessed
Link 1	Oxford Nanopore website https://nanoporetech.com/products	October 2019
Link 2	Thermofisher website https://www.thermofisher.com/dk/en/home/brands/ion-torrent.html	October 2019
Link 3	Methods for isolation of ESBL, ampC and carbapenemase-producing E. coli from meat and caecal samples https://www.eurl-ar.eu/protocols.aspx	October 2019
Link 4	Method for detection of <i>Salmonella</i> in food and animal feed https://www.eurlsalmonella.eu/publications/eurl-manual	October 2019
Link 5	DNA extraction protocol EasyDNA https://assets.thermofisher.com/TFS-Assets/LSG/manuals/easydna_man.pdf	October 2019
Link 6	Automated DNA extraction Magna Pure https://lifescience.roche.com/en_dk/products/magna-pure-96-instrument-382411-1.html	October 2019
Link 7	Overview of applications of Qubit www.invitrogen.com/qubit	October 2019
Link 8	Protocol for Qubit 4 DNA quantification https://assets.thermofisher.com/TFS-Assets/BID/manuals/MAN0017210_Qubit_4_Assays_QR.pdf	October 2019
Link 9	Library prep Nextera XT http://support.illumina.com/downloads/nextera_xt_sample_preparation_guide_15031942.html	October 2019
Link 10	Library prep Nextera DNA Flex https://emea.support.illumina.com/downloads/nextera-dna-flex-library-prep-reference-guide-100000025416.html	October 2019
Link 11	Guide for quantification and QC of library prep https://emea.support.illumina.com/bulletins/2016/05/library-quantification-and-quality-control-quick-reference-guide.html	October 2019



Link 12	Illumina instrument-specific sequencing reagents, flow cells, cluster generation reagents https://emea.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents.html	October 2019
Link 13	MiSeq https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-10.pdf	October 2019
Link 14	NextSeq https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/nextseq/nextseq-denature-dilute-libraries-guide-15048776-09.pdf	October 2019
Link 15	Illumina experiment manager software guide https://support.illumina.com/downloads/illumina-experiment-manager-user-guide-15031335.html	October 2019
Link 16	Access to QC tool: FoodQCPipeline https://bitbucket.org/RolfKaas/foodqcpipeline/src/master/	October 2019
Link 17	Guide to available web tools for assessment of genome quality http://www.engage-europe.eu/resources/protocols-and-training	October 2019
Link 18	Link to web tool ResFinder 4.0 (currently in Beta testing) https://cge.cbs.dtu.dk/services/ResFinder-4.0/	October 2019
Link 19	CGE web tools https://cge.cbs.dtu.dk/services/	October 2019
Link 20	DTU Genomic Proficiency Test 2020 http://www.globalsurveillance.eu/projects/genomic-proficiency-test-2020	January 2020
Link 21	WGS online course on Coursera https://www.coursera.org/learn/wgs-bacteria	January 2020
Link 22	AMR online course on Coursera https://www.coursera.org/learn/antimicrobial-resistance	January 2020