Foreword

The WG has been established by the European Commission with the aim to promote the use of NGS across the EURLs' networks, build NGS capacity within the EU and ensure liaison with the work of the EURLs and the work of EFSA and ECDC on the NGS mandate sent by the Commission. The WG includes all the EURLs operating in the field of the microbiological contamination of food and feed and this document represents a deliverable of the WG and is meant to be diffused to all the respective networks of NRLs.

Overview of Conducted and Planned Proficiency Tests

Introduction

The next-generation sequencing (NGS)-based proficiency tests (PTs) performed by the European Union Reference Laboratories (EURLs), i.e., EURL-*Listeria monocytogenes*, EURL-VTEC, EURL-*Salmonella*, EURL-*Campylobacter*, and EURL-Antimicrobial Resistance (AR), have been established to assist National Reference Laboratories (NRLs) in implementation of whole-genome sequencing (WGS). Generally, the PTs include quality assessment of the sequence data, phylogenetic cluster analysis (except for EURL-AR) and genotyping analysis (e.g., MLST determination, gene detection, serotyping), with the ultimate aim of helping NRLs to jointly solve national and multi-country outbreaks. The PTs also play a function in facilitating the implementation of WGS as alternative to traditional methods for bacterial characterisation and typing. Three reference laboratories, namely EURL-Coagulase Positive Staphylococcus (EURL-CPS), EURL-Foodborne Viruses and EURL-Parasites, are part of the inter-EURLs working group (WG) on NGS but have not yet organised PTs on NGS.

In the current document, the activities of the five EURLs performing PTs on NGS are summarised, specifically addressing distribution of material, data collection and analysis (Figure 1). This overview is an update of version 01 of Deliverable 1 of inter-EURLs WG NGS and aims at presenting the approaches used by EURLs on NGS PTs, which could be useful for other EURLs and NRLs intending to organise future NGS PTs.

MATERIAL
- Bacterial isolates (strains)
- DNA
- FASTQ files (raw data)

DATA COLLECTION
- FASTQ files (incl. QC)
- Fasta files
- Characterisation: *AMR profile*, *typing* (MLST, serotyping, virulotyping, *stx*-typing); *cluster analysis* (distance matrix or alignment and phylogenetic trees or minimum spanning trees (MSTs) based on cgMLST, wgMLST or wgSNPs analysis)

ASSESSMENT (ANALYSIS/DATA VALIDATION)
- FASTQ and/or fasta files submitted by the NRLs
- FASTQ and fasta quality evaluation to defined thresholds
- Assessment of the characterisation performance by comparison to pre-defined values
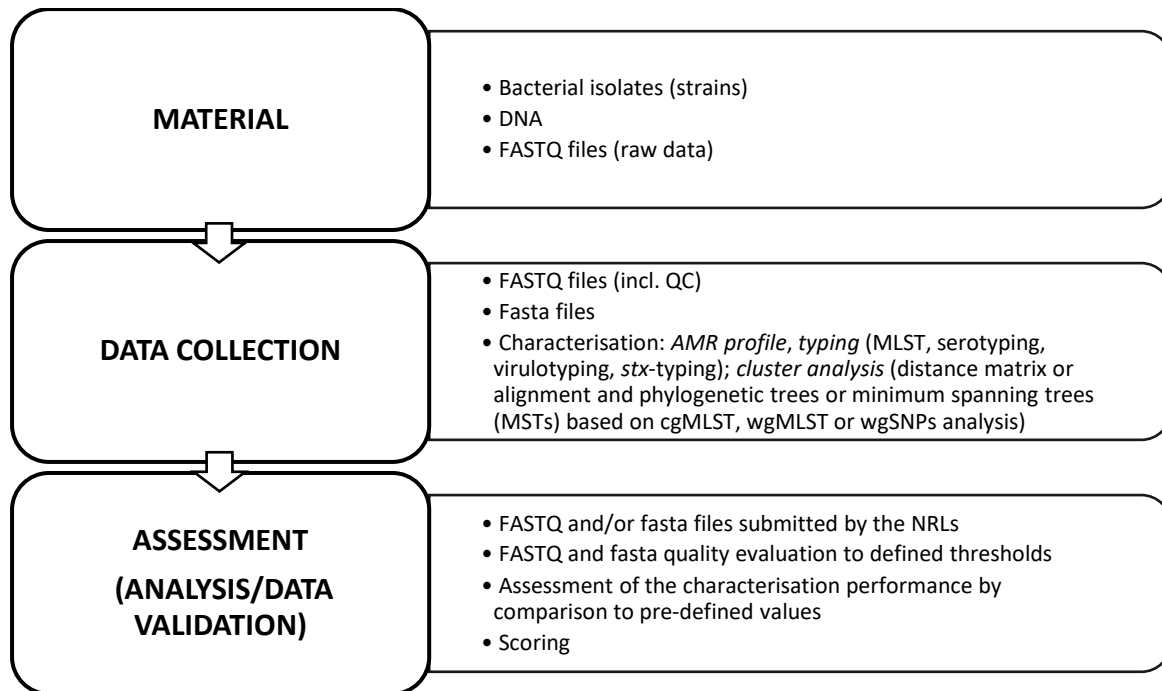- Scoring

Figure 1. Overall summary of the activities performed by the EURLs in the framework of NGS PTs organisation, including material distribution, data collection and data analysis.

In the following summary, the individual EURLs' PTs on NGS are presented in more detail.

EURL-Listeria monocytogenes

In 2016, EURL-*Listeria monocytogenes* (*Lm*) introduced in its PTs NGS analysis dedicated to molecular typing. Since 2022, the EURL-*Lm* PTs have been conducted under accreditation according to EN ISO 17043 and focusing on molecular serotyping and cgMLST (Pasteur scheme) for clustering. Generally, the EURL-*Lm* distributes four strains and six FASTQ files to their NRLs network. The laboratory collects quality assessment of the raw data, raw sequencing data (FASTQ files) with minimum depth of 20X, assemblies (fasta files) and distance matrix (in .csv and/or .tsv format) and a phylogenetic tree (newick file) for cluster determination. Additionally, EURL-*Lm* collects data on MLST clonal complexes (CCs) and serotypes. The cgMLST is assessed based on genomic DNA extraction, quality of the sequences, while the distance matrix is evaluated by the Mantel test in comparison with reference matrix distance. The results are considered satisfactory if the Mantel coefficient is at least equal to 0.70. The phylogenetic tree is evaluated by a cophenetic distance, and the

result is considered acceptable if the cophenetic correlation coefficient is at least equal to 0.70. Serotypes, CCs and clusters are determined by cgMLST method with a threshold of seven allelic differences (AD) (equal or less than seven AD). The results are satisfactory if the distance matrix, phylogenetic tree, serotype, CC and cluster determination comply with predefined results.

Plans for upcoming PTs: The pipeline established by EFSA will be used as a reference and assessment of the allelic profile for individual *L. monocytogenes* strains will be included and will allow limited number of strains to be sequenced.

EURL-VTEC

The activities of the EURL-VTEC are focused on STEC and other pathogenic foodborne or zoonotic *E. coli*. So far, the EURL-VTEC PTs have included shipment of six to eight STEC strains and collection of raw sequencing data (2017-2018) or characterisation results (2018-2022). For the 2017-2018 PT, the EURL-VTEC evaluated the quality of the sequences (assembly coverage, N50 and raw reads depth) and their effect on the WGS-based characterisation of STEC. The data analysis was performed at the EURL-VTEC using the ARIES Galaxy platform and include determination of MLST profile, serotype, Shiga toxin (*stx*) genes subtyping, virulotyping and cluster analysis using cgMLST, reference-free SNPs and wgSNPs. Since 2018, EURL-VTEC has been collecting characterisation results obtained either by conventional methods or WGS. The participants are asked to indicate which of the two options was used, however, no FASTQ files or details of the bioinformatics pipelines for data analysis are collected, thus only focusing on the interpretation of results. As a matter of fact, the participants are requested to report characterisation results in terms of virulence genes identification, serotyping, *stx* genes subtyping and cluster analysis. As for the latter, in 2018 (PT23), results obtained through PFGE were accepted, as well as NGS results obtained through cgMLST or wgSNPs. The latter two have been the only accepted methods since 2019 (PT26). In detail, the participants are requested to report the samples identified to be part of a cluster and the range of allelic or SNPs distances identified in the detected cluster. The results are collected through dedicated online forms built with Microsoft Forms application. This approach has been used for four rounds of PTs (PTs 26, 28, 31 and 35) and the corresponding

reports are available through the EURL-VTEC website, in the page dedicated to PTs: https://www.iss.it/en/vtec-proficiency-tests under subpage 'PTs on the identification and typing of pathogenic *E. coli* strains'.

Plans for upcoming PTs:

The approach in use in the last few years has proved valuable to assess the proficiency of the NRLs in the characterisation of pathogenic *E. coli* strains, regardless of the method used. Moreover, the choice to avoid the transfer of sequencing files allows saving time for both the NRLs and the EURL-VTEC. PTs on the identification and typing of pathogenic *E. coli* strains allowing the use of NGS as analytical method are routinely planned for the last months of each year (October – December). The next round (PT38) is foreseen for October 2023, with deadline set in December 2023.

EURL-Salmonella

EURL-*Salmonella* performs an annual "Serotyping PT" which has been mandatory for all MSs since 1995, and an optional "Cluster analysis PT" since 2019. Previously, EURL-*Salmonella* has also provided optional "AMR PTs" (2003-2005), "Phage typing PTs" (1999-2014), and "PFGE typing PTs" (2013-2018). For the annual "Cluster analysis PT", EURL-*Salmonella* distributes 10-12 strains or FASTQ files (including FASTQ files of a reference outbreak strain and technical duplicates). The NRLs choose their own routine method of typing (PFGE, MLVA and/or WGS) depending on the year of the PT. In the latest PTs (2021 and 2022), participants used WGS as a preferred typing method and performed either cgMLST or SNP-based analysis. The results were submitted in an online Result Form (formdesk application), one cluster analysis per strain. The raw reads were uploaded via sftp, and the distance matrix sent by e-mail. The EURL-*Salmonella* evaluated the participants raw sequencing data by in-house developed assembly pipelines. The cluster analysis was performed by the Ridom SeqSphere+ with the cgMLST Enterobase scheme. Participants were asked to report per strain if a clustering match was found with the reference outbreak strain. For example, in 2021 *S.* Enteritidis ST11, MLVA type 3-10-4-4-1 was used as a reference strain and the cluster definition was set at ≤7 allelic differences. The evaluation of the cluster analysis was performed by comparing the results submitted by the participants to pre-defined results by the EURL-*Salmonella*. In general, no

specific performance criteria are set on cluster analysis, however, the participants are expected to at least report technical replicates as part of the same cluster. Protocols and reports regarding the optional parts on NGS-based cluster analysis within the EURL-*Salmonella* PTs Typing (2019 – 2022) are available on the website, respectively at https://www.eurlsalmonella.eu/proficiency-testing/typing-studies and https://www.eurlsalmonella.eu/publications/proficiency-test-reports.

Plans for upcoming PTs: The annual EURL-*Salmonella* PT Typing (obligatory part on serotyping) takes place in November/December and will continue to have an optional part on cluster analysis. In 2023, the latter will be WGS-based only, since numbers of MLVA-participants dropped below significance. Therefore, *Salmonella* serovars not being Typhimurium or Enteritidis may now be used within cluster analysis as well.

EURL-Campylobacter

All three NGS PTs organised so far by the EURL-*Campylobacter* (2019, 2020 and 2022) have focused on sequence quality and two PTs (2019 and 2022) included additionally cluster analysis. The material has been *Campylobacter* isolates and/or DNA stocks. The EURL-*Campylobacter* has created complete gap-free assemblies for each strain used in the PTs. These have served as reference for the quality assessment.

The NRLs perform library preparations, sequencing and, when included, cluster analysis using their own routine methods and a cluster cut-off value for cluster analysis. The EURL-*Campylobacter* uses Questback Essentials surveys to collect results and detailed information on the NRLs methods to follow the process of implementation and to guide NRLs in need of improvement. The EURL also collects the generated raw sequence data as well as assemblies (if this step was part of the routine) for sequence quality assessment, and the phylogenetic tree or minimum spanning tree (MST) and raw sequencing data used to create the tree (distance matrix or alignment) for assessment of cluster analysis. These data are uploaded onto a personal cloud-based folder created by the EURL for each participant.

The assessed quality parameters of the raw sequence data have included proportion of: Q30 bases, contamination of non-target species, coverage over reference genome and GC

deviation in raw reads compared to the reference. Assembly-QC values (when data is assembled by the EURL) have also been used. In PT28 (2020), performance thresholds were based on median absolute deviation values, while in PT33 (2022) they were based on values from the ISO 23418. The assessment of cluster analysis in PT33 was based on the ability to reproduce major topological features of MSTs or phylogenetic trees when compared to the reference topology defined by the EURL. Each individual step is assessed as satisfactory or needs improvement and no overall performance grade has been given until now. In case of an outlying result, a root-cause analysis is performed by the EURL and communicated to the NRL in the individual NRL report. Reports from each PT on NGS can be found at the EURL-*Campylobacter* website (https://www.sva.se/en/about-us/eurl-campylobacter/proficiency-tests/).

Plans for upcoming PTs: The EURL-*Campylobacter* intends to continue to organise NGS PTs every other year. The upcoming NGS PT in 2024 will assess quality of both sequencing data from a small number of DNA samples and cluster analysis of a larger dataset provided as FASTQ files.

EURL-AR

EURL-AR distributes live cultures and corresponding pre-prepared DNA of six strains each year (e.g., 2021 two *Salmonella*, two *E. coli*, two *Campylobacter* and in 2022 two Enterococci, two *Staphylococcus aureus* and two *E. coli* strains) and collects FASTQ files from the participants for both types of samples. The PT provider (EURL-AR) assembles the FASTQ files using in-house pipeline and calculates quality metrics. For genomic characterisation, the EURL-AR collects results from analysis of chromosomal mutations, including *ampC* promoter mutations for *E. coli*, gene variant detection and predicted AMR profile (phenotype). Other information includes MLST and for *E. coli* and *Salmonella* optional serotyping. The assessed quality parameters of the assemblies include NG50, number of contigs, assembly size compared to a reference genome (completeness) and proportion of reads mapped to a reference.

Plans for upcoming PTs: In the upcoming PT (2023), the EURL-AR will include plasmid replicon typing as part of the data analysis. EURL-AR will also continue the established setup to offer discussions to their network on the obtained results recognising that the use of WGS data

have important limitations when considering their applicability for routine confirmatory, diagnostics or surveillance procedures. The EURL-AR will continue to compare bioinformatics approaches used and to benchmark those approaches with the purpose of discovering potential problems or differences between the available pipelines, and to identify local, national and European opportunities for improvement and harmonisation for analysis of WGS data.

EURL-CPS, EURL-Foodborne Viruses and EURL-Parasites

In the coming years, EURL-CPS aims at organising PTs on the NGS analysis of *Staphylococcus aureus* strains, previously isolated in the framework of staphylococcal food poisoning outbreaks. The PT will focus on the assessment of the quality of sequencing datasets produced from the strains, as well as of the results concerning enterotoxin gene content and MLST typing of assembled genomes.

The EURL-Foodborne Viruses has no plans for organising PTs in the next coming years. The EURL is currently working on a workflow for NGS typing of norovirus and will share the protocol with their NRLs in the beginning of next year.

In the future, the EURL-Parasites aims at organising a PT on the analysis of WGS data of protozoan genomes (*Cryptosporidium* and *Giardia*) focusing on the recognition of epidemiologically related cases (e.g., outbreaks).

Considerations and conclusions

NGS is a powerful technique that has already been established as the preferred method for bacterial typing and outbreak investigation in food microbiology, and concordantly, the majority of the EURLs of the inter-EURLs WG NGS are providing NGS PTs for their NRL networks. The majority of the PTs involving NGS focused on quality assessment of the sequence data and phylogenetic cluster analysis (except EURL-AR that did not include cluster analysis and EURL-VTEC that did not ask for sequencing files at all, basing the PTs on the interpretation of results by the NRLs) and have a genotyping approach included as well.

Other EURLs, i.e., EURL-CPS, EURL-Foodborne Viruses and EURL-Parasites, have not provided NGS PTs so far, however, EURL-CPS and EURL-Parasites plan future NGS PTs implementation, while EURL-Foodborne Viruses is currently preparing a NGS workflow to share with their NRL network. In this regard, it is considered important to exchange plans and experiences between the EURLs and to divulge the lessons learned to ease the work of NRLs intending to organise NGS PTs at a national level and to those EURLs without prior experience in NGS PTs implementation.

When evaluating existing PTs or planning new PTs, the Joint Integrative Project, EJP CARE, provide a "Guidance document for cross sectorial proficiency testing" (https://zenodo.org/record/7467622) and "SOPs for specific WGS proficiency testing distributions" (https://zenodo.org/record/7467902) may prove relevant for inspiration. Moreover, in 2022, ISO 23418 was published ("Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance"), which can be used as a guidance for future implementation/application of NGS in the laboratories.

Since the interest in participation in NGS PTs has already been shown to be large among several EURL/NRL networks, it is anticipated that the number of participants in NGS PTs - or NGS components of PTs - may increase as more laboratories are expected to implement NGS. Upcoming as part of the inter-EURLs WG NGS, the EURLs will identify synergies among the different EURLs NGS PTs and exchange experiences to take advantage of the lessons learnt in the set-up, evaluation procedures, and especially submission of sequences, data and evaluation analysis, to avoid unnecessary workloads for both the EURLs and the NRLs.

--- --- ---