Guidance for Implementing QC Strains and Negative Controls in Whole Genome Sequencing (WGS) Runs

Table of Contents

1. Purpose and scope

This document provides technical guidance for the implementation of quality control (QC) strains and negative controls in bacterial whole genome sequencing (WGS) workflows used by laboratories working with antimicrobial resistance (AMR).

The purpose of this guidance is to support internal quality control, workflow verification, and ongoing performance monitoring of the wet-lab and dry-lab

DTU Food
National Food Institute
Techinical University of Denmark

Anker Engelunds Vej 101
Building 101
2800 Kongens Lyngby
Denmark

Tel. +45 45 25 25 25
Dir. +45 00 00 00 00

www.dtu.dk

components of WGS workflows. It is intended to help laboratories assess whether their WGS workflow performs consistently and is fit for its intended use.

This document provides practical guidance for the harmonized implementation of **internal quality control (QC), workflow verification, and ongoing performance monitoring** for bacterial whole genome sequencing (WGS) workflows. The QC strains, tools, and thresholds described here are **recommendations only** and are intended to support laboratories in establishing or strengthening their local WGS quality system. Laboratories may use other tools, databases, and acceptance criteria, provided that these are justified, documented, and appropriate for the intended use of the workflow. This document does **not** constitute formal method validation, accreditation, or proof of compliance with **ISO/IEC 17025, ISO 23418, or other regulatory requirements**. Each laboratory remains responsible for defining, documenting, and verifying its own workflow according to the intended application, target organism(s), sequencing platform, and local quality system.

2. General principles

The inclusion of standardized QC strains can support laboratories in:

- monitoring sequencing accuracy and reproducibility;
- verifying bioinformatics pipelines from raw-read QC through assembly and downstream analysis;
- detecting cross-contamination or technical artefacts;
- benchmarking results against reference genomes and expected genotypic characteristics.

In addition to positive QC strains, laboratories should include an appropriate negative control to monitor contamination introduced during DNA extraction, library preparation, and/or sequencing.

The tools and thresholds listed in this document are examples only and are not mandatory. Laboratories may use other suitable tools, databases, and thresholds, provided that these are documented and shown to be appropriate for the intended use of the workflow.

QC strains and negative controls may be included in each WGS run or at a laboratory-defined frequency, depending on throughput, workflow stability, previous performance, and local quality-management procedures.

1. Recommended QC strains

The following four QC strains may be used as positive controls for routine monitoring of bacterial WGS workflows:

Escherichia coli

A. **EURL-AMR-WGS-QC-EC01**
   - **Genome size**: 5.16 Mb
   - **MLST**: ST410
   - **Key AMR Genes**: *aac(6')-Ib3*, *bla*$_{\text{CMY-6}}$, *bla*$_{\text{NDM-1}}$, *bla*$_{\text{OXA-181}}$, *qnrS1*, *sul1*
   - **Plasmids**: IncC, IncFIA, IncFIB, IncFII, IncX3, IncX4

B. **EURL-AMR-WGS-QC-EC02**
   - **Genome size**: 5.49 Mb
   - **MLST**: ST405
   - **Key AMR Genes**: *bla*$_{\text{NDM-5}}$, *bla*$_{\text{CTX-M-15}}$, *mph(A), dfrA12, sul1, tet(B)*
   - **Plasmids**: IncFIB, IncFII, IncX4

Staphylococcus aureus

A. **EURL-AMR-WGS-QC-SA01**
   - **Genome size**: 2.93 Mb
   - **MLST**: ST398
   - **Spa-type:** t34
   - **Key AMR Genes**: *mecA*, *tet*(M), *dfrG*, *lsa*(B)

B. **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

2. Negative control

A suitable negative control may be:

- an **extraction blank**, for example nuclease-free water carried through the full DNA extraction procedure and then through library preparation and sequencing; and/or
- a **no-template library control**, for example nuclease-free water added at the library-preparation step when an extraction control is not included.

2.1. How the negative control should be included

- At least one negative control should be included per sequencing batch/run, or at a laboratory-defined frequency.
- The negative control should be processed using the same reagents, workflow, and handling conditions as routine samples.
- The negative control should receive its own index/barcode and be analysed together with the rest of the run.
- The negative control should be reviewed for evidence of contamination, including:
  - unexpected bacterial reads;

- recovery of a meaningful assembly;
- detection of AMR genes, plasmid replicons, or MLST results inconsistent with a true blank.

2.2. Expected outcome for the negative control

The negative control should yield no meaningful bacterial WGS result. Any substantial signal in the negative control should trigger investigation of possible reagent contamination, sample carry-over, index crosstalk, or laboratory handling issues.

NOTE: Each laboratory should define and document its own acceptance criteria and corrective actions for negative-control failures.

3. Suggested implementation in WGS runs

Laboratories may include one or more of the positive QC strains together with routine isolates to monitor run performance and compare results against expected reference data.
For routine implementation, laboratories may consider:

- including at least **one positive QC strain** and **one negative control** in each run; or
- using a laboratory-defined rotation scheme for the four QC strains across runs, while maintaining traceability and trend analysis over time.

Where possible, laboratories should record:

- sequencing platform and chemistry;
- library-preparation method;
- software tools and versions;
- database versions;
- reference genome version;
- run-level QC metrics;

- acceptance criteria;
- deviations and corrective actions.

4. Suggested Data Analysis Workflow

The following steps are recommended as part of routine QC review. Equivalent tools and locally defined workflows may be used.

4.1. Species Confirmation

Confirm the species identity of each QC strain using a suitable method. Examples of suitable tools include **KmerFinder** or other species-identification tools.

Expected result:
The species result should match the expected species for each QC strain.

4.2. Contamination Assessment

Assess contamination before and/or after assembly using suitable tools. Examples include:

- **Kraken** or equivalent tools for read-based taxonomic screening;
- **ConFindr** for detection of intra-species contamination;
- **CheckM2** for assembly-based assessment.

Example acceptance criteria:

- **>95%** of reads assigned to the target family in read-based analysis; and
- contamination approximately **≤5%** in assembly-based assessment.

NOTE: Laboratories may define different thresholds based on platform, organism, and workflow performance.

4.3. MLST

Determine the MLST profile using an appropriate tool and database.
The expected MLST schemes are:
For *E. coli*
adk, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, *recA*
For *S. aureus*
arcC, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, *yqiL*

Expected result:
The recovered MLST profile should match the expected allelic profile for the QC strain.

4.4. AMR gene detection

Use a suitable tool and database to detect AMR genes.
Examples of suitable tools include **ResFinder**, **AMRFinderPlus**, or equivalent tools.
Example reporting thresholds:

- nucleotide identity **≥90%**
- gene coverage **≥60%**

Expected result:
The expected AMR genes should be detected for each QC strain, within the limitations of the selected method, database, and threshold settings.

4.5. Plasmid Identification

Use a suitable tool to identify plasmid replicons where relevant.
Examples of suitable tools include **PlasmidFinder** or equivalent tools.
Example reporting thresholds:

- nucleotide identity **≥95%**
- coverage **≥60%**

Expected result:

Expected plasmid replicons should be detected in the relevant QC strains.

4.6. Coverage (depth and breadth)

Coverage should be assessed using two complementary metrics: depth of coverage and breadth of coverage against the appropriate reference genome.

Example acceptance criteria:

- average depth of coverage of at least 30× (routine analysis);
- a target depth of ≥95% of the reference genome covered at ≥10× for routine analysis;
- Laboratories may define additional acceptance criteria, such as mean depth, proportion of mapped reads, or stricter coverage thresholds, according to their workflow and intended use.

4.7. Genome Size Assessment

Verify that the assembled genome size falls within the expected range.

Example acceptance criterion:

- assembled genome size within ±10% of the expected genome size.

This check may help identify contamination, poor assembly, or incomplete sequencing.

5. Interpretation and follow-up

QC results should be interpreted in the context of the laboratory's own workflow, platform, and intended application. A single deviation does not necessarily invalidate all results, but it should be assessed and documented.

Examples of situations that should trigger investigation include:

- unexpected species assignment;

- incorrect MLST result;
- absence of expected AMR genes or plasmid replicons;
- excessive contamination;
- insufficient depth of coverage;
- genome size outside the expected range;
- evidence of contamination in the negative control.

Where QC results fall outside predefined acceptance criteria, laboratories should follow their internal procedures for:

- investigation of root cause;
- documentation of non-conformity;
- repeat extraction, library preparation, sequencing, or analysis, as appropriate;
- evaluation of whether the run or selected samples should be accepted, repeated, or rejected.

6. Example troubleshooting guide

Issue	Resolution
Low coverage (<30×)	Repeat library prep with increased DNA input or re-sequence
Unexpected species or MLST result	Check sample identity, potential mix-up, and contamination; review software/database versions
Missing expected AMR genes	Verify analysis settings, assembly quality, and database version; repeat analysis if needed
High contamination	Review extraction and library-preparation workflow; assess reagent contamination; repeat extraction if necessary
Genome size out of range	Review contamination, poor assembly, low coverage, or incomplete

Issue	Resolution
	sequencing; consider repeat sequencing
Signal detected in the negative control	Investigate reagent contamination, carry-over, barcode/index leakage, or handling error; assess whether the run should be repeated

Contact: EURL-AR (joam@food.dtu.dk) for strain-specific questions, for access to the QC strains, or for the associated reference data.