EURL-*Campylobacter* Proficiency Test Report

PT 38. Whole genome sequencing and cluster analysis of *Campylobacter*

Publication history

Version	Date
Final version	2024-12-12
Final version 2	2024-12-20

SWEDISH VETERINARY AGENCY

EURL-*Campylobacter* Proficiency Test Report

PT 38. Whole genome sequencing and cluster analysis of *Campylobacter*.

Final version 2 edits Appendix A was corrected for columns "Megabases"

Authors Bo Segerman, Ásgeir Ástvaldsson and Hanna Skarin

Cover image Ida Olsson, Linda Svensson, EURL-*Campylobacter*, SVA, 2024

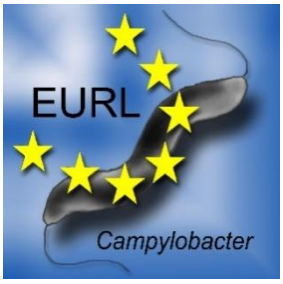
ISSN 2004-9188

Publication ID SVAKOM225.2.v2

Web https://www.sva.se

This publication may be cited "PT 38. Whole genome sequencing and cluster analysis of *Campylobacter*. EURL-*Campylobacter* Proficiency Test Report. SVA, 2024".

The European Commission officially designated the Swedish Veterinary Agency as the European Union reference laboratory (EURL) for *Campylobacter* on July 1st, 2006. The EURL regularly organises proficiency tests (PTs) for the national reference laboratories (NRLs) on methods of laboratory analysis for *Campylobacter* in different matrices of food or animal origin.



Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor HaDEA can be held responsible for them.

Summary

The EU reference laboratory for *Campylobacter* organised proficiency test (PT) number 38 on whole genome sequencing (WGS) and cluster analysis of *Campylobacter* March 2024. The PT included sequencing of provided DNA and sequence analysis of genomic sequence data provided as a pre-sequenced dataset. The objective was to assess the quality of WGS data and accuracy of quality control, species identification, Multi-Locus Sequence Type (MLST) determination and cluster detection of *Campylobacter* performed by participating laboratories.

Participation in PT 38 was voluntary for all NRLs. Twenty-four NRLs in 17 EU Member States (MS, some MS have more than one NRL) and in four non-EU countries received the PT, and responses were reported from all of them. One NRL only reported results on the sequence analysis part.

The individual parts (WGS and sequence analysis) were assessed through different criteria as satisfactory/needs improvement and no overall performance criteria was applied for this PT.

In summary, all but two of the NRLs met the criteria for satisfactory performance of WGS and 14 out of 24 NRLs met the criteria for satisfactory performance in sequence analysis. The most frequent issue was failure to exclude contaminated samples, but a few NRLs were also classified as needing improvement on species identification, MLST determination or cluster detection.

Contents

Abbreviations

C.	*Campylobacter*
EU	European Union
EURL	European Union reference laboratory
ISO	International Organization for Standardization
MLST	Multi-Locus Sequence Type
cgMLST	core genome MLST
wgMLST	whole genome MLST
NRL	national reference laboratory (in this report used for all participating laboratories, also in non-EU Member States)
PT	proficiency test
QC	quality control
SNP	single nucleotide polymorphism
SD	standard deviation
ST	sequence type
WGS	whole genome sequencing

Introduction

Proficiency test (PT) number 38 on whole genome sequencing (WGS) and cluster analysis of *Campylobacter* was organised by the EU reference laboratory (EURL) for *Campylobacter* in March 2024. Participation was voluntary. Twenty-four national reference laboratories (NRLs) in 17 EU Member States (MS, some MS have more than one NRL) and in four non-EU countries received the PT. All 24 NRLs reported the test results and operational details to the EURL. One NRL only reported results on the sequence analysis part.

Part 1 of the PT included to perform library preparations and sequencing of two DNA samples, and part 2 to perform quality control, identify species, determine Multi-Locus Sequence Type (MLST) and to perform cluster analysis of a dataset of 18 raw-data sequence samples. The NRLs were instructed to use their standard laboratory procedures for all parts of the analysis. Cluster analysis could be performed using SNP (single nucleotide polymorphism) analysis or gene-by-gene (wgMLST or cgMLST) analysis or other types of comparisons. The participants were instructed to use their own interpretation (cut-off value) of a cluster.

The objective was to assess the quality of WGS data and accuracy of the sequence analysis performed by the participating laboratories. The main purpose of the PT was to help laboratories to implement and evaluate their capacity of WGS and sequence analysis of *Campylobacter*.

TERMS AND DEFINITIONS

Only some selected terms are defined here. For additional definitions of terms used in this document, please see ISO 23418:2022 (Anonymous, 2022).

- Assembly: output from a process of aligning and merging sequencing reads into larger contiguous sequences (contigs).
- Coverage: number of times that a given base position is read in a sequencing run.
- Library: collection of genomic DNA fragments from a single isolate intended for determining genome sequence(s).
- N50: length (N) such that sequence contigs of N or longer include half the bases in the assembly.

Outline of the proficiency test

SELECTION OF STRAINS

One *C. jejuni* and one *C. coli* strain were selected for part 1 (WGS) of PT 38 (Table 1).

TABLE 1. Identity of the two DNA samples distributed to the NRLs in proficiency test No. 38, 2024.

Sample ID	Species	Sampling year	Sequence type (ST)	GC content (%)	Genome size
PT38-1	*Campylobacter jejuni*	2016	464	30.29	1.8 Mb
PT38-2	*Campylobacter coli*	2017	4709	31.18	1.8 Mb + 0.1 Mb plasmid

For part 2 (sequence analysis), 18 strains, including both *C. jejuni* and *C. coli* and representing several different sequence types (STs), were selected. The strains had been isolated at different timepoints and mostly from different farms and were sequenced with different Illumina instruments using different read lengths (Table 2). The selection was based on cgMLST and SNP analysis, aiming for a relevant challenge in cluster analysis and confirming similar topology with different comparison approaches. In the cgMLST analysis the PubMLST *C. jejuni* / *C. coli* cgMLST v1 (1343 loci) and v2 (1142 loci) (Cody et al., 2017; Jolley et al., 2018), Ridom cgMLST.org core (637 loci) and Ridom cgMLST.org core + accessory (1595 loci) and Innuendo cgMLST (678 loci) schemes were used (Rossi et al., 2018). The cgMLST was performed with both chewBBACA (Silva et al., 2018) and Ridom SeqSphere+ software and SNP analysis was made using Snippy (Seemann, 2015). The data also included samples with different types of quality issues that were expected to be identified and handled by the NRLs.

PREPARATION AND QUALITY OF THE SAMPLES

Preparation and quality control of DNA samples (for part 1)

Strains were cultivated on horse blood agar and overnight cultures were prepared using one or two colonies inoculated in Mueller-Hinton broth. The cultures were grown until OD600 values reached >0.7. The cultures were collected by centrifugation, washed with PBS and the pellets were frozen at -20 °C. Two or three pellets were prepared for each sample. The overnight cultures were checked for absence of contamination by cultivation on blood agar plates.

DNA was extracted using the Genomic Tip 20/G kit (Qiagen) according to the kit protocol, except that Ready-Lyse reagent (Biosearch Technologies) was used for cell lysis instead of lysosome. The concentration of the extracted DNA was measured using a Qubit 2.0 with a DNA BR kit (Thermo Fisher Scientific) and quality checked with a Nanodrop instrument. Multiple DNA solutions were pooled to generate a homogeneous stock solution. The stock was quantified and quality checked as described above.

For stabilisation, DNA was aliquoted into 0.5 ml screw cap tubes containing GenTegra DNA (GenTegra). The DNA was dried by leaving the cap off in a biosafety hood for at least 48 h. The tubes were then closed

TABLE 2. Identity of the 18 raw-data samples distributed to the NRLs in proficiency test No. 38, 2024.

Sample ID	Species	Location	Sampling time	Sequencer	Amount of data (Mbases)	% Q30+	Q30+ base coverage (X)	% reads with adapters ≥12 bp
PT38-3	*C. jejuni*	Farm A	Oct., 2020	NovaSeq (151+151)	455	90	250	19
PT38-4	*C. jejuni*	Farm B	Oct., 2020	MiSeq (251+251)	212	93	120	0.1
PT38-5	*C. jejuni*	Farm C	Sep., 2019	NovaSeq (151+151)	417	86	220	73
PT38-6	*C. jejuni*	Farm C	Sep., 2021	MiSeq (76+76)	149	95	91	0.003
PT38-7	*C. jejuni*	Farm D	Jun, 2021	NextSeq 500 (151+151)	366	89	223	0.14
PT38-8	*C. jejuni*	Farm E	Nov., 2020	NovaSeq (151+151)	438	93	268	29
PT38-9	*C. jejuni*	Farm D	Aug., 2018	MiSeq (76+76)	499	95	290	0.002
PT38-10	*C. jejuni*	Farm F	Jul., 2020	MiSeq (76+76)	198	90	121	0.008
PT38-11	*C. coli*	Farm G	Sep., 2021	NovaSeq (151+151)	400	90	230	36
PT38-12	*C. jejuni*	Farm C	Sep., 2021	NovaSeq (151+151)	423	88	256	73
PT38-13	*C. coli*	Farm G	Sep., 2021	NovaSeq (151+151)	485	90	280	56
PT38-14	*C. jejuni*	Farm C	Sep., 2019	NovaSeq (151+151)	413	85	253	66
PT38-15	*C. jejuni*	Farm H	Aug., 2017	NovaSeq (151+151)	88	13	7	
PT38-16	*C. jejuni*	Farm I	Sep., 2020	MiSeq (251+251)	250	93	152	0.1
PT38-17	*C. jejuni*	Farm J	Oct., 2020	MiSeq (251+251)	192	93	117	0.1
PT38-18	*C. jejuni*	Farm K	Oct., 2020	NovaSeq (151+151)	242	92	220	73
PT38-19	*C. jejuni*	Farm L	Aug., 2019	MiSeq (76+76)	124	95	91	0.03
PT38-20	*C. jejuni*	Farm C	Sep., 2021	NovaSeq (151+151)	450	90	275	45

and stored at room temperature. The expected yield for each tube was >500 ng.

The whole test, including reconstitution of DNA samples, DNA quantification and quality check, library preparation and sequencing was performed by the EURL before dispatch and one week after final date to report the results to control that the test was stable. The output of both DNA and sequence quality was satisfactory.

Reference genomes of the two strains were available from a previous PT (EURL-*Campylobacter* PT 28).

Preparation and quality of the raw-data files (for part 2)

The FASTQ files were generated from different types of Illumina machines and had different read lengths and coverage depths (Table 2). The files contained a variable amount of adapter sequences, which in all cases

were of Illumina Nextera type.

Three samples were in-silico modified to inflict major quality issues:

- The PT38-10 sample was *in-silico* contaminated with 10 % *Pseudomonas aeruginosa*.
- The PT38-15 sample was modified so that only 13 % of the bases were Q30+. The sample contained too little high-quality data to be useful.
- The PT38-19 sample was *in-silico* contaminated with 40 % *Salmonella enterica*.

In the other samples, the amount of Q30+ bases was acceptable and only traces or small amounts of contaminations were present (0-0.3 %). The contamination levels were considered acceptable but could for some samples negatively affect assembly quality values due to small low-coverage contigs generated in the assembly process. This would mainly be seen if excessive amount of data is used and/or short or low-coverage contigs are not filtered.

DISTRIBUTION OF THE PROFICIENCY TEST

The PT38-1 and PT38-2 samples were distributed from the EURL together with PT 36 on the 11th of March 2024. The samples were placed in foam boxes along with freezing blocks. The foam boxes were packed in cardboard boxes for transport and were sent from the EURL using courier service. One test was distributed by ordinary mail and was sent to an NRL that did not participate in PT 36.

Each participant received a plastic bag containing two numbered tubes, each containing stabilised and dried DNA from *Campylobacter*. A temperature logger was included in each package to record the temperature every second hour during transport.

Fifteen NRLs received the PT within one day after the packages had been dispatched from the EURL, seven NRLs within two days and one NRL received the package within four days. The one sent by ordinary mail was received within five weeks.

The samples were recommended to be stored at room temperature until start of analysis. Instructions for rehydration of each sample were included in the packages and were also sent out by e-mail a few days before the PT distribution.

The raw data files for the cluster analysis part were distributed in a OneDrive folder together with md5 checksums.

All results and information about the procedures had to be reported in the Questback Essentials system by 15th of May 2024. Supplementary data requested were: raw sequence files (i.e., FASTQ files), assembly files in FASTA format (only requested if assembling was part of the analysis), the tree used to draw conclusions from the analysis (e.g. phylogenetic tree or minimum spanning tree) and raw clustering data used to create the tree (distance matrix or alignment). Participants were instructed to upload requested files into a personal OneDrive folder. Each NRL was given a unique Lab ID number that was used as an identifier for reporting and uploading of sequence data. Lab ID has been shortened to L# in the text and figures of this report.

Methods for analysis of results

PART 1: WGS

Quantifications and quality control (QC) measurements were made with the tngs script (Segerman, 2022). Contamination levels were estimated using the Kraken2 software (Wood et al., 2019) with the 30GB standard database to obtain metagenomic information about the sequencing datasets. Kraken2 classifies reads as belonging to different phylogenetic taxa and this indicates if the correct species was sequenced and if the samples contained contaminating reads from a different organism. The percentage of *k*-mers present in the reference genomes that was covered by the read data submitted by participants was quantified using the tngs script (Segerman, 2022).

The quality of assemblies made from the participants' sequence data was measured with a pipeline using 80X Q30 data, default fastp trimming (Chen et al., 2018) and assembly with spades (--isolate option) (Bankevich et al., 2012). No post processing or filtering of the assemblies were made before QC metric quantifications. Assembly quality parameters analysed were:

- Total number of contigs
- Total size of assembly (bp)
- GC content
- Longest contig
- N50 length
- Percent of assembly in contigs with low coverage
- Percent targets called by chewBBACA with the PubMLST *C. jejuni* / *C. coli* cgMLST v2 scheme (1142 loci)

PART 2: SEQUENCE ANALYSIS

Data from the Questback Essentials system and uploaded data to the OneDrive folder describing detected contaminations, quality problems, excluded samples, species identification, MLST determination and identified clusters was collected and compiled into tables.

ASSESSING THE PERFORMANCE OF THE NRLS

Different criteria for the individual steps of each part (WGS and sequence analysis) were assessed. The results were 'Satisfactory' when all criteria were met for all the samples, whereas failure to reach one or more criteria for one or more samples was marked as 'Needs improvement'. No overall performance grade was applied for this PT. Overall comments on the data and possible focus areas for improving performance were commented further in each laboratory's individual report.

Assessment of sequence quality (part 1)

Cut-off values were defined for six different criteria to assess the sequence quality through submitted FASTQ files for each sample (Table 3). The criteria were: total amount of data, percentage of Q30+ bases, percentage of contaminating reads, percentage coverage of the corresponding reference genome, and percentage GC-deviation in the sequence reads from the corresponding reference genome and the percentage of targets found when reads were assembled and analysed with cgMLST.

TABLE 3. Overview of the criteria and cut-off values used for assessment of sequence quality in proficiency test No. 38 (2024).

Criteria	Cut-off value for satisfactory performance
Total amount of data	>30X or 80X depending on library preparation kit (80X for Nextera XT)
Q30+	>70 %, 75 % or 80 % depending on read length (300, 250, 150-100 bp)
Contamination	<5 % from non-target species
Reference coverage	>98 % of reference genome[a]
GC deviation	<4 % deviation from reference genomes
Assembly targets	>95 % of targets found

[a]The maximum amount of data used for the assessment was 80X coverage for NRLs using Nextera XT and 30X coverage for NRLs using other library preparation kits.

Assessment of sequence analysis (part 2)

Cut-off values were defined for four different criteria, which were all assessed separately (Table 4). The criteria were ‘evaluation of sequence quality’, ‘species identification’, ‘MLST determination’ and ‘cluster detection’.

TABLE 4. Overview of the criteria and cut-off values used for assessment of sequence analysis in proficiency test No. 38 (2024).

Criteria	Cut-off value for satisfactory performance
Evaluation of sequence quality	Identify and exclude (or ‘clean up’) sample PT38-10, PT38-15 and PT38-19
Species identification	All samples analysed[a] should be correctly species identified
MLST determination	All samples analysed[a] should be designated with correct ST
Cluster detection	Cluster A (or AB), C and D should be identified

[a]Samples PT38-10, PT38-15, PT38-19 excluded from the assessment.

Results

Proficiency test number 38 was distributed to 24 NRLs and 23 of them reported results from part 1 (WGS) and all reported results from part 2 (sequence analysis).

The analysis was started at different timepoints between March and mid-April 2024. The sequence quality measures were calculated from the raw data submitted by participants. Twenty out of 23 NRLs submitted data where adapters already had been removed. A summary of all sequence quality measures in each submitted dataset can be found in Appendix A.

PART 1: WGS

DNA library preparation and sequencing

To estimate the DNA concentration of the samples, 16 NRLs used Qubit, two used Thermo Fisher Nanodrop, one used both Qubit and Nanodrop, one used Agilent BioAnalyzer, one used BMG ClarioStar, one used GloMax Discover, and one used Quant-iT dsDNA Assay Kit with a Tecan Infinite 200 Pro microplate reader.

DNA quality control was performed by 15 NRLs; nine used Thermo Fisher Nanodrop, two used DeNovix spectrophotometer, one used Agilent TapeStation, one used Agarose gel electrophoresis, one used nanodrop and Agilent Fragment Analyzer, and one used IMPLEN NanoPhotometer P-Class P330-30.

For library preparations, 15 NRLs used the Illumina DNA Prep kit (previously known as Nextera DNA Flex Library Preparation kit), three used the Illumina Nextera XT, one used the Illumina TrueSeq DNA Nano/PCR-Free, one used the Invitrogen Collibri ES DNA Library Prep, one used the Roche KAPA HyperPlus Library Prepp kit and one used the Qiagen QIAseq / GeneRead kit. One NRL did not know which kit was used (outsourced). Four NRL using DNA Prep kit and one using Nextera XT used 1/2 or 1/2.5 of the volume of reagents recommended by the manufacturer.

For quantification of the library preparations, 14 NRLs used Qubit, two used Quantitative real time PCR, one used Agilent Tapestation, one used BMG ClarioStar, one used Glomax with Quant-It high sensitivity kit, and one used Quant-iT dsDNA Assay Kit with a Tecan Infinite 200 Pro microplate reader. Two NRLs did not know (outsourced) and one did not perform quantification.

Library quality control was performed by 18 NRLs; four used Agilent BioAnalyzer, six used Agilent TapeStation, one used Invitrogen Qubit, one used Quantitative real time PCR, two used Agilent Fragment Analyzer, and two used capillary electrophoresis. Two NRLs did not know (outsourced).

All the participating NRLs used Illumina technology for sequencing. Fourteen NRLs used Illumina MiSeq for sequencing, seven used NextSeq, one used MiniSeq, and one used NovaSeq. The read lengths were: 2×75 (2 NRLs), 2×100 (2 NRLs), 2×150 (12 NRLs), 2×250 (3 NRLs) and 2×300 (4 NRLs). The targeted theoretical coverage (e.g. minimum sequencing depth) reported by the NRLs were: 20X-40X (6 NRLs), 40X-60X (8 NRLs), 60X-80X (3 NRLs), 80X-100X (3 NRLs) and >100X (3 NRLs). The three NRLs using Nextera XT

for library preparation aimed for 20X-40X, 40X-60X or 80X-100X, respectively, in theoretical coverage.

Sample mix-ups

The submitted data was checked for mix-up mistakes. One NRL (L183) had submitted sample P38-1 as PT38-2 and vice versa. Another NRL (L144) had submitted the R2 reads of sample PT38-1 as the R2 reads of sample PT38-2 and vice versa but had not mixed-up the R1 reads. These mistakes were noted and corrected before further analysis.

Total amount of data: sequence coverage

A minimum coverage threshold of 30X was applied for unbiased library preps (all except Nextera XT) and 80X for library preps generated by the Nextera XT, which has a GC-dependent coverage bias and therefore requires higher coverage for similar performance (Segerman et al., 2022). All NRLs submitted data with coverage over the expected threshold (Figure 1A). The QC values are listed in Appendix A.

Quantification of high-quality bases

The percentage of bases with at least a quality score Q30 was calculated. The minimum quality threshold was set depending on the number of cycles sequenced (read length). Short read lengths (2×75, 2×100 and 2×150 bp) were expected to have at least 80 % Q30 bases, 2×250 bp were expected to have at least 75 % Q30 bases and 2×300 bp at least 70 % Q30 bases [1]. Two NRLs (L107 and L178) produced data of quality below the threshold (Figure 1B). The QC values are listed in Appendix A.

Deviation from expected GC content

The GC-content of each reference genome sequence was used as an expected value. Deviation from this value is seen in Nextera XT data because of GC-dependent coverage bias but it can also arise by large number of contaminating reads from a species with a different GC-content (Segerman et al., 2022). The cut-off value applied for this criterion was a deviation larger than 4 % (Anonymous, 2022). All NRLs were below this threshold (Figure 1C). The QC values are listed in Appendix A.

Sequence contamination

Contamination levels were estimated using the Kraken2 software using the 30 GB standard database. The threshold was set to 5 % reads from a non-target species (Anonymous, 2022). All NRLs had low contamination levels (<0.5 %) for both samples (Figure 1D). The QC values are listed in Appendix A.

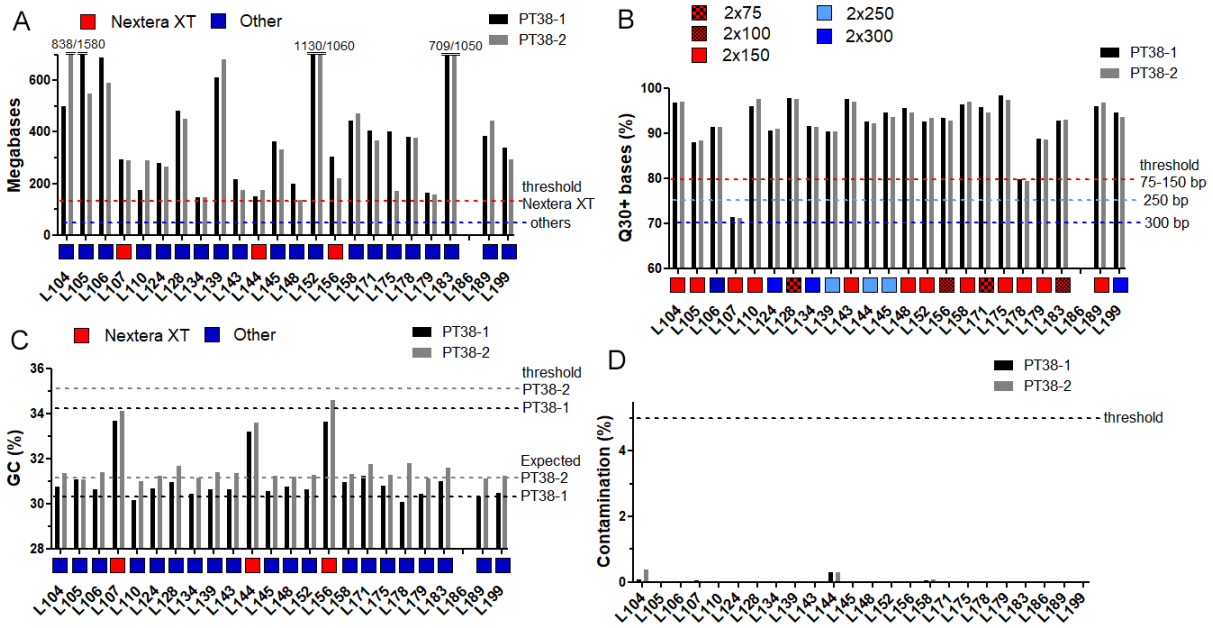


FIGURE 1. **A:** Total amount of data submitted by the NRLs. Thresholds were set at 30X for unbiased library preps (Blue) and 80x for Nextera XT preps (Red). **B:** Percentage of Q30+ bases in the WGS data submitted by the NRLs. Thresholds are set based on the read length used. **C:** Deviation of the GC content in the reads from the expected GC content (the GC content of the reference genome). **D:** Contamination levels estimated by the Kraken2 software using the 30 GB standard database.

Coverage of reference genome *k*-mers

The percentage of *k*-mers present in the reference genomes that was found in the raw data was quantified using 30X of raw data (Figure 2A) and 80X of raw data (Figure 4B). The threshold was set to 98 % reference coverage, but the NRLs using Nextera XT were evaluated at 80X and the NRLs using unbiased library preps at 30X of data. All NRLs reached the expected coverage of reference genome k-mers.

Assemblies

The QC metrics for each assembly is summarised in Appendix B. The NRLs that used Nextera XT (L107, L144, L156) got slightly inferior assembly QC values. The NRL with the low-level contamination (L144) got a slightly elevated assembly size and a larger amount of sequence data in low-coverage contigs. The acceptance threshold for the allele calling was set at 95 % targets found. All NRLs reached over that threshold.

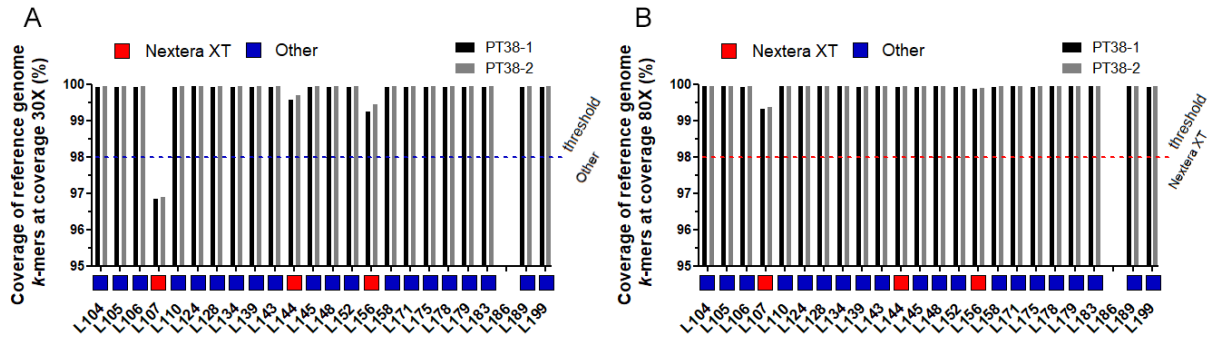


FIGURE 2. **A:** Coverage of the reference genome (k-mers) in the read data at 30X sequencing depth. The threshold applies for non-Nextera XT library kits. **B:** Coverage of the reference genome (k-mers) in the read data at 80X sequencing depth. The threshold applies for Nextera XT library kits.

TABLE 5. Overview of assessment of the sequence quality of each NRL in proficiency test No. 38 (2024). The number indicate number of samples out of two reaching the criteria cut-offs.

Lab ID	Amount of data	Q30+	Contamination	Reference coverage	GC deviation	Assembly targets	Overall evaluation sequence quality
L104	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L105	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L106	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L107[a]	2/2	0/2	2/2	2/2	2/2	2/2	Needs improvement
L110	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L124	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L128	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L134[a]	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L139	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L143	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L144	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L145	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L148	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L152[a]	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L156[a]	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L158	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L171	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L175	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L178	2/2	0/2	2/2	2/2	2/2	2/2	Needs improvement
L179	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L183	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L189	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory
L199	2/2	2/2	2/2	2/2	2/2	2/2	Satisfactory

[a]non-EU country

Assessment of NRL performance

The results using the defined criteria for assessment of sequence quality of each NRL is summarised in Table 5. According to the assessment, 21 NRLs fulfilled the criteria for satisfactory performance and two NRLs scored below the criteria.

PART 2: SEQUENCE ANALYSIS

Evaluation of sequence quality

To evaluate the sequence quality, 15 NRLs used FastQC, three used Ridom SeqSphere+, one used Fastp, one used BioNumerics, one used FastQC+SeqKit+Quast, one used the INNUca pipeline, one used FastQC, Qualimap and MultiQC and one used Quast, Qualimap and ConFinder. To evaluate contamination levels, ten NRLs used Kraken2, four used Ridom SeqSphere+, three used ConFindr, one used Kraken2 and ConFindr, one used Kraken2 and QUAST, one used Kraken and KmerID (in-house), one used BioNumerics, one used bbmap, sendsketch and fileSize, one used PubMLST, and one did not do contamination check.

Two samples had been in-silico contaminated, PT38-10 with 10 % *Pseudomonas aeruginosa* and PT38-19 with 40 % *Salmonella enterica*. The participants were expected to identify these samples as contaminated and exclude them or effectively remove the contamination. Eighteen out of 24 NRLs excluded PT38-10 and 22 out of 24 excluded PT30-19.

One sample had been modified to have low quality values. The participants were expected to identify this sample and exclude it. All NRLs excluded this low-quality sample.

The participants were not expected to exclude any other samples. Seven out of 24 participants excluded additional samples with variable explanations. L105 performed cluster analysis on the *C. jejuni* isolates only. L199 categorized many samples as being cross contaminated between *C. coli* and *C. jejuni*. The remaining NRLs that excluded samples did it mostly based on assembly quality criterions. Variations in assembly quality values could be related to the use of the wrong adapter file during trimming and/or use of excessive amount of data without filtering low-coverage contigs. Three NRLs (L124, L148 and L156) excluded PT38-16 because of bad assembly statistics. Several other NRLs that did not exclude this sample had contaminations of short *Bifidobacterium* contigs in their assemblies. *Bifidobacterium* is a small genome that assembles well if excess data is used or low-coverage contigs are not filtered. One NRL (L148) found too few targets (<97 %) in the PT38-5 assembly, and one NRL (L183) failed to assemble PT38-20 for unknown reasons.

TABLE 6. Overview of results from the participants' evaluation of sequence quality in Part 2 of proficiency test No. 38 (2024).

Lab ID	Excluded PT38-10 (10 % Pseudomonas)	Excluded PT38-19 (40 % Salmonella)	Excluded PT38-15 (low QC score)	Number of other exclusions
L104	Yes[a]	Yes[a]	Yes	0
L105	No	Yes	Yes	2
L106	Yes	Yes	Yes	0
L107	Yes	Yes	Yes	0
L110	No	No	Yes	0
L124	Yes	Yes	Yes	1
L128	Yes	Yes	Yes	0
L134	No	Yes	Yes	0
L139	No	No	Yes	0
L143	Yes	Yes	Yes	3
L144	Yes	Yes	Yes	0
L145	Yes	Yes	Yes	0
L148	Yes[a]	Yes	Yes	2
L152	No	Yes	Yes	0
L156	Yes	Yes	Yes	1
L158	Yes	Yes	Yes	0
L171	Yes	Yes	Yes	0
L175	Yes	Yes	Yes	0
L178	Yes	Yes	Yes	0
L179	Yes	Yes	Yes	0
L183	Yes	Yes	Yes	1
L186	Yes[a]	Yes	Yes	0
L189	No	Yes	Yes	0
L199	Yes	Yes	Yes	8

[a]Sample excluded from the analysis in the supplementary data, but exclusion was not reported in Questback.

Species identification and MLST determination

To identify the species, nine NRLs used Kraken2, four used pubMLST, four used Ridom SeqSphere+, two used kmerFinder, one used CGE at DTU, one used in-house KmerID, one used Kraken2 and kmerFinder, one used SpeciesFinder and one did not perform species identification.

Twenty-two of 24 NRLs identified the correct species in all samples. One NRL (L128) did not perform species identification and one NRL (L189) identified PT38-7 as "*Campylobacter* sp. CFSAN093227". Species identification was not always reported on excluded samples.

To determine ST, eight NRLs used mlst (T. Semann), eight used Ridom SeqSphere+, three used Galaxy, one used BioNumerics, one used MOST, two used pubMLST, and one used Ridom SeqSphere+ and BakCharak.

Twenty-one NRLs correctly determined ST from the samples they included in the analysis (Table 7). One NRL (L183) did not specify ST for PT38-11 and PT38-13 (the *C. coli* samples) but gave the correct allelic numbers. One NRL (L107) could only identify five out of seven alleles in PT38-16. One NRL (L189) reported wrong ST for PT38-20, but the correct ST was specified in the supplementary uploaded data. One NRL (L128) reported wrong ST for five samples because clonal complexes (CC) were given instead of STs.

TABLE 7. Overview of results from the MLST determination in Part 2 of proficiency test No. 38 (2024).

Lab ID	PT38 -3	PT38 -4	PT38 -5	PT38 -6	PT38 -7	PT38 -8	PT38 -9	PT38 -10	PT38 -11	PT38 -12	PT38 -13	PT38 -14	PT38 -15	PT38 -16	PT38 -17	PT38 -18	PT38 -19	PT38 -20
L104	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L105	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L106	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L107	257	257	257	148	257	257	257		854	148	854	257		ND[a]	257	257		148
L110	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L124	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L128	257	257	257	21	257	257	257		828	21	828	257		257	257	257		21
L134	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L139	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L143	257	257	EXCL	148	257	257	257		854	EXCL	854	EXCL		257	257	257		148
L144	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L145	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L148	257	257	EXCL	148	257	257	257		854	148	854	257		EXCL	257	257		148
L152	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L156	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L158	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L171	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L175	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L178	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L179	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L183	257	257	257	148	257	257	257		ND[b]	148	ND[b]	257		257	257	257		EXCL
L186	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148
L189	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148[c]
L199	257	257	257	148	257	257	257		854	148	854	257		257	257	257		148

[a] Missing alleles.

[b] All seven alleles were correctly identified, but ST was not determined.

[c] Wrongly reported in Questback, but correctly determined in the supplementary uploaded data.

	PT38-3	PT38-4	PT38-8	PT38-16	PT38-18	PT38-17	PT38-9	*PT38-14*	*PT38-5*	PT38-7	PT38-6	*PT38-12*	*PT38-20*	*PT38-11*	*PT38-13*	PT38-15	PT38-10	PT38-19	cut-off	schema	software
L158	A	A	A	A	A	A	A	A	A	A	C	C	C	D	D	EXCL	EXCL	EXCL	NA	PubMLST v1	cgMLSTFinder
L107	A	A	A	A	A	A	A	A	A	A	C	C	C	D	D	EXCL	EXCL	EXCL	NA	NA	snippy
L199	A	A	A	A	A	A	EXCL	EXCL	EXCL	EXCL	NO	EXCL	EXCL	EXCL	EXCL	EXCL	EXCL	EXCL	7 AD	PubMLST v1	ChewBBACA
L106	A	A	A	A	A	A	A	A	A	NO	C	C	C	D	D	EXCL	EXCL	EXCL	13 AD	Ridom core	Ridom SeqSphere
L183	A	A	A	A	A	A	A	A	A	NO	C	C	EXCL	D	D	EXCL	EXCL	EXCL	13 AD	Ridom core	Ridom SeqSphere
L152	A	A	A	A	A	A	A	A	A	NO	C	C	C	D	D	EXCL	C	EXCL	13 AD	Ridom core	Ridom SeqSphere
L189	A	A	A	A	A	A	A	A	A	NO	C	C	C	D	D	EXCL	NO	EXCL	7-10 AD	Ridom core	Ridom SeqSphere
L186	A	A	A	A	A	A	A	A	A	NO	C	C	C	D	D	EXCL	EXCL	EXCL	13 AD	Ridom core	Ridom SeqSphere
L143	A	A	A	A	A	A	A	EXCL	EXCL	NO	C	EXCL	C	D	D	EXCL	EXCL	EXCL	6 SNPs /6 AD	Innuendo core	ChewBBACA/in house SNP
L145	A	A	A	A	A	A	A	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	0.5% AD	Innuendo WG	ChewBBACA
L178	A	A	A	A	A	A	A	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	10 AD	PubMLST v1	ChewBBACA
L139	A	A	A	A	A	A	A	B	B	NO	C	C	C	D	D	EXCL	NO	NO	13 AD	PubMLST v1	Ridom SeqSphere
L156	A	A	A	EXCL	A	A	A	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	10 SNPs	NA	SnapperDB
L105	A	A	A	A	A	A	A	B	B	NO	C	C	C	EXCL	EXCL	EXCL	NO	EXCL	10 AD	Innuendo core	ChewBBACA
L148	A	A	A	EXCL	A	A	A	NO	EXCL	NO	C	C	C	D	D	EXCL	EXCL	EXCL	10 AD	PubMLST v1	Ridom SeqSphere
L124	A	A	A	EXCL	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	5 AD	PubMLST v1	Ridom SeqSphere
L128	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	5 AD	PubMLST v1	BioNumerics
L179	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	10 AD	PubMLST v1	Ridom SeqSphere
L175	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	10 AD	PubMLST v1	Ridom SeqSphere
L110	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	C	NO	NA	NA	Samtools SNP Phylogeny
L134	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	NO	EXCL	14 AD	PubMLST v1	ChewBBACA
L104	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	10 SNPs / 31 AD	Innuendo WG	CSI Phylogeny
L171	A	A	A	A	A	A	NO	B	B	NO	C	C	C	D	D	EXCL	EXCL	EXCL	14 AD	Ridom core+Acc	Ridom SeqSphere
L144	A	A	A	A	A	NO	NO	NO	NO	NO	C	NO	C	D	D	EXCL	EXCL	EXCL	5 AD	PubMLST v1	in house

FIGURE 3. Depiction of the different clusters identified by the NRLs

Cluster detection

Most NRLs agreed on a common larger cluster ‘A’ with PT38-3, PT38-4, PT38-8, PT38-16, PT38-18 and PT38-17. Some NRLs included also PT38-9 in this cluster. Many NRLs identified a second closely related cluster ‘B’ including PT38-14 and PT38-5. Some NRLs merged this cluster with cluster ‘A’. Those who did this used the comparatively small Ridom core scheme or did not use a cut-off value. Most NRLs also defined a third cluster ‘C’ including PT38-6, PT38-12 and PT38-20, and a fourth cluster ‘D’ including PT38-11 and PT38-13. NRLs were expected to identify cluster ‘A’ (or ‘AB’), ‘C’ and ‘D’. No exact boundary defining the clusters was defined, but the clusters identified by the NRLs are depicted in Figure 3 so that the NRLs can compare their cluster definition stringency to that of the other NRLs. Two NRLs failed to detect the expected clusters because of too many excluded samples (L105 and L199).

Assessment of NRL performance

The results using the defined criteria for assessment of ‘evaluation of sequence quality’, species identification, MLST determination and cluster detection of each NRL is summarised in Table 8. According to the assessment, 14 NRLs fulfilled the criteria for satisfactory performance and 10 NRLs scored below the criteria.

TABLE 8. Overview of assessment of the sequence analysis of each NRL in proficiency test No. 38 (2024). The number indicate number of samples out of analysed samples reaching the criteria cut-offs.

Lab ID	Evaluation of sequence quality	Species identification	MLST determination	Cluster detection	Overall evaluation sequence quality
L104	3/3	15/15	15/15	AB, C and D	Satisfactory
L105	2/3	15/15	15/15	AB and C	Needs improvement
L106	3/3	15/15	15/15	A, C and D	Satisfactory
L107[a]	3/3	15/15	14/15	A, C and D	Needs improvement
L110	1/3	15/15	15/15	AB, C and D	Needs improvement
L124	3/3	15/15	15/15	AB, C and D	Satisfactory
L128	3/3	Not performed	10/15	AB, C and D	Needs improvement
L134[a]	2/3	15/15	15/15	AB, C and D	Needs improvement
L139	1/3	15/15	15/15	AB, C and D	Needs improvement
L143	3/3	15/15	12/12	A, C and D	Satisfactory
L144	3/3	15/15	15/15	A, C and D	Satisfactory
L145	3/3	15/15	15/15	AB, C and D	Satisfactory
L148	3/3	13/13	13/13	A, C and D	Satisfactory
L152[a]	2/3	15/15	15/15	A, C and D	Needs improvement
L156[a]	3/3	15/15	15/15	AB, C and D	Satisfactory
L158	3/3	15/15	15/15	A, C and D	Satisfactory
L171	3/3	15/15	15/15	AB, C and D	Satisfactory
L175	3/3	15/15	15/15	AB, C and D	Satisfactory
L178	3/3	15/15	15/15	AB, C and D	Satisfactory
L179	3/3	15/15	15/15	AB, C and D	Satisfactory
L183	3/3	15/15	12/14	A, C and D	Needs improvement
L186[a]	3/3	15/15	15/15	A, C and D	Satisfactory
L189	2/3	14/15	15/15	A, C and D	Needs improvement
L199	3/3	15/15	15/15	A	Needs improvement

[a]non-EU country

References

Anonymous (2022). ISO 23418:2022 - Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of bacteria — General requirements and guidance. 1st ed. *International Organization for Standardization (ISO)*.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19, 455–477. doi: 10.1089/cmb.2012.0021

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cody, A. J., Bray, J. E., Jolley, K. A., McCarthy, N. D., and Maiden, M. C. J. (2017). Core Genome Multilocus Sequence Typing Scheme for Stable, Comparative Analyses of *Campylobacter jejuni* and *C. coli* Human Disease Isolates. *Journal of Clinical Microbiology* 55, 2086–2097. doi: 10.1128/JCM.00080-17

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 3, 124. doi: 10.12688/wellcomeopenres.14826.1

Rossi, M., Silva, M. S. D., Ribeiro-Gonçalves, B. F., Silva, D. N., Machado, M. P., Oleastro, M., et al. (2018). INNUENDO whole genome and core genome MLST schemas and datasets for *Campylobacter jejuni*. doi: 10.5281/zenodo.1322564

Seemann, T. (2015). Snippy: fast bacterial variant calling from NGS reads https://github.com/tseemann/snippy. Available at: https://github.com/tseemann/snippy

Segerman, B. (2022). tngs. Available at: https://github.com/BoSegerman/tngs

Segerman, B., Ástvaldsson, Á., Mustafa, L., Skarin, J., and Skarin, H. (2022). The efficiency of Nextera XT tagmentation depends on G and C bases in the binding motif leading to uneven coverage in bacterial species with low and neutral GC-content. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.944770

Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microbial Genomics* 4. doi: 10.1099/mgen.0.000166

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology* 20, 257. doi: 10.1186/s13059-019-1891-0

Appendix A – QC metrics for sequence read data

Lab ID	Megabases PT38-1	Megabases PT38-2	Q30+ (%) PT38-1	Q30+ (%) PT38-2	GC (%) PT38-1	GC (%) PT38-2	Cont. (%) PT38-1	Cont. (%) PT38-2
104	499	838	96.8	96.8	30.73	31.36	0.07	0.37
105	1580	547	87.9	88.4	31.07	31.05	0.02	0.01
106	688	592	91.4	91.2	30.64	31.37	0.02	0
107	292	289	71.4	71.2	33.68	34.11	0.04	0
110	174	291	95.9	97.5	30.16	30.98	0.02	0
124	278	264	90.4	90.9	30.66	31.22	0.02	0.01
128	483	450	97.8	97.6	30.96	31.68	0.01	0
134	145	146	91.6	91.4	30.42	31.13	0.03	0
139	611	681	90.4	90.3	30.62	31.37	0.02	0
143	217	173	97.5	97.0	30.63	31.33	0.02	0
144	149	173	92.5	92.1	33.19	33.60	0.28	0.28
145	363	332	94.6	93.6	30.53	31.22	0.02	0
148	198	135	95.6	94.6	30.76	31.17	0.01	0
152	1130	1060	92.5	93.4	30.61	31.27	0.02	0
156	305	221	93.3	92.8	33.61	34.57	0.02	0
158	445	473	96.2	96.9	30.96	31.32	0.04	0.07
171	405	367	95.8	94.5	31.21	31.76	0.01	0
175	401	169	98.3	97.4	30.78	31.28	0.02	0
178	382	376	79.6	79.3	30.06	31.78	0.02	0
179	165	155	88.7	88.5	30.44	31.09	0.02	0
183	709	1050	92.8	92.9	30.98	31.60	0.01	0
189	384	443	96.0	96.7	30.29	31.09	0.01	0
199	340	293	94.6	93.6	30.46	31.24	0.03	0

Appendix B – QC metrics for assemblies (made by EURL)

Deviations from the mean with 2-3 Standard deviations (SD) are marked with light colour and more than 3 with dark colour

Sample ID	Total number of contigs	Total size of assembly (bp)	GC content (%)	Longest contig (bp)	N50 length (bp)	% of assembly in contigs with low coverage	% targets called[a]
PT38-1-L104	66	1,746,200	30.2	287,185	174,979	0.15	99.2
PT38-1-L105	87	1,751,125	30.2	287,185	154,617	0.38	99.2
PT38-1-L106	63	1,748,852	30.2	287,285	176,918	0.10	99.2
PT38-1-L107	173	1,736,127	30.3	106,388	39,389	1.47	98.5
PT38-1-L110	68	1,746,619	30.2	287,185	154,617	0.17	99.2
PT38-1-L124	54	1,748,757	30.2	287,285	176,918	0.13	99.2
PT38-1-L128	76	1,741,862	30.2	287,141	154,893	0.00	99.2
PT38-1-L134	46	1,749,480	30.2	287,285	176,918	0.18	99.2
PT38-1-L139	58	1,749,120	30.2	287,285	176,918	0.18	99.2
PT38-1-L143	55	1,744,232	30.2	287,185	154,937	0.03	99.2
PT38-1-L144	215	1,814,095	31.2	210,963	118,171	3.73	99.2
PT38-1-L145	54	1,748,483	30.2	287,306	176,949	0.12	99.2
PT38-1-L148	81	1,744,758	30.2	287,184	176,818	0.06	99.2
PT38-1-L152	58	1,743,812	30.2	287,185	154,937	0.04	99.2
PT38-1-L156	143	1,790,625	30.2	202,633	68,329	0.05	97.2
PT38-1-L158	62	1,750,390	30.2	287,185	154,937	0.36	99.2
PT38-1-L171	85	1,740,777	30.2	287,134	154,893	0.00	99.2
PT38-1-L175	76	1,744,267	30.2	287,185	175,811	0.04	99.2
PT38-1-L178	73	1,746,623	30.2	287,185	175,811	0.13	99.2
PT38-1-L179	81	1,742,104	30.2	287,141	154,893	0.05	99.2
PT38-1-L183	87	1,741,071	30.2	287,141	154,893	0.00	99.2
PT38-1-L189	77	1,745,741	30.2	287,185	154,937	0.12	99.2
PT38-1-L199	51	1,746,928	30.2	287,285	176,918	0.05	99.2
PT38-2-L104	71	1,796,704	31.1	352,379	203,691	0.2549	99.5
PT38-2-L105	87	1,791,751	31.1	352,916	203,731	0.3581	99.5
PT38-2-L106	62	1,798,202	31.1	391,185	203,846	0.0994	99.5
PT38-2-L107	202	1,789,958	31.2	79,699	37,100	1.1619	98.5
PT38-2-L110	84	1,795,328	31.1	352,379	203,691	0.2137	99.5
PT38-2-L124	55	1,799,028	31.1	391,185	203,846	0.1759	99.5
PT38-2-L128	88	1,789,244	31.1	352,335	203,647	0.0000	99.5
PT38-2-L134	59	1,803,553	31.1	391,185	203,846	0.4232	99.5
PT38-2-L139	68	1,802,314	31.1	391,185	203,846	0.3020	99.5
PT38-2-L143	72	1,793,407	31.1	352,379	203,746	0.0790	99.5
PT38-2-L144	206	1,849,727	32.0	279,007	130,492	3.3368	99.2
PT38-2-L145	61	1,800,018	31.1	391,185	203,846	0.2289	99.5
PT38-2-L148	83	1,795,422	31.1	352,379	203,691	0.1948	99.5
PT38-2-L152	80	1,795,099	31.1	352,379	203,691	0.2012	99.5
PT38-2-L156	134	1,789,360	31.2	278,660	130,348	0.0968	99.5
PT38-2-L158	68	1,798,221	31.2	429,360	212,632	0.3286	99.5

PT38-2-L171	90	1,790,007	31.1	352,335	203,647	0.0127	99.5
PT38-2-L175	66	1,792,241	31.1	352,379	203,691	0.0411	99.5
PT38-2-L178	90	1,797,017	31.1	352,651	203,691	0.3061	99.5
PT38-2-L179	100	1,792,151	31.1	352,335	203,647	0.1476	99.5
PT38-2-L183	101	1,789,810	31.1	381,476	203,647	0.0000	99.5
PT38-2-L189	85	1,795,851	31.1	352,379	203,691	0.2119	99.5
PT38-2-L199	58	1,799,612	31.1	352,829	203,691	0.2059	99.5

[a]Targets were called with chewBBACA using the PubMLST *C. jejuni* / *C. coli* cgMLST v2 scheme (1142 loci)

The Swedish Veterinary Agency (SVA) is an expert authority with contingency missions. SVA promotes animal and human health, Swedish animal husbandry and our environment through diagnostics, research, preparedness, and advice. The authority is under the Ministry of Rural Affairs and Infrastructure.

Visitor address: Ulls väg 2B, postal address: 751 89 Uppsala
Tel: +46 18 67 40 00, e-mail: sva@sva.se
Web: sva.se

SWEDISH
VETERINARY
AGENCY