



## Guidelines for Implementing QC Strains in Whole Genome Sequencing (WGS) Runs

### Purpose of QC Strains

The inclusion of four standardized quality control (QC) strains—two *Escherichia coli* (EURL-AMR-WGS-QC-EC01, EURL-AMR-WGS-QC-EC02) and two *Staphylococcus aureus* (EURL-AMR-WGS-QC-SA01, EURL-AMR-WGS-QC-SA02)—ensures consistent performance evaluation of WGS workflows. These strains enable laboratories to:

- Monitor sequencing accuracy and reproducibility.
- Validate bioinformatics pipelines from QC, to assembly, through downstream bioinformatic analysis that can include antimicrobial resistance (AMR) gene detection.
- Detect cross-contamination or technical artifacts.
- Benchmark against reference genomes for assembly quality.

### QC Strain Characteristics

#### *Escherichia coli*

##### 1. EURL-AMR-WGS-QC-EC01

- **Genome size:** 5.16 Mb
- **MLST:** ST410
- **Key AMR Genes:** *aac(6′)-Ib3*, *bla<sub>CMY-6</sub>*, *bla<sub>NDM-1</sub>*, *bla<sub>OXA-181</sub>*, *qnrS1*, *sul1*
- **Plasmids:** IncC, IncFIA, IncFIB, IncFII, IncX3, IncX4

##### 2. EURL-AMR-WGS-QC-EC02

- **Genome size:** 5.49 Mb
- **MLST:** ST405

- **Key AMR Genes:** *aadA2*, *bla<sub>NDM-5</sub>*, *bla<sub>CTX-M-15</sub>*, *dfrA12*, *sul1*, *tet(B)*
- **Plasmids:** IncFIB, IncFII, IncX4

### ***Staphylococcus aureus***

#### **1. EURL-AMR-WGS-QC-SA01**

- **Genome size:** 2.93 Mb
- **MLST:** ST398
- **Spa-type:** t34
- **Key AMR Genes:** *mecA*, *tet(M)*, *dfrG*, *lsa(B)*

#### **2. EURL-AMR-WGS-QC-SA02**

- **Genome size:** 2.94 Mb
- **MLST:** ST239
- **Spa-type:** t1155
- **Key AMR Genes:** *mecA*, *erm(A)*, *aac(6')-Ie/aph(2'')-Ia*

### **Protocol for QC Strain Implementation**

Laboratories can include the QC strains (EURL-AMR-WGS-QC-EC01, EURL-AMR-WGS-QC-EC02, EURL-AMR-WGS-QC-SA01, EURL-AMR-WGS-QC-SA02) in every WGS run to ensure consistent quality control across sequencing batches and to validate the performance of their specific sequencing platforms and bioinformatics tools against the provided reference data.

### **Proposed Data Analysis Workflow**

#### **1. Species Confirmation**

- Confirm the species identity of each QC strain using a suitable tool. Recommended tools include **KmerFinder**.
- Ensure the species result matches the expected species for each QC strain.

## 2. Contamination Assessment

- Assess the level of contamination (pre- and/or post-assembly) in each QC strain assembly using a suitable tool. Recommended tools include **Kraken** (pre-assembly) and/or **CheckM2** (post-assembly). Intra-species contamination can also be assessed using **ConFinder**.
- Usually contamination should be  $\leq 5\%$  (CheckM) or  $< 5\%$  of reads assigned to other taxa (Kraken).

## 3. MLST

- Determine the MLST profile using an appropriate tool and database.
- The MLST of QC strains was determined using the following schemes:
  1. *E. coli* – *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*
  2. *S. aureus* – *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, *yqiL*

## 4. AMR gene detection

- Use a suitable tool with thresholds to detect AMR genes: %identity  $\geq 90$ , %coverage  $\geq 60$ . Recommended tools include **ResFinder** or **AMRFinderPlus**.
- Confirm the presence of expected resistance genes.

## 5. Plasmid Identification

- Use a suitable tool with thresholds to identify and characterize plasmid replicons: %identity  $\geq 95$ , %coverage  $\geq 60$ . Recommended tools include **PlasmidFinder**.
- Confirm the presence of expected plasmid replicon genes.

## 6. Mapping and Coverage

- $\geq 90\%$  of reads should map to reference genomes.
- Depth of coverage should be at least 30x, with a target of 50x or higher for optimal accuracy in further bioinformatic analysis.

## 7. Genome Size Assessment

- Verify that the assembled genome size falls within  $\pm 10\%$  of the expected size for each strain to detect potential contamination or incomplete sequencing.

## Troubleshooting

Issue	Resolution
Low coverage (<30×)	Repeat library prep with increased DNA input or re-sequence
Mismatched AMR genes	Verify analysis tool settings; check for updated resistance databases
High contamination	Re-extract DNA using column-based kits (e.g., Qiagen DNeasy)
Genome size out of range	Review for contamination or incomplete data; consider re-sequencing or re-extraction if outside $\pm 10\%$ of expected size

**Contact:** EURL-AR ([Food-AMR-WGS@food.dtu.dk](mailto:Food-AMR-WGS@food.dtu.dk)) for strain-specific queries or to request the QC strains and the complete reference genomes (including a table with all bioinformatic analysis).