Harmonisation of WGS AMR data

EURL-AR WGS protocol

Jette S. Kjeldgaard

DTU Food
National Food Institute

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

AUTHORED BY THE EURL-AR

3RD VERSION - DECEMBER 2021(UPDATED)

RENE S. HENDRIKSEN & JETTE S. KJELDGAARD

Background for protocol

- The Commission Implementing **Decision 2020/1729** on the monitoring and reporting of antimicrobial resistance in zoonotic and commensal bacteria
- Per January 2021 authorising the **use of WGS as an alternative method** for prediction of resistance in relation to the specific monitoring of ESBL- or AmpC- or carbapenemase-producing *E. coli* and *Salmonella*
- The EURL-AR has produced the present protocol for guidance in these matters

- 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
  – Sequence QC and bioinformatics analyses

Purpose of protocol

- **Ensure that WGS data reported to EFSA is obtained in a harmonised and comparable way**

  – Less important
    - How the bacteria, DNA and sequences are obtained

  – Very imporant
    - Assure the sequence quality control
      – Using the same QC criteria
    - Harmonised AMR gene analysis
      – using the same methods and settings for analysis
    - Reporting adequate data

Table of Contents

Links

Generic protocol – not one method that fits all

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

Link#	Method or content	Last accessed
Link 1	Illumina website https://emea.illumina.com/	November 2021
Link 2	Oxford Nanopore website https://nanoporetech.com/products	November 2021
Link 3	Thermofisher website https://www.thermofisher.com/dk/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-products-services.html	November 2021
Link 4	EURL-AR website – Inter-EURLs WG on NGS https://www.eurl-ar.eu/inter-eurls-working-group-on-ngs.aspx	November 2021
Link 5	Document on bioinformatics tools for basic analysis of Next Generation Sequencing data https://www.iss.it/documents/20126/0/Bioinformatics_tools_for_basic_analysis_of_Next_Generation_Sequencing_data_Del4.pdf/02c8f77b-db2c-6b8d-e2ba-416144f89f7e?t=1602603602556	November 2021
Link 6	Methods for isolation of ESBL, ampC and carbapenemase-producing *E. coli* from meat and caecal samples https://www.eurl-ar.eu/protocols.aspx	November 2021
Link 7	Method for detection of *Salmonella* in food and animal feed https://www.eurlsalmonella.eu/publications/analytical-methods	November 2021
Link 8	Method for detection of Campylobacter https://www.sva.se/en/about-us/eurl-campylobacter/laboratory-procedures/	November 2021
Link 9	DNA extraction protocol EasyDNA https://assets.thermofisher.com/TFS-Assets/LSG/manuals/easydna_man.pdf	November 2021
Link 10	Automated DNA extraction Magna Pure https://lifescience.roche.com/en_dk/products/magna-pure-96-instrument-382411-1.html	November 2021
Link 11	Overview of applications of Qubit https://www.thermofisher.com/dk/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html	November 2021
Link 12	Protocol for Qubit 4 DNA quantification https://assets.thermofisher.com/TFS-Assets/BID/manuals/MAN0017210_Qubit_4_Assays_QR.pdf	November 2021
Link 13	Library prep Nextera XT http://support.illumina.com/downloads/nextera_xt_sample_preparation_guide_15031942.html	November 2021

Links to protocols

- Continuously updated
- Dependent on lab
  - Equipment
  - Throughput
  - Prerequisites

Table of Contents

Bacterial isolation, DNA preparation and DNA quality and quantity assessment

- Methods for isolation of ESBL, ampC and carbapenemase-producing *E. coli* from meat and caecal samples (EURL-AR)
- Method for detection of *Salmonella* in food and animal feed (EURL-Salmonella)

  – Ensure purity and correct species

- Examples of DNA extraction kits
  – Laboratory routine methods

- Examples of DNA quality/quantity assessment

Library preparation, library quality and quantity assessment and sequencing

- Dependent on the laboratory equipment
  - Majority using Illumina sequencing equipment

At present the EURL-AR recommends Illumina sequencing
- QC of sequences
- Tools for analysis

- Suggestion for library preparation
- Quantification and QC of library prep
- Illumina instrument-specific sequencing reagents, flow cells, cluster generation reagents
  - MiSeq and NextSeq

Sequence QC and bioinformatics analyses

- Trimming of raw reads
  - Can be performed, but is not crucial for Illumina sequences

- File format
  - it is recommended to perform the assembly of fastq files into fasta files
    - part of the quality control

- Check for contamination
  - E.g. using KmerFinder for species determination and look into QC parameters

- Assembly
  - Using SPAdes 3.14 or newer
  - Accessible as CGE tool with output of important QC parameters (for EURL-AR network)

QC parameters

- The process of raw reads assembly into contigs outputs a range of QC paramters

- **number of reads**
- **depth of coverage**
- **average read length** (as specified by the sequencing equipment)
- **size of assembled genome** (+/- 0.5 million bases deviation from expected size)
- **total number of contigs** (<500 contigs)
- **N50** (>30.000 bp)

Assembly with SPAdes v 3.14

- The SPAdes 3.14 tool will output the contigs file (.fasta) and additionally a .txt file with some basic statistics and QC parameters.

- The output file contains data on:
- Input files :
  - Total number of reads
  - Total number of bases
- Contigs file :
  - Number of contigs
  - Number of bases (assembled genome size)
  - N50

- Using this output, it is also possible to calculate the average read length= Number of bases/Number of reads (input files)

AMR gene and point mutation prediction

- The EURL-AR recommends using ResFinder v4.1 or newer
- For harmonisation of the AMR data reported by different laboratories, it is important to use the defined settings.
- The EURL-AR recommends running the ResFinder analysis on the contigs **assembly files (.fasta**) using specific **settings**

- ResFinder can be run as a web-tool (CGE) or as local installation (available on BitBucket)
  – Web-tool limited to analysing one sequence at a time

ResFinder settings

For chromosomal point mutations:

- Select threshold for % ID: **90 %**
- Select minimum length: **60 %**

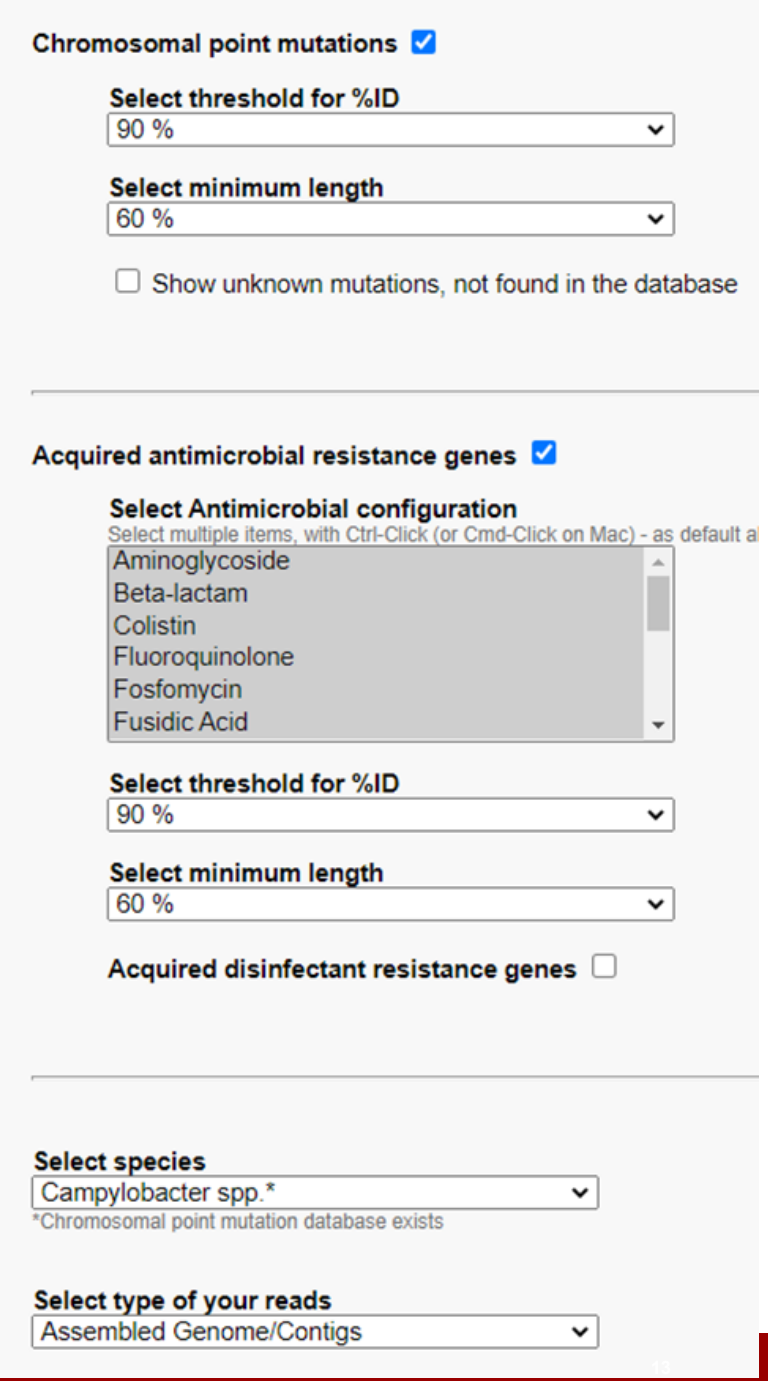
For acquired antimicrobial resistance genes:

Select all antimicrobial databases (default setting)

- Select threshold for % ID: **90 %**
- Select minimum length: **60 %**

Select species: as appropriate

Select type of your reads: Assembled genome/Contigs



Data to report to EFSA

- Beyond the sampling and isolate data, the results reported in relation to Decision 2020/1729 should include:

- Date of sequencing
- Sequencing technology used
- Library preparation used
- Version of the predictive tool (ResFinder)
- AMR-conferring genes data:
  - Gene name
  - Output information on % identity
  - Output information on % coverage (length)
- Date of ResFinder analysis

- The protocol will be added a template sheet for collection of metadata, including examples of how to report data.

Questions and discussion

Contact:

DTU FOOD – Research group of Global Capacity Building & EURL-AR Food-AMR-WGS@food.dtu.dk

Jette Sejer Kjeldgaard (jetk@food.dtu.dk)