Proficiency test number 28

Whole Genome Sequencing of *Campylobacter*

Ásgeir Ástvaldsson

PT28 Objectives

To assess the performance of DNA extraction and whole genome sequencing (WGS) of *Campylobacter*

Quantify differences between whole genome sequence (WGS) data from *Campylobacter*, produced at different laboratories

NRL Participation in PT28

- 20 labs signed up for participation
  - Due to Covid-19 the deadline was postponed from 1st of June until 1st of August
  - 11 labs submitted results within the deadline
  - 3 labs submitted Questback responses within the deadline but failed to upload data in time (not included in the presentation)
  - 2 labs did not submit Questback responses and uploaded data after the deadline (not included in the presentation)

PT28 samples

PT28-1 – 30 µl of stabilized genomic DNA extracted from *Campylobacter jejuni*

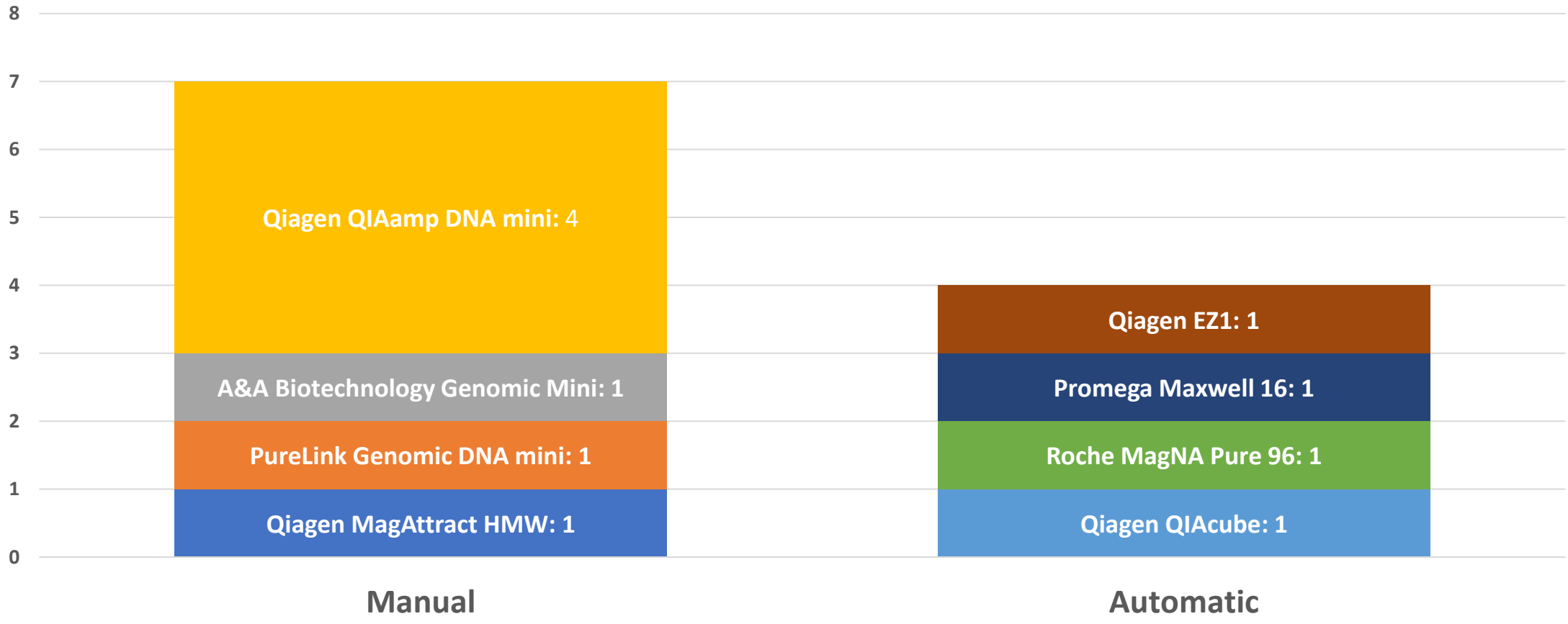
PT28-2 – 30 µl of stabilized genomic DNA extracted from *Campylobacter coli*

PT28-3 – lyophilised *Campylobacter jejun*i corresponding to PT28-1

PT28-4 – lyophilised *Campylobacter coli* corresponding to PT28-2

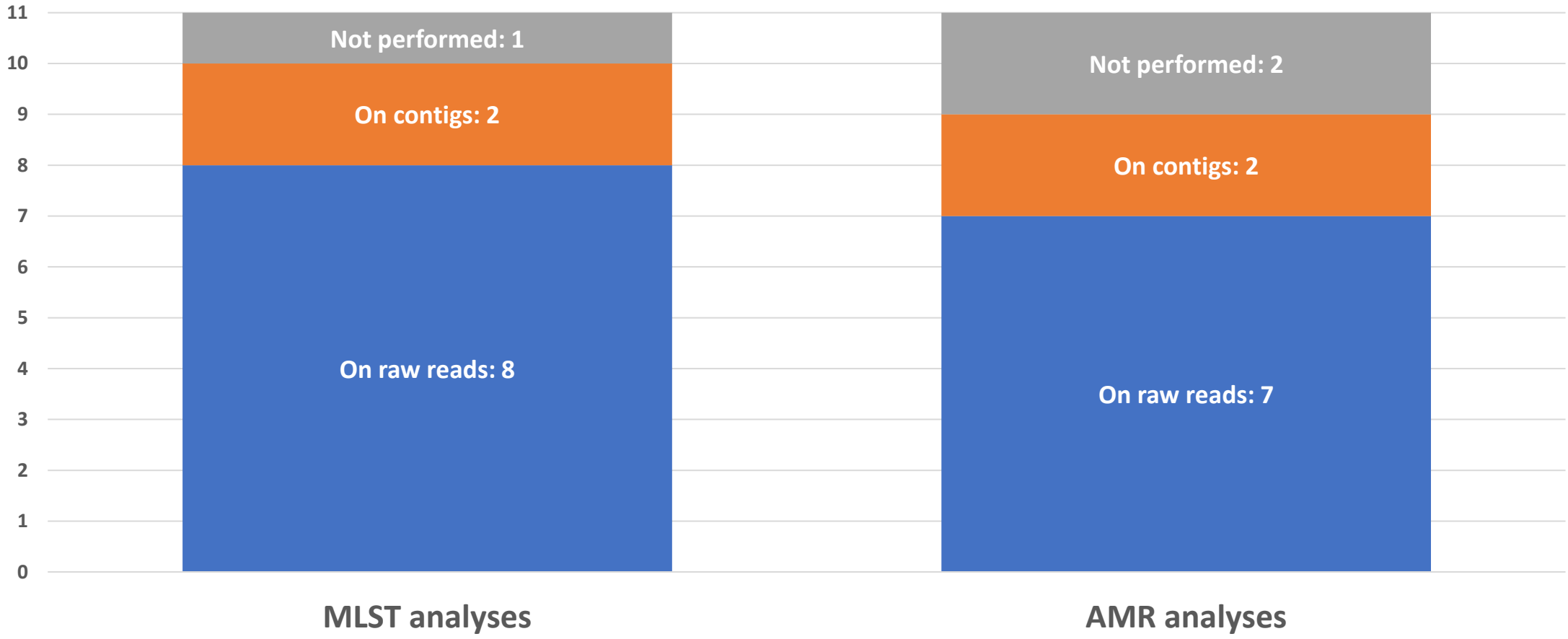
Data reported by the NRLs through Questback questionnaire

DNA extraction method



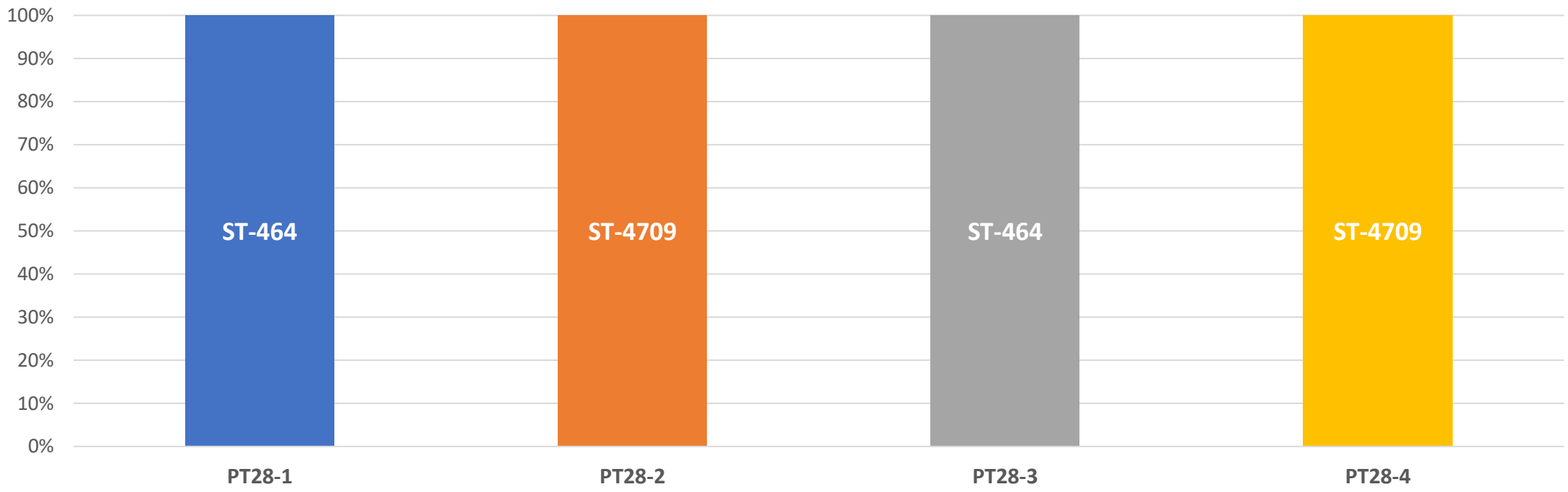
Data reported by the NRLs through Questback questionnaire

MLST and AMR analyses



Data reported by the NRLs through Questback questionnaire

Determined ST-number



Reference strains ST-number

PT28-1 and PT28-3, *C. jejuni* : **ST-464**

PT28-2 and PT28-4, C. coli : **ST-4709**

Data reported by the NRLs through Questback questionnaire

AMR genes identified

LabID	PT28-1 *C. jejuni*	PT28-3 *C. Jejuni*	PT28-2 *C. coli*	PT28-4 *C. coli*
18	tetO		blaOXA-61	
19	tetO		blaOXA-61 family gene	
23	tetO and cmeR		blaOXA-193 or blaOXA-61 like	
24	tetO		blaOXA-like	
35	tetO and cmeR		blaOXA-61 and cmeR	
49	tetO		blaOXA	
58	tetO		blaOXA	
61	tetO and cmeR		blaOXA-61	
65	tetO		blaOXA-193	
Ref. strains	tetO – tetracycline resistance acr3 – asenite efflux (stress) arsP – organoarsenical efflux (stress)		blaOXA-193 (OXA-61 family class) – Beta-lactam resistance	

Referense strains – AMRFinderPlus v3.2.3, database v. 2019-10-30.1

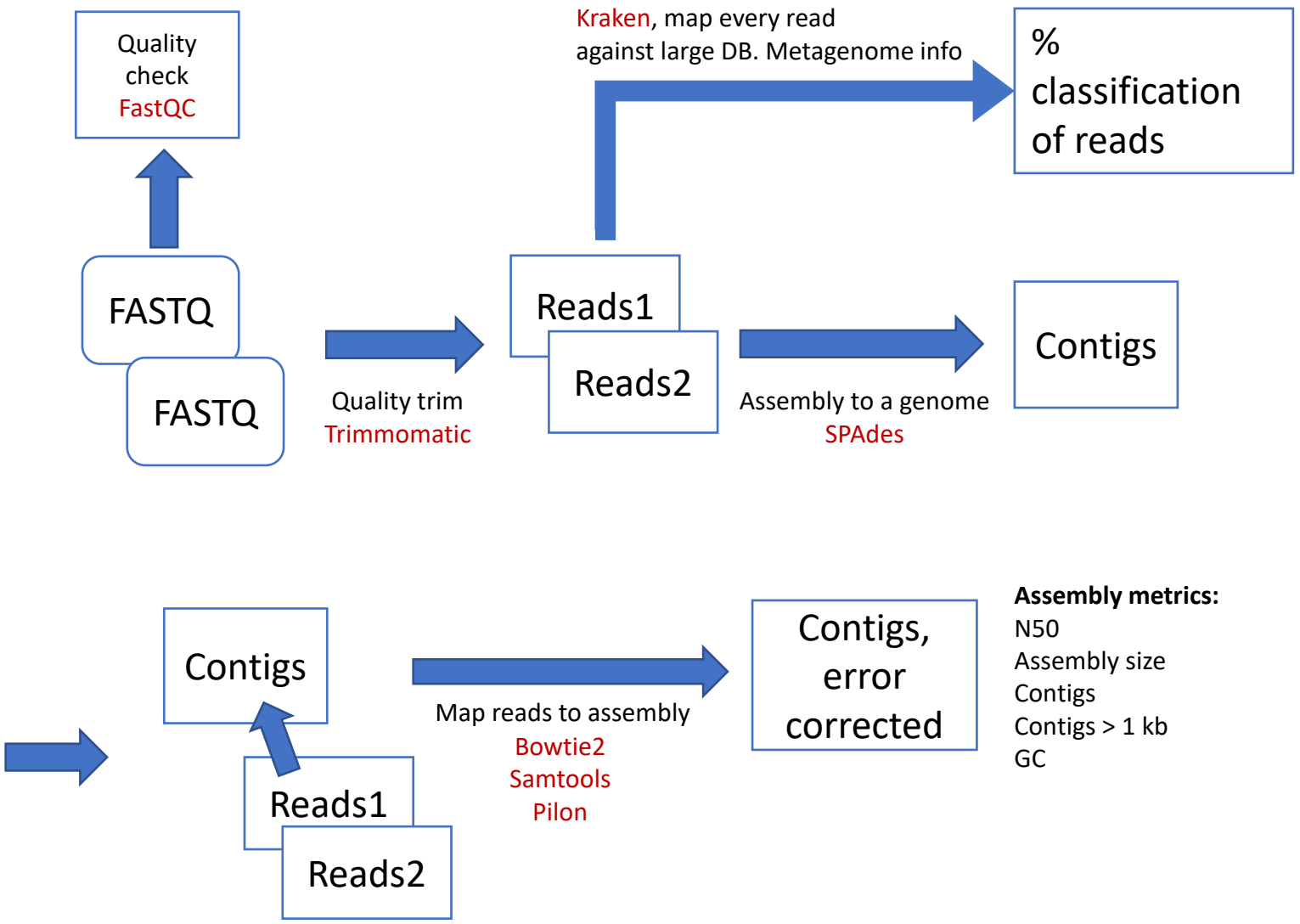
Data reported by the NRLs through Questback questionnaire

Point mutations possibly leading to AMR

LabID	PT28-1 *C. jejuni*	PT28-3 *C. Jejuni*	PT28-2 *C. coli*	PT28-4 *C. coli*
18	gyrA p.T86I		None	
19	gyrA p.T86I		None	
23	gyrA p.T86I ACA>ATA		None	
24	gyrA p.T86I – gyrA p.Q863* - cmeR p.T6I – cmeR p.G144D – cmeR p.P183R – cmeR p.S207G		rpsL pA119T GCT>ACT	
35	gyrA p.T86I ACA>ATA		None	
49	gyrA p.T86I		None	
58	gyrA p.T86I		None	
61	gyrA p.T86I		None	
65	gyrA p.T86I ACA>ATA		None	
Ref. strains	gyrA p.T86I – Quinolone resistance 50S_L22_A103V – Macrolide resistance		None	

Referense strains – AMRFinderPlus v3.2.3, database v. 2019-10-30.1

EURL-*Campylobacter* assembly pipeline (simplified)



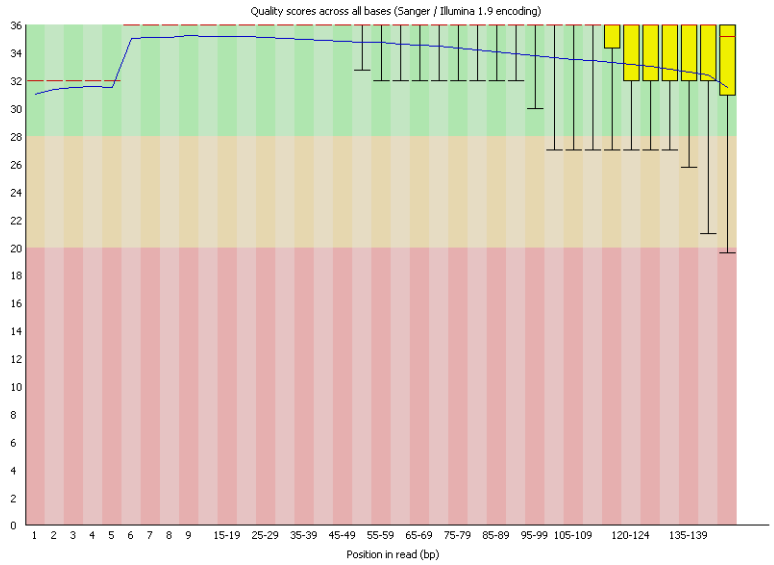
Assembly metrics:
N50
Assembly size
Contigs
Contigs > 1 kb
GC

PT28 data from all the NRLs was analyses using this pipeline

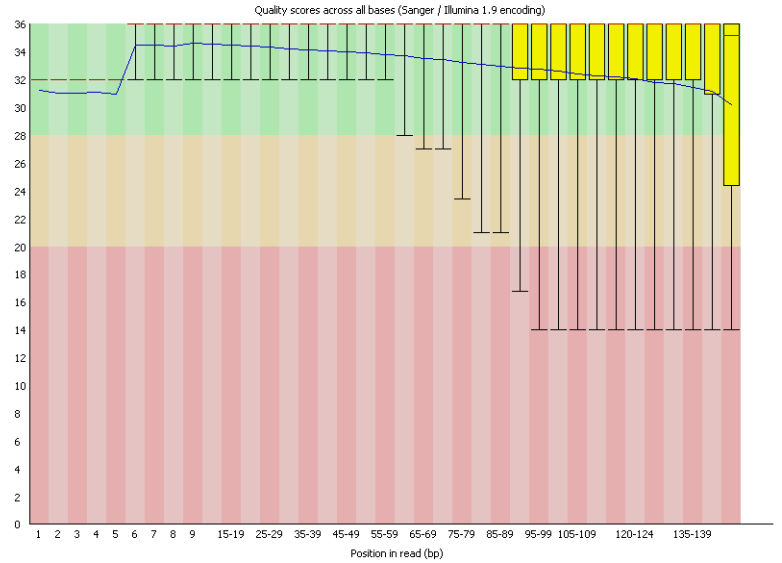
No performance threeshold has been set –> labs are not judged by the performance

Assembly pipeline results - examples

LabID-35 – high quality in both forward and reverse reads



PT28-1-35.R1.fastqc



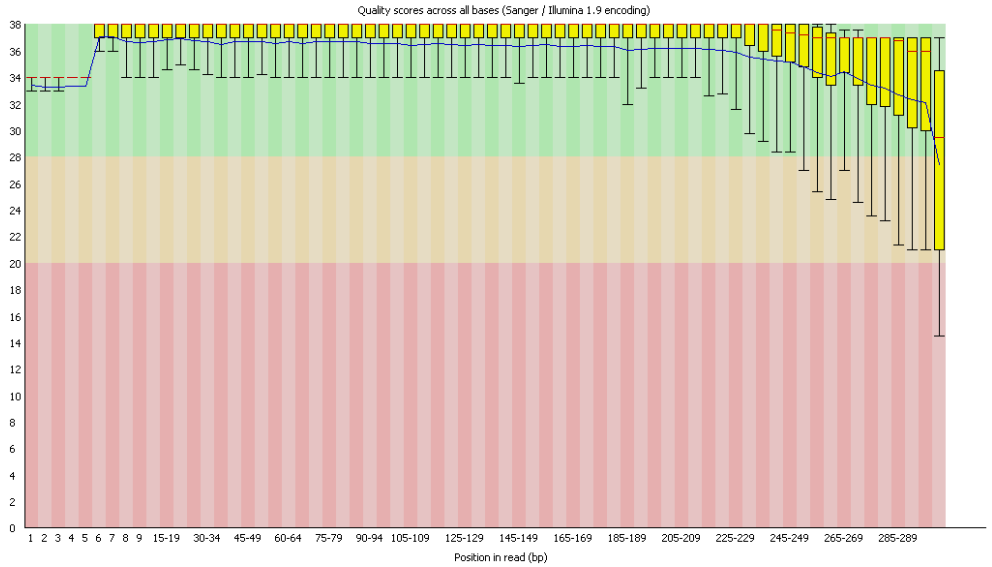
PT28-1-35.R2.fastqc

	Total reads	Coverage	No. of contigs	N50	Assembly size
PT28-1-35	2335200	>100x	61	154573	1742313
PT28-2-35	2232404	>100x	79	203647	1790040
PT28-3-35	2335200	>100x	64	154893	1742550
PT28-4-35	1982024	>100x	86	203647	1791117

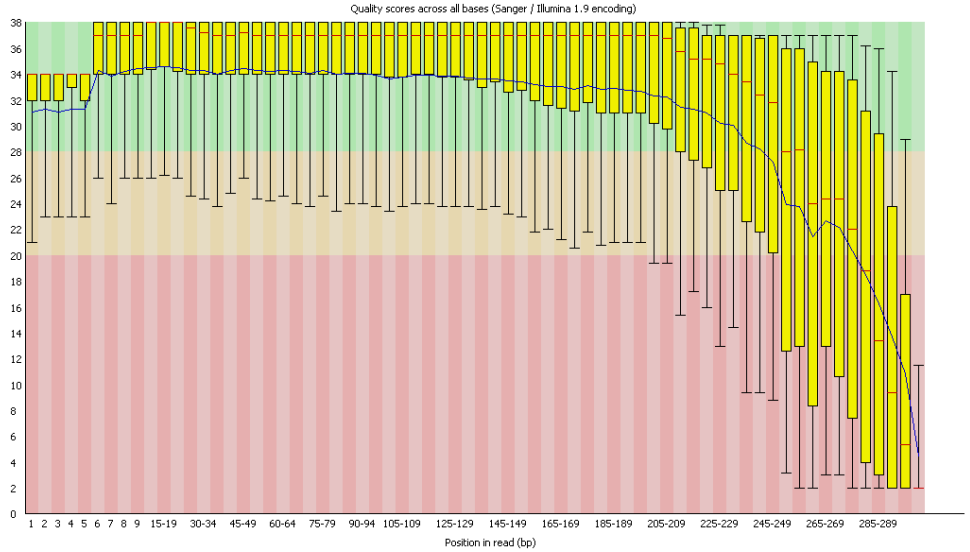
Many reads ➡ High coverage ➡ Few contigs ➡ High N50 size ➡ Lower variations

Assembly pipeline results - examples

LabID-18 – high quality in forward reads but bad quality in reverse reads



PT28-1-18.R1.fastqc



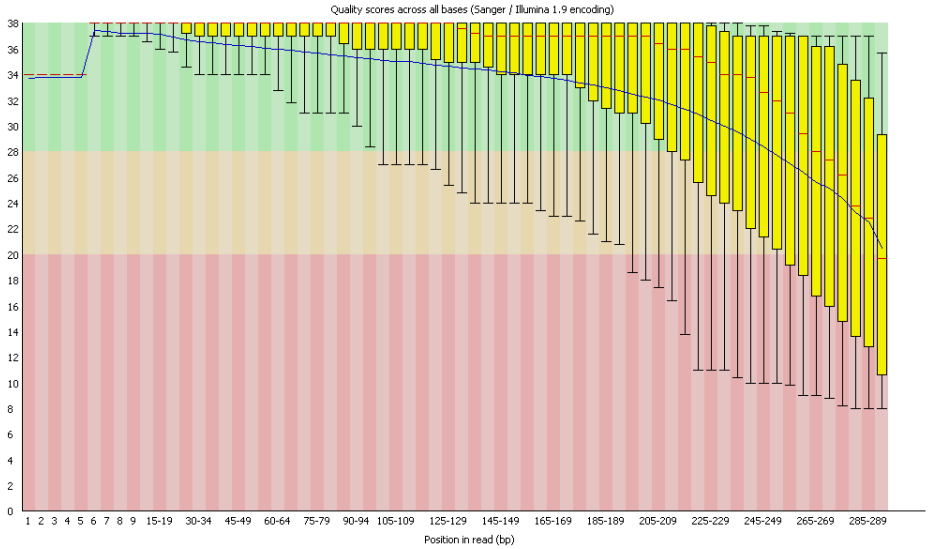
PT28-1-18.R2.fastqc

	Total reads	Coverage	No. of contigs	N50 (bp)	Assembly size (bp)
PT28-1-18	274481	64x	121	29113	1726994
PT28-2-18	232622	44x	206	16506	1762737
PT28-3-18	251635	50x	153	24460	1735322
PT28-4-18	344552	72x	107	62457	1779227

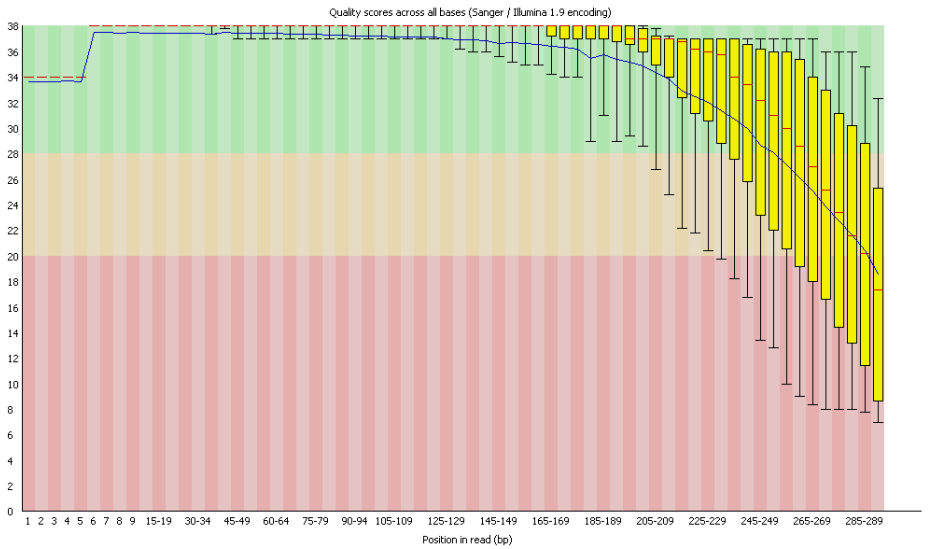
Few reads ➡ Low coverage ➡ Many contigs ➡ Low N50 size ➡ High variations

Assembly pipeline results - examples

LabID-65 – low quality in both forward and reverse reads



PT28-1-65.R1.fastqc



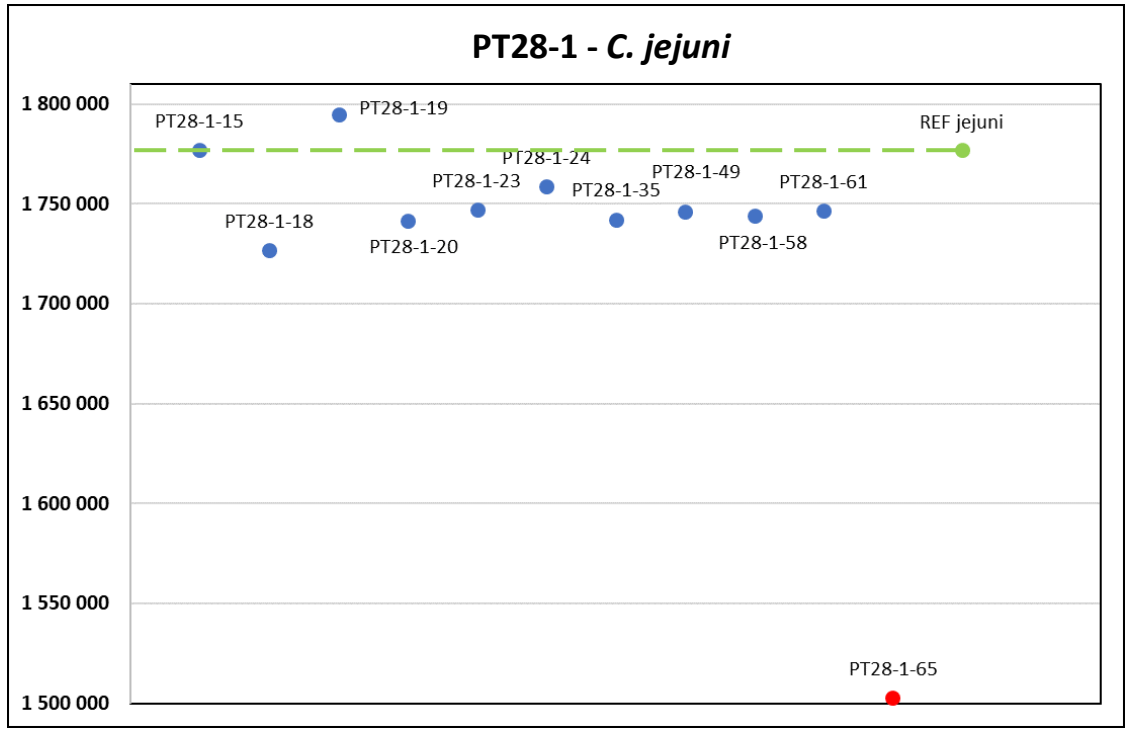
PT28-1-65.R2.fastqc

	Total reads	Coverage	No. of contigs	N50	Assembly size
PT28-1-65	1095510	>100x	643	3974	1502770
PT28-2-65	737758	>100x	466	7107	1683691
PT28-3-65	939141	>100x	400	7906	1644818
PT28-4-65	841077	>100x	302	11904	1719654

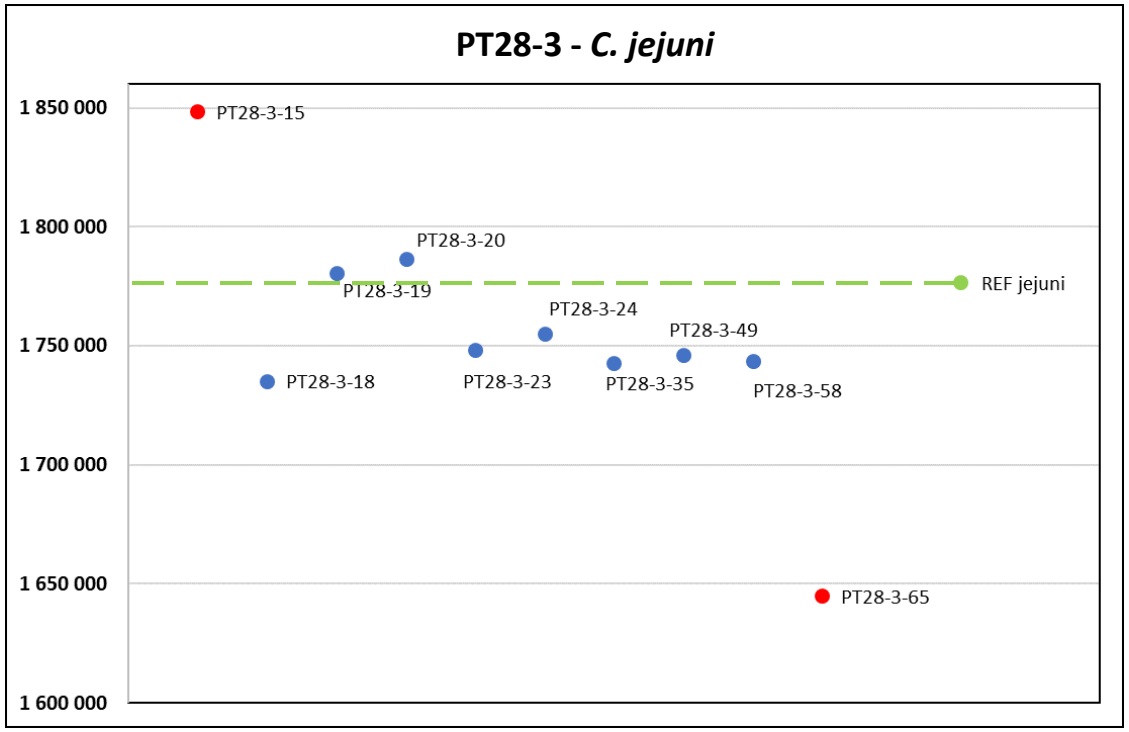
Ok number of reads ➡ Ok coverage ➡ Many contigs ➡ Very low N50 size ➡ Low size and big variations

Assembly pipeline results

Assembly size



DNA

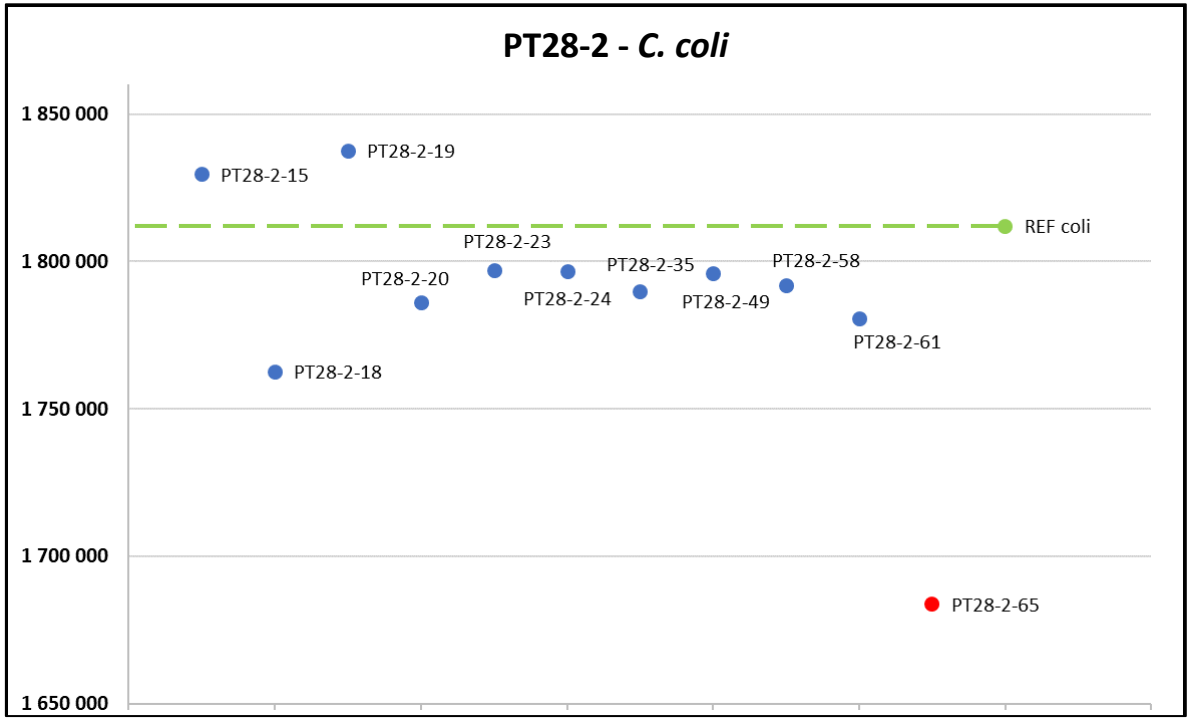


STRAIN

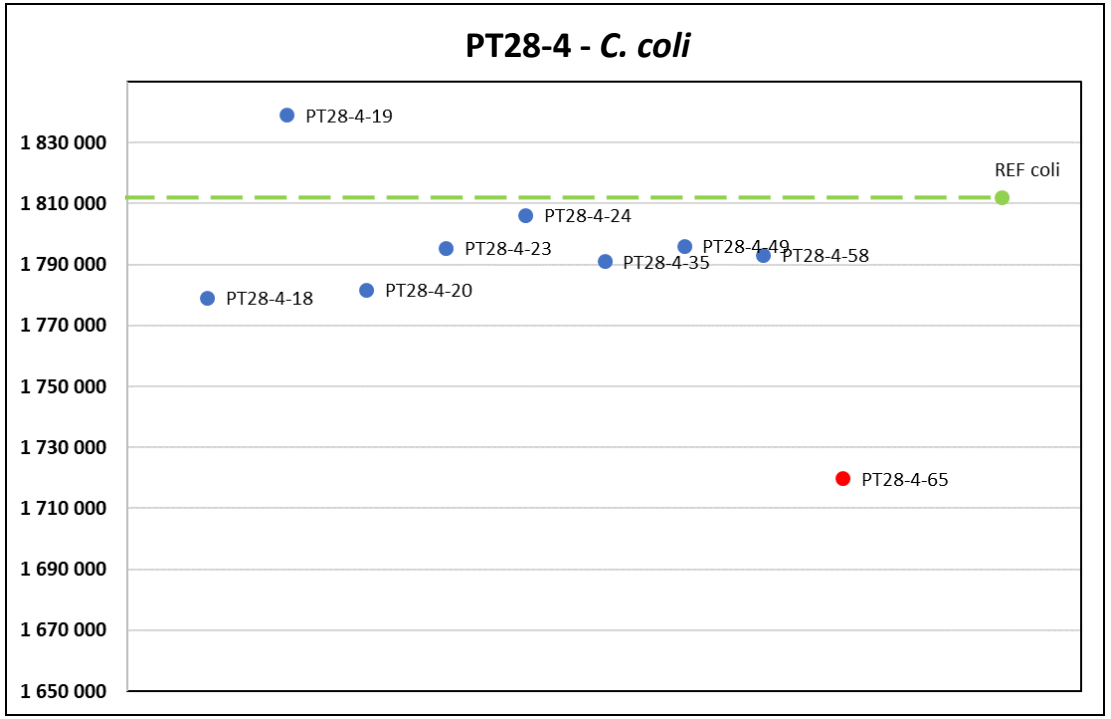
The assembly for the *C. jejuni* reference strain has one gap and is assembled using both short read Illumina data and long read Oxford Nanopore data.

Assembly pipeline results

Assembly size



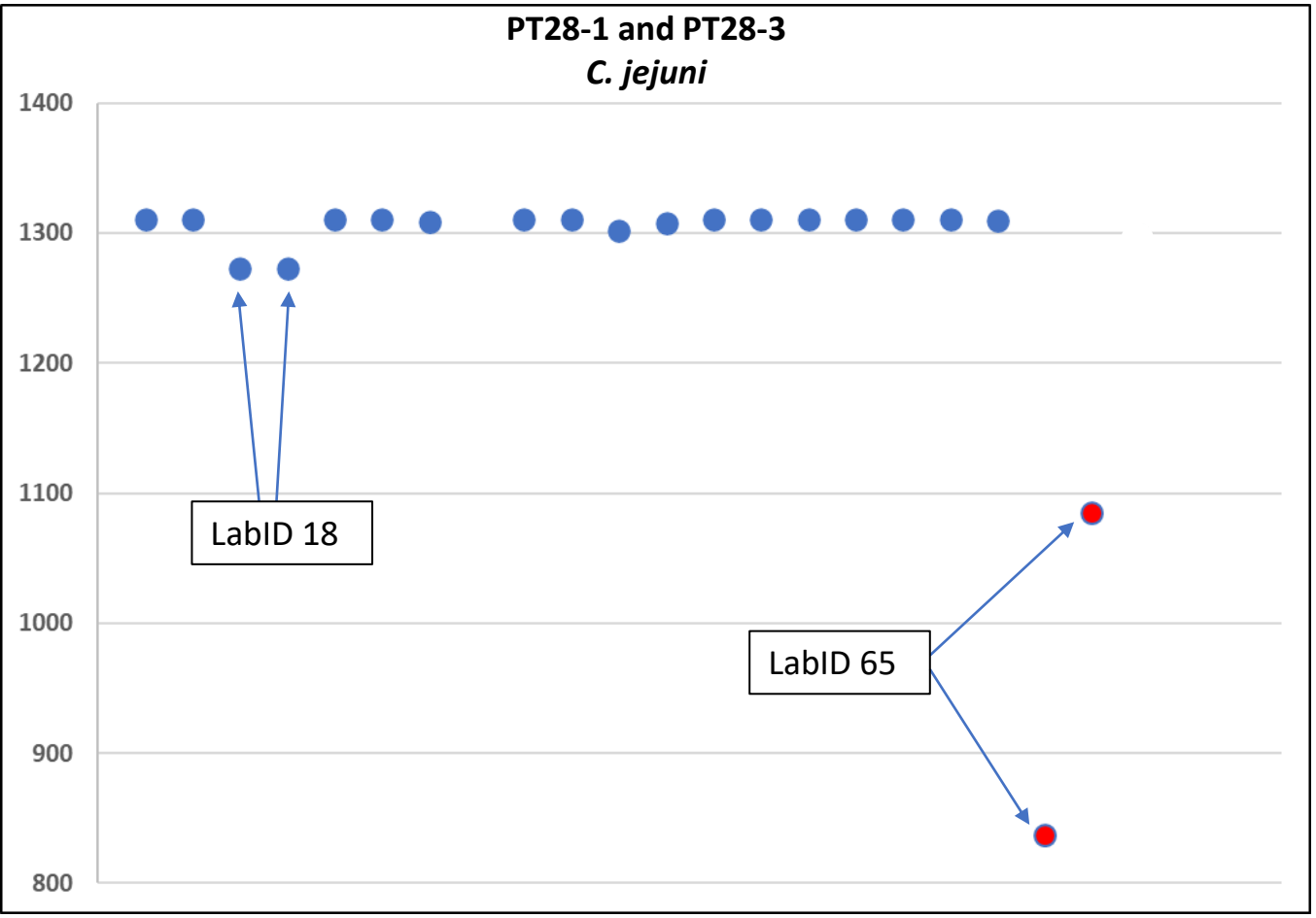
DNA

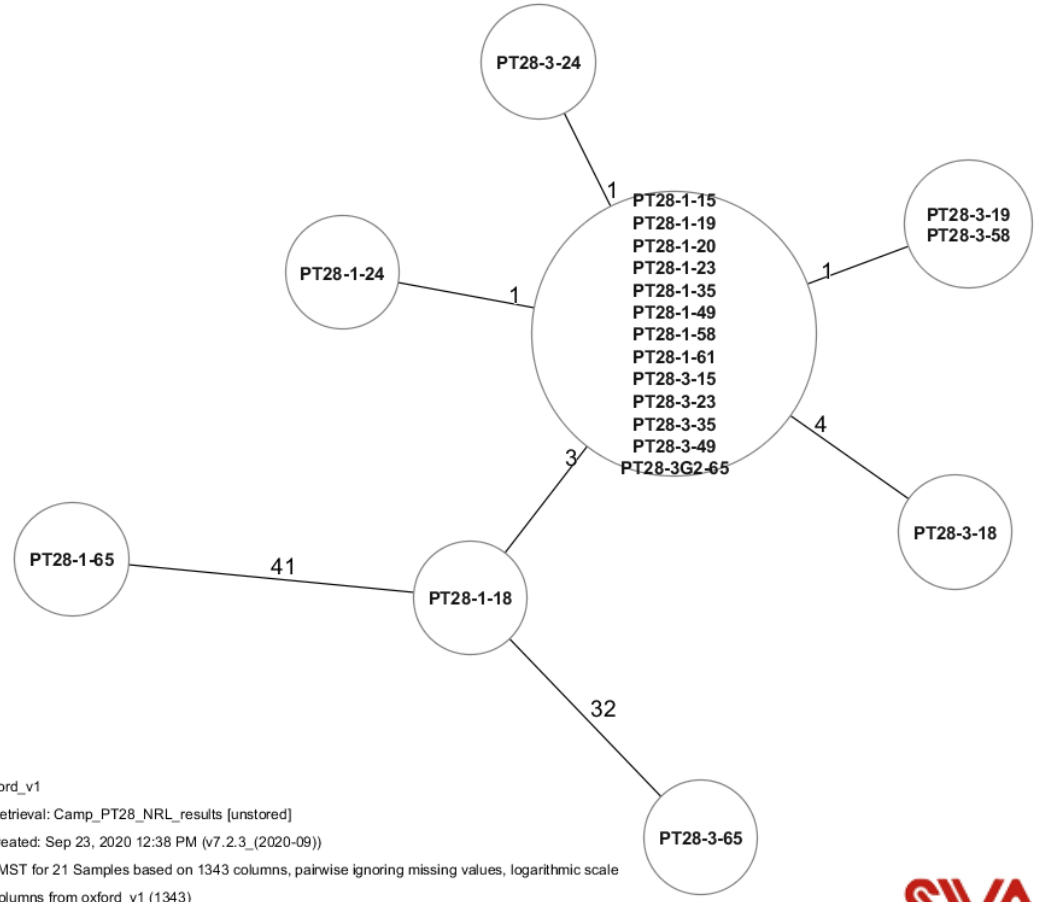


STRAIN

The assembly for the *C. coli* reference strain is gap free and is assembled using both short read Illumina data and long read Oxford Nanopore data.

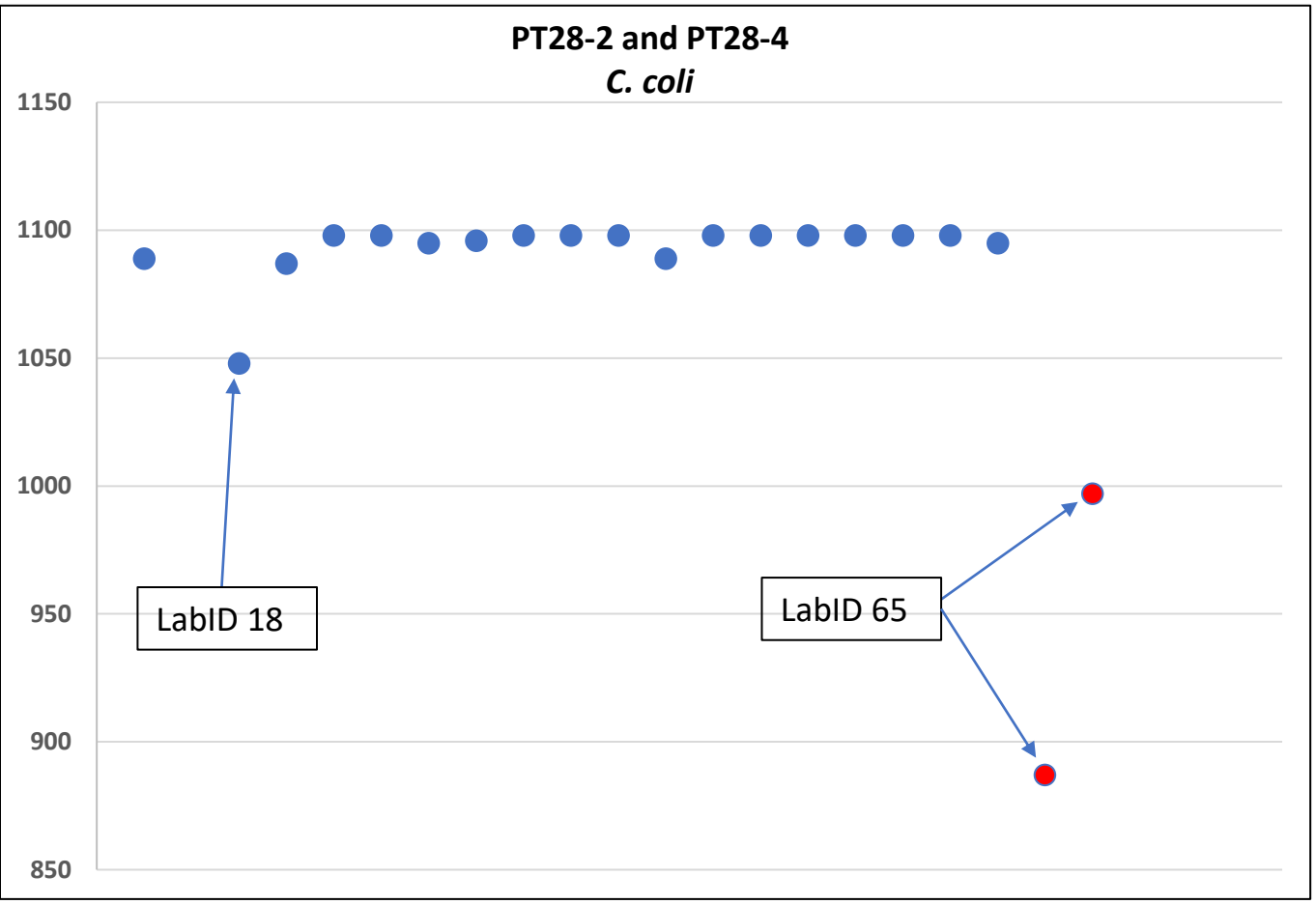
Allele calling - results

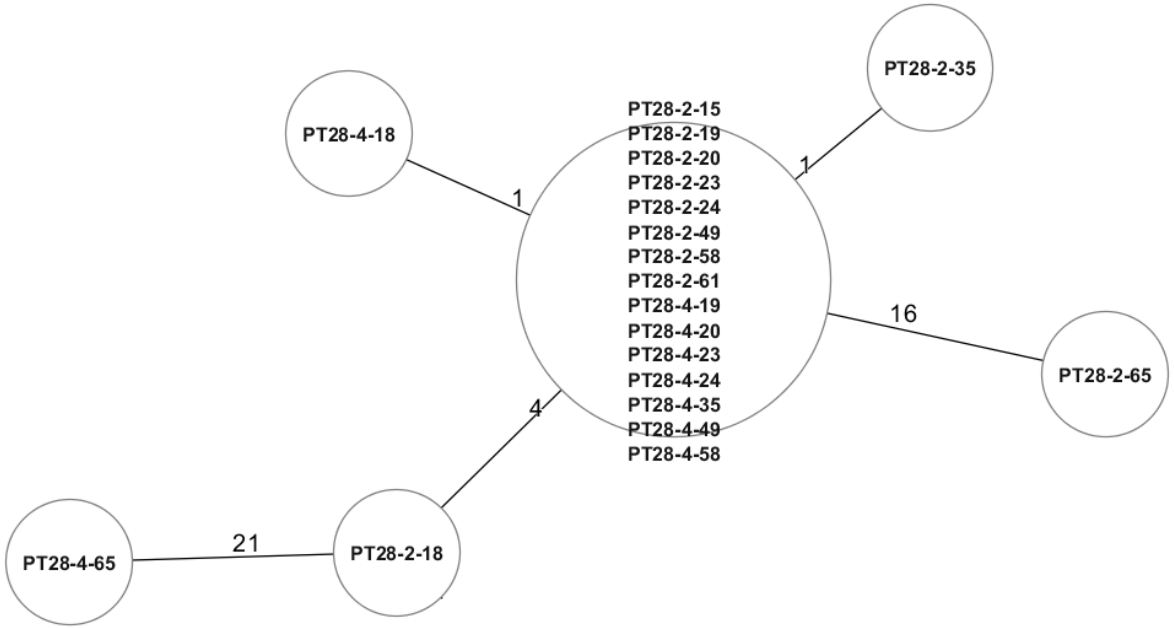




Alleles called for *C. jejuni* using Oxford cgMLST-scheme (1343 targets)

Allele calling - results





Alleles called for *C. coli* using an ad-hoc cgMLST-scheme (1121 targets)

Task Templates: C. coli cgMLST 1121 targets
Comparison Table Retrieval: Camp_PT28_NRL_results [unstored]
Comparison Table created: Sep 23, 2020 3:04 PM (v7.2.3_(2020-09))
Ridom SeqSphere+ MST for 21 Samples based on 1121 columns, pairwise ignoring missing values, logarithmic scale
Distance based on columns from C. coli cgMLST 1121 targets (1121)
For citing correctly in publications the tools used for this analysis see menu Help | Citations and Licenses.

Contamination in reads

- Many datasets had reads derived from something else than *Campylobacter*
  - Possible due to carry-over from previous sequencing runs
  - Contaminated buffers

Contamination examples

Contaminant	Number of labs
Alteromonas macleodii	Many labs had this contamination – EURL buffers?
Listeria	3 labs
E. coli	3 labs
Klebsiella pneumonaie	1 lab
Neisseriae	1 lab
Salmonella enterica	4 labs
Mycobacterium tuberculosis	1 lab

Sequencing contamination can lead to poor assemblies with many contigs

Summary

- NRLs that performed MLST and AMR analyses could identify correct ST-number and AMR genes for all samples
- Only small variations between corresponding samples (DNA and strain) were detected > NRLs capable of extracting and generating quality DNA for sequencing
- Sequencing reads quality and genome coverage are very important factors to obtain quality assemblies
- Further analyses on the PT28 NRL data, e.g. mapping of reads to the reference genome
  - Presented in the final report