IZS TERAMO

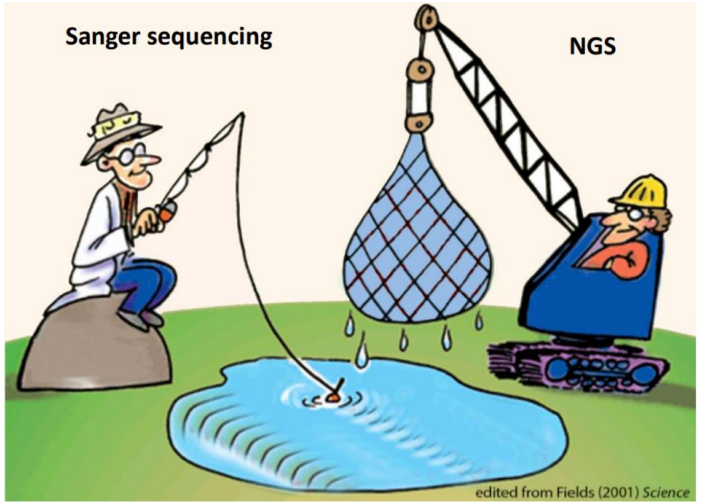
LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

Perspectives and challenges in accreditation of NGS-based methods

Giuliano Garofolo

The 18th EURL – Campylobacter workshop:
Videoconference 26-27th September 2023

INTRODUCTION



In recent decades, **Whole genome sequencing (WGS)** has increased the understanding of the evolutionary and epidemiological dynamics of foodborne pathogens and improved surveillance and outbreak detection because of its high discriminatory power, compared to traditional molecular techniques.

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

MOLECULAR EPIDEMIOLOGY

TRADITIONAL MICROBIOLOGICAL TECHNIQUES versus NGS FOR SURVEILLANCE:

The molecular epidemiology has been defined as the use of molecular typing methods for infectious agents in order to study the distribution, dynamics, and determinants of health and disease

Traditional typing

Often many methods are requested for each isolate (e.g. Serotyping, Antimicrobial susceptibility testing, MLST, PFGE)

- Time required
- Limited accuracy
- Expensive

WGS ADVANTAGES:

- Massive parallel sequencing
- Reasonable times
- Cost are becoming affordable
- High resolution

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

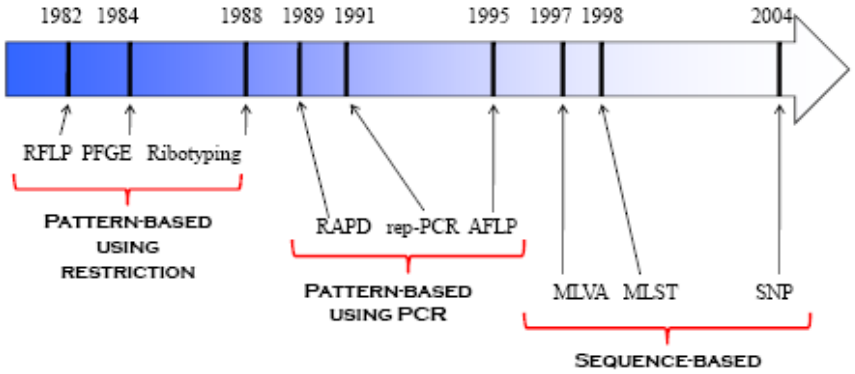
Campylobacter typing

MLST

WGS -cgMLST

Genome-based:
Within the last years there has been an ongoing revolution in bacterial typing with a move away from 'band-based' approaches such as those described above to 'genome-based' approaches

1. Multi locus approach reduce mistakes
2. Easy to compare and store
3. Makers with different mutation rates



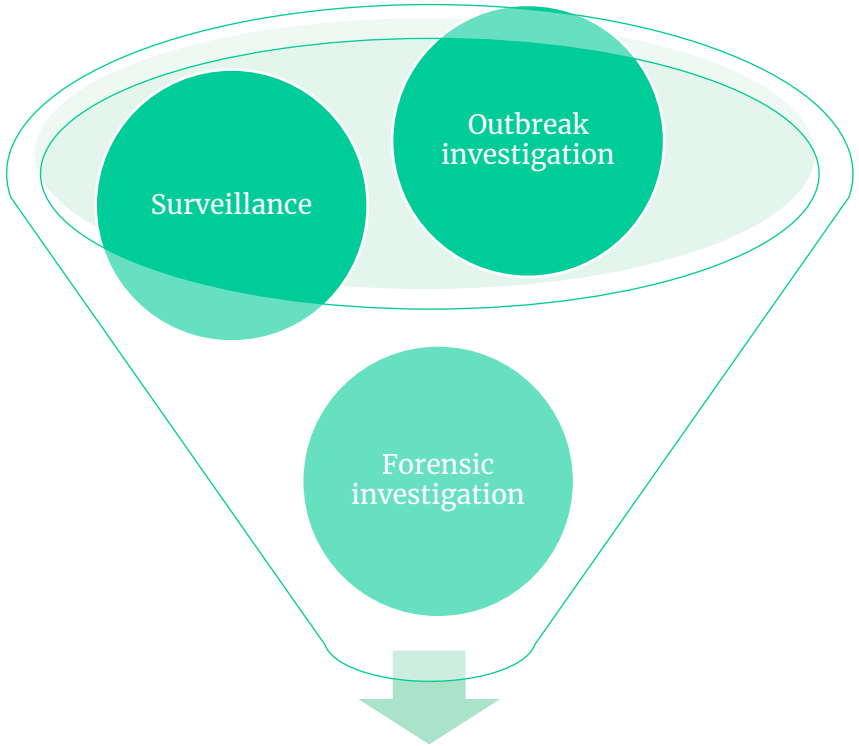
IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

PERSPECTIVE NRL-IT *Campylobacter*

Early application of genomics for investigation of foodborne outbreaks demonstrated the utility of the technology

How can we ensure that WGS methods are fit for purpose?



The validation of new technologies WGS is essential to ensure reliable results even for legal proceedings

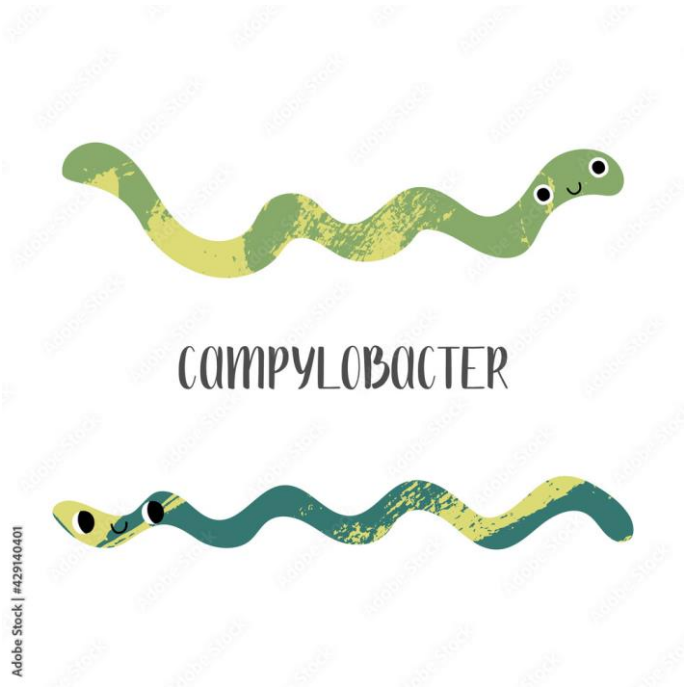
1. Legal action against companies involved in food production or distribution (foodborne outbreaks)
2. Use of genomics data to inform public health authorities.....

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

PERSPECTIVE – NRL-IT *Campylobacter*

THE EFSA ONE HEALTH WGS SYSTEM



"……..but could be extended to include other food-borne pathogens such as Campylobacter and food-borne viruses, upon agreement between EFSA, ECDC, Data Providers and the European Commission"

https://doi.org/10.2903/sp.efsa.2022.EN-7413

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

NRL-IT Campylobacter – challenges

Genomic data in a public health context should conform to best practices:

1. Wet-laboratory NGS workflows
2. Bioinformatic analysis

Validation end to end process versus separate steps:

1. Maintenance of software
2. Demonstration of data integrity
3. Version traceability
4. Documentation of process (standard operating procedures)

Validation for accreditation:

1. Accreditation is the procedure by which an authoritative body gives formal recognition
2. Following ISO or ISO-equivalent standards
3. ISO 17025:2017 General requirements for the competence of testing and calibration laboratories
4. ISO 23418:2022 Microbiology of the food chain — whole-genome sequencing for typing and genomic characterization of bacteria — general requirements and guidance

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

Quality System

Since 1995 the Institute is accredited according to:

ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories

Qualitative continual improvement:
Quality, policies, Vision, mission
SOPs
Document control
Quality control
Proficiency Testing
Effective quality indicator
Internal audit
Management review
Accreditation Visits

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

AIM OF THE STUDY

Validation WGS workflow, according to performance criteria including repeatability, reproducibility and epidemiological concordance, for *C. jejuni* and *C. coli* :

- The process was divided into two steps:
- Wet-WGS workflow (DNA extraction to NGS sequencing)
- Bioinformatic pipeline

Two SOPs:

B3.1.3 -SOP071 ANALISI DATI DI SEQUENZIAMENTO NGS PER LA GENOTIPIZZAZIONE DI CAMPYLOBACTER JEJUNI E CAMPYLOBACTER COLI

B2.1.9 -SOP032 SEQUENZIAMENTO GENOMICO DI ISOLATI BATTERICI

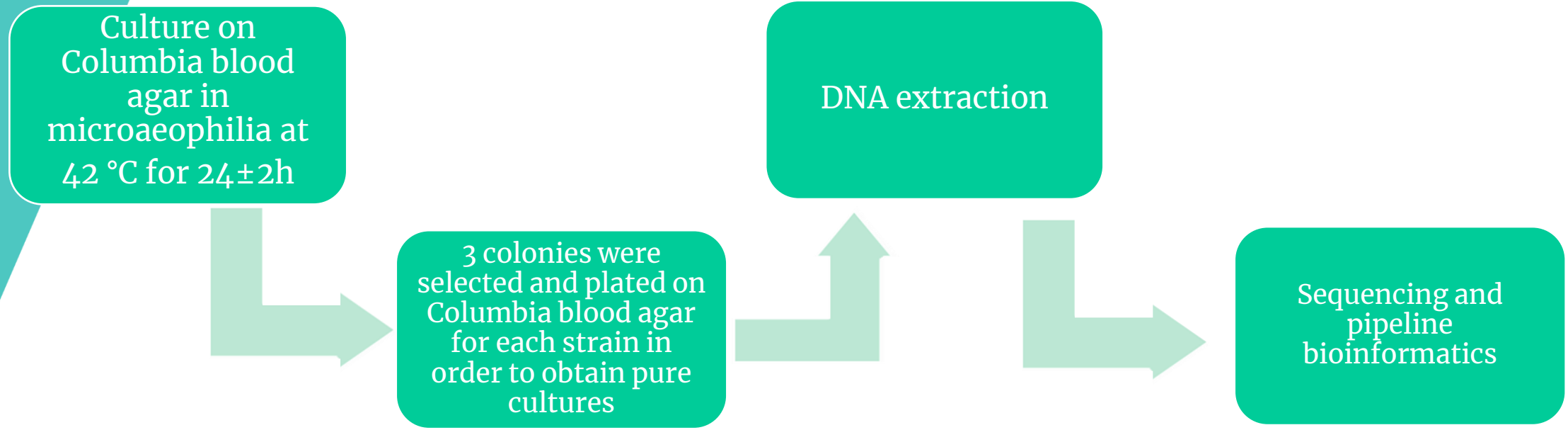
IZS
TERAMO
LABORATORIO
NAZIONALE
DI RIFERIMENTO PER
CAMPYLOBACTER

MATERIALS AND METHODS – WGS WORKFLOW

STRAINS

3 Bacterial strains used (from the ATCC and NCTC collections)

ATCC 43431	***C. jejuni***
ATCC 33291	***C. jejuni***
NCTC 11353	***C. coli***



IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

DNA EXTRACTION AND QUALITY/CONC. EVALUATION

- Bacterial DNA was extracted using QIAamp DNA Mini Kit (QIAGEN)

Quantity

Purity

Integrity

Bacterial DNA concentrations were measured using Qubit fluorometric quantitation with Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific)

Bacterial DNA purity was estimated using NanoDrop

The integrity of the extracted DNA was assessed by determining the DNA Integrity Number (DIN) with the Agilent 2200 Tape station

- Repeatibility
- Reproducibility

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

SEQUENCING

- 27 DNA samples extracted from C. jejuni and C. coli strains were sequenced on 3 different run on Illumina NextSeq500 and NextSeq2000 platforms, after preparation of the genomic libraries with the Illumina DNA prep kit.

		RUN	RUN	RUN
Refer. strain	Species	220503_NS500787_0269 AHKN2LAFX3	220617_VH00572_8 AAC5CG7M5	220629_VH00572_10 AAAV2FHM5
ATCC 43431	C. jejuni	2022.TE.34347	2022.TE.36349	2022.TE.36351
ATCC 33291	C. jejuni	2022.TE.34346	2022.TE.36350	2022.TE.36353
NCTC 11353	C. coli	2022.TE.34342	2022.TE.36346	2022.TE.36352
Run data		03/05/2022	17/06/2022	29/06/2022

Ref. strain		species	link
ATCC 43431	fasta	C. jejuni	https://www.atcc.org/products/43431
ATCC 33291	fasta	C. jejuni	https://www.atcc.org/products/33291
NCTC 11353	fasta	C. coli	https://www.ncbi.nlm.nih.gov/assembly/GCA_001495315.1

ATCC

NIH National Library of Medicine
National Center for Biotechnology Information

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

DNA EXTRACTION

Parameters	Acceptable values*	Acceptance
DNA concentration	≥ 3.5 ng/µl	✓
A260/230	2.0-2.2	✓
A260/280	1.75-2.05	✓
DNA integrity	≥ 7	✓

RUN ACCEPTANCE PARAMETERS

Parameters	Acceptable values*	Acceptance
PhiX error rate %	< 6%	✓
% ≥ Q30 Total	2x150bp ≥ 80%	✓
N. reads negative control	< 10.000	✓

*values according to ISO 23418:2022

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

SAMPLE ACCEPTANCE PARAMETERS

Parameter	Acceptable values*	Acceptance
Mean Phred score (Q-score)	≥ 30	✓
% ≥ Q30	≥80%	✓
Estimated coverage	≥ 30X**	✓
Repeatability	CV ≤ 5%	✓
Reproducibility	CV ≤ 5%	✓

ASSEMBLY QUALITY

Parameters	Acceptable values	Acceptance
C. jejuni genome length	1.641.481 bp ± 5% (ISO 23418:2022)	
C. coli genome length	1.724.380 bp ± 5% (ISO 23418:2022)	
Contigs no.	<300 (Timme et al. 2020)	

*values according to ISO 23418:2022

**more stringent according to EFSA guidelines for data providers

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

BIOINFORMATIC

ridom BIOINFORMATICS

- MLST *in silico*
- *flaA-SVR*
- cgMLST
- wgMLST

Software NGSmanager

(https://github.com/genpat-it/ngsmanager).

- trimmomatic (trimming)
- fastQC (quality check)
- spades (assembly)
- chewbbaca (MLST, cgMLST, wgMLST)
- GrapeTree - minimum spanning tree (clustering)

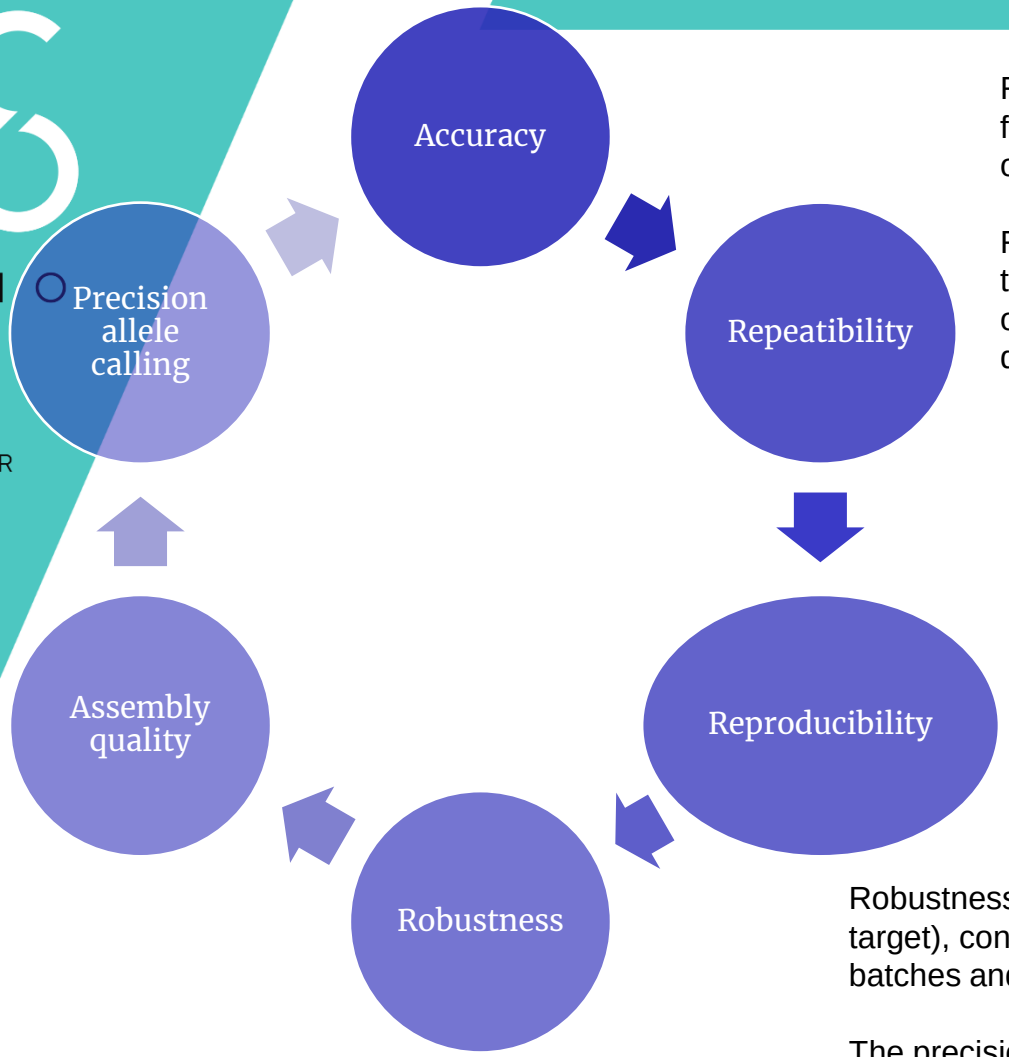
cgMLST scheme	No. loci
C. Jejuni - Innuendo	678
C. Coli - Innuendo	528
C. Jejuni – C. coli Ridom Seqsphere	637

wgMLST scheme	No. loci
C. Jejuni - Innuendo	**2,795**
C. Coli - Innuendo	**2,477**
C. Jejuni – C. coli Ridom Seqsphere	**1,595**

Rossi M., et al. (2018). INNUENDO whole genome and core genome MLST schemas and datasets for Campylobacter jejuni (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1322564

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*



Repeatability expresses the number of locus identified (percentage of good target) , from DNA extracted the same day by the same operator and isolates were sequenced on the same sequencing run with the same library preparation

Reproducibility expresses the number of locus identified (percentage of good target), from DNA extractions of each set of isolates by two (or more) operators. DNA obtained by each operator was split in on different Illumina sequencing runs using different sequencers and different batches for library preparation

Robustness expresses the number of locus identified (percentage of good target), considering all the changes in the experimental conditions (strains, operators, batches and sequencers).

The precision expresses the closeness of agreement between allelic calls considering all the changes in the experimental conditions (operators and Sequencers).

The accuracy expresses the closeness of agreement between allelic calls of our isolates against the reference assemblies

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

Bioinformatic pipeline ACCEPTANCE PARAMETERS

Number of good target ST	100%
Number of good target cgMLST	> 98%
Number of good target wgMLST	> 30%
REPEATIBILITY	$CV \leq 0{,}05$
RIPRODUCIBILITY	$CV \leq 0{,}05$
Accuracy of ALLELE CALLING cgMLST	Allelic distance detected in replicates < 10*

SeqSphere + v.6 (Ridom GmbH, Münster, Germany)

NGSmanager (https://github.com/genpat-it)

* Stringent criteria SCHÜRCH, A. C., et al. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection*, 2018, 24.4: 350-354.

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

WHOLE GENOME SEQUENCING VALIDATION

No. of good targets

MLST	100%
cgMLST	> 98%
wgMLST	> 30%
flaA-SVR	100%

Repeatability Reproducibility

TRIPLICATES

cgMLST	$CV \leq 0,05$
wgMLST	$CV \leq 0,05$

Campylobacter jejuni	No. Loci cgMLST	No. Loci wgMLST
Median	677.6	1071.0
Standard Deviation	0.5	0.6
Co-efficient of variation (CV)	0.0008	0.0005

Robusteness

27 SAMPLES

cgMLST	$CV \leq 0,05$
wgMLST	$CV \leq 0,05$

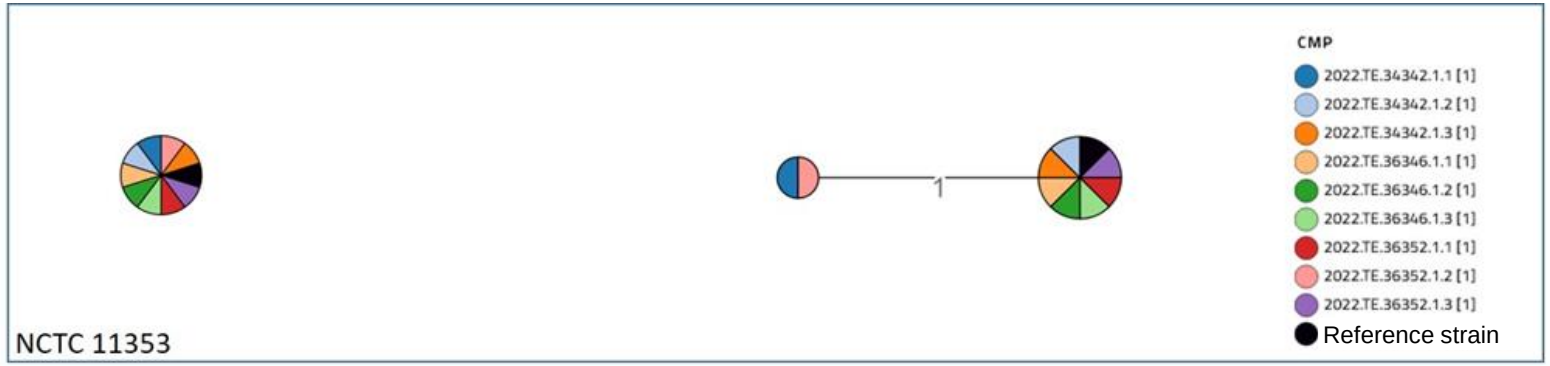
Campylobacter coli	No. Loci cgMLST	No. Loci wgMLST
Median	528.0	912.0
Standard Deviation	0.0	2.6
Co-efficient of variation (CV)	0.0000	0.0029

Bioinformatic pipeline Accuracy

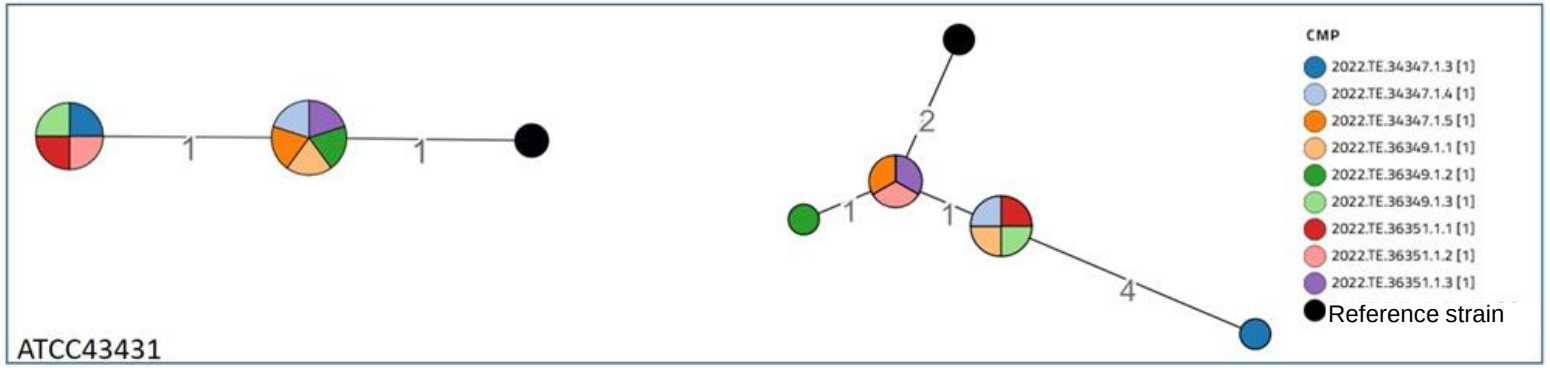
IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

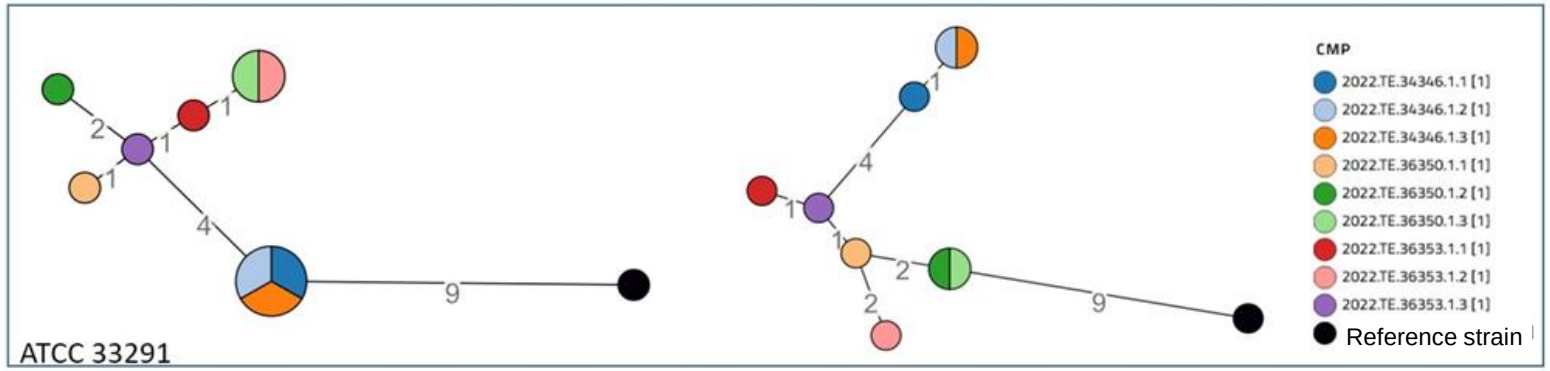
C.coli



C.jejuni



C.jejuni



SeqSphere + v5.0.90 (Ridom GmbH, Münster, Germany)

NGSmanager (https://github.com/genpat-it).

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

Epidemiological concordance

> J Med Microbiol. 2021 Mar;70(3). doi: 10.1099/jmm.0.001262. Epub 2021 Jan 20.

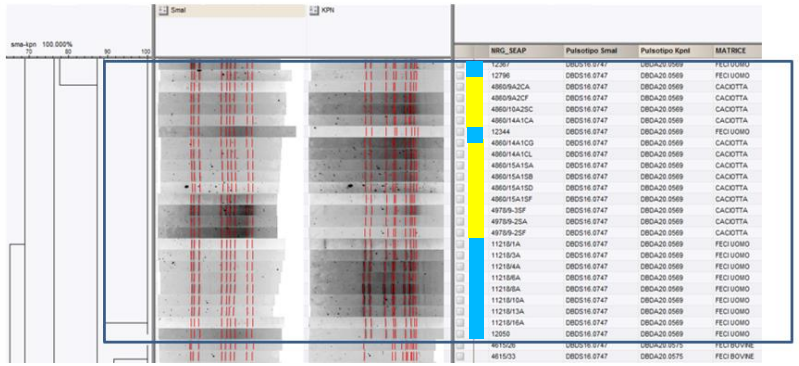
A large food-borne outbreak of campylobacteriosis in kindergartens and primary schools in Pescara, Italy, May-June 2018

Simona Sorgentone [1], Luca Busani [2], Paolo Calistri [3], Giorgio Robuffo [1], Stefania Bellino [2], Vicdalia Acciari [3], Maurizio Ferri [1], Caterina Graziani [4 2], Salvatore Antoci [3], Fabrizio Lodi [1], Valeria Alfonsi [5 2], Cesare Cammà [3], Paolo Fazii [6], Xanthi Andrianou [2], Francesca Cito [3], Giuliano Lombardi [6], Gabriella Centorotola [3], Massimo D'Amario [1], Nicola D'Alterio [3], Vincenzo Savini [6], Fabrizio De Massis [3], Anna Pelatti [6], Marco Di Domenico [3], Guido Di Donato [3], Elisabetta Di Giannatale [3], Lisa Di Marcantonio [3], Violeta Di Marzio [3], Gabriella Di Serafino [3], Anna Janowicz [3], Cristina Marfoglia [3], Francesca Marotta [3], Daniela Morelli [3], Giacomo Migliorati [3], Diana Neri [3], Francesco Pomilio [3], Silvia Scattolini [3], Giovanni Rezza [7 2], Antonio Caponetti [1], Patrizio Pezzotti [2], Giuliano Garofolo [3]

Affiliations + expand

PMID: 33475480 DOI: 10.1099/jmm.0.001262

PFGE



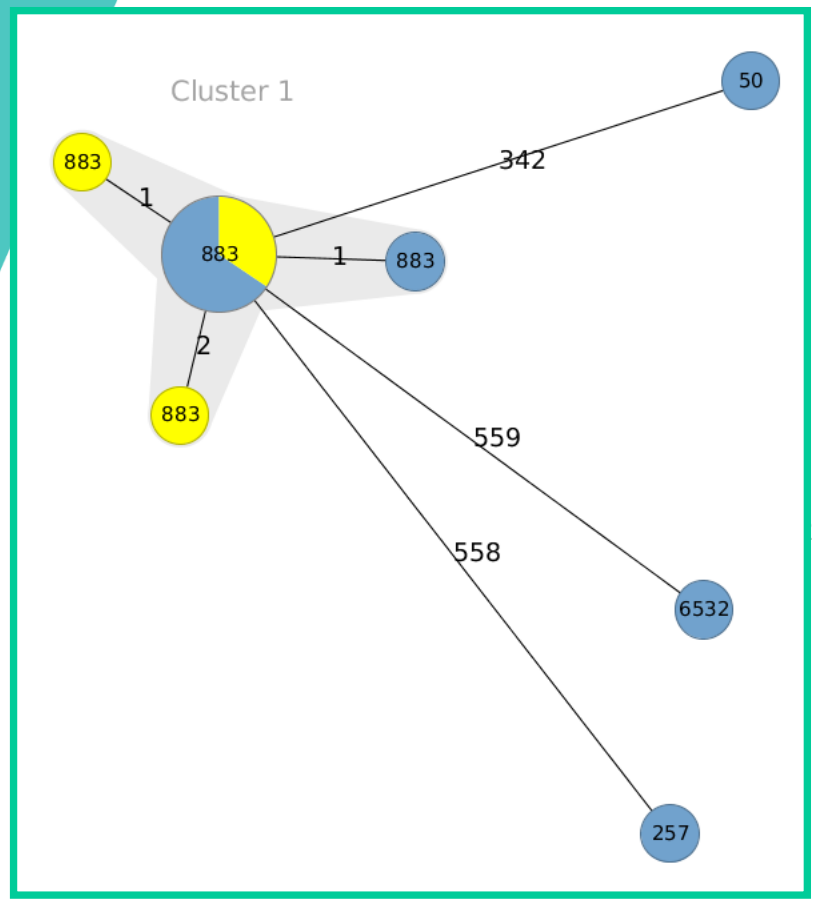
SeqSphere + v5.0.90 (Ridom GmbH, Münster, Germany)

NGSmanager (https://github.com/genpat-it).

Epidemiological concordance

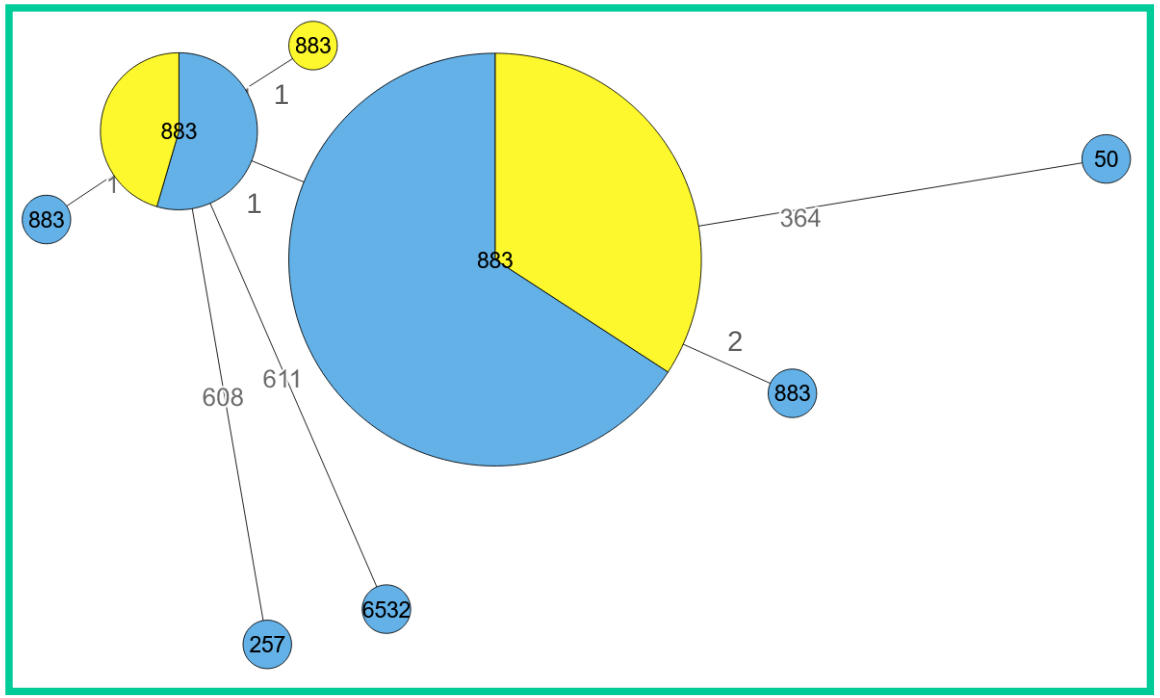
SeqSphere + v.6 (Ridom GmbH, Münster, Germany)

Schema: cgMLST_c_jejuni,637



NGSmanager (https://github.com/genpat-it)

Schema “cgmlst_c_jejuni,678



LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

IZS.IT

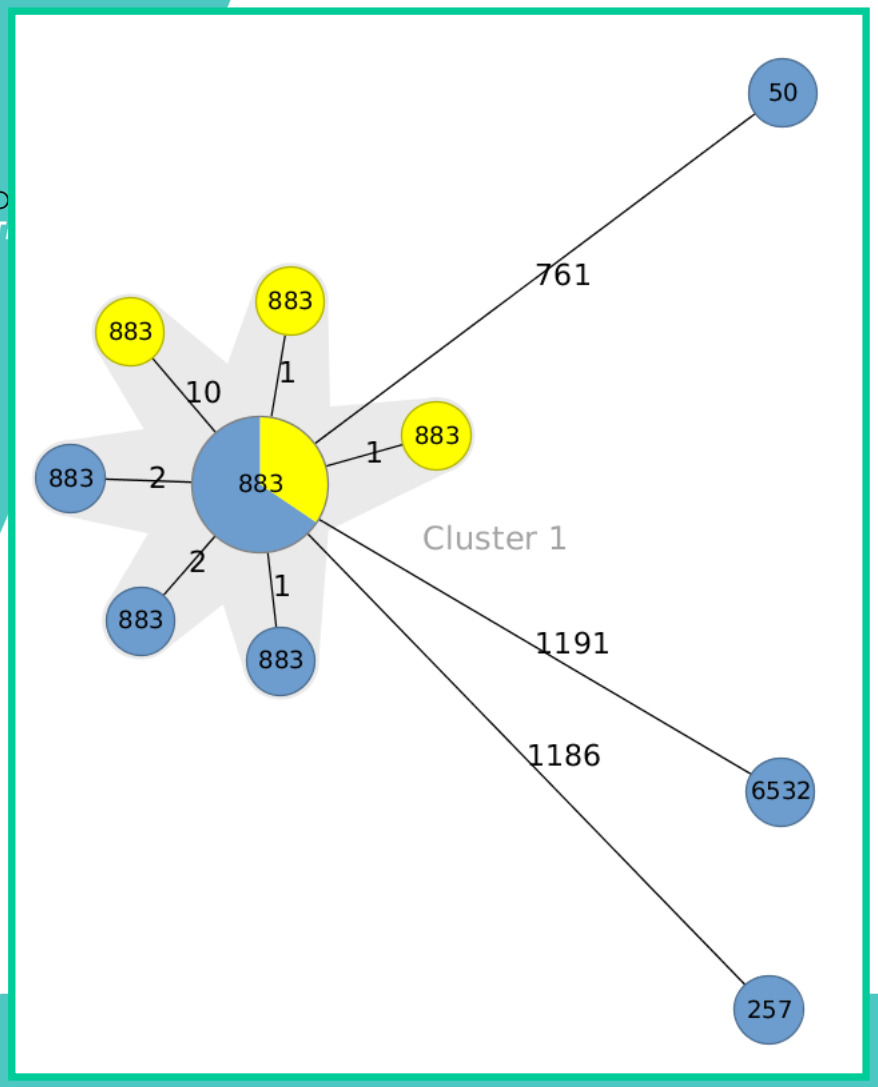
Epidemiological concordance

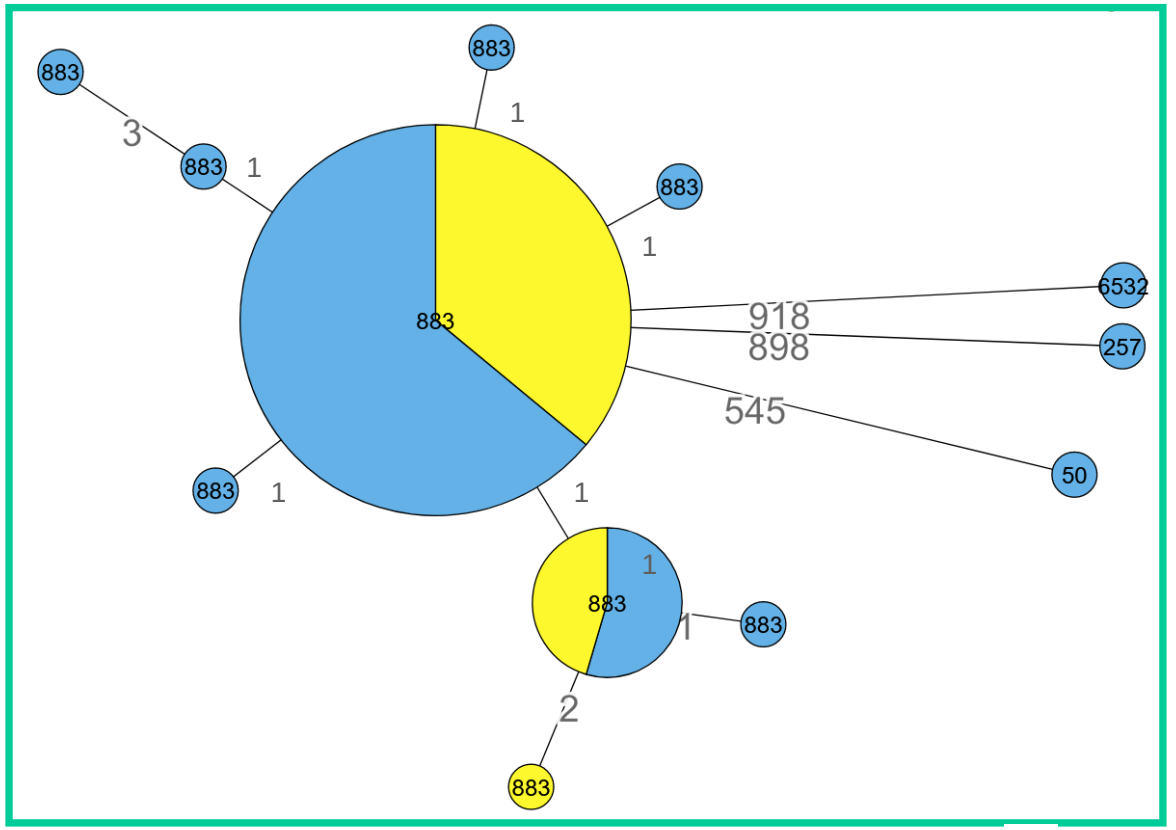
SeqSphere + v.6 (Ridom GmbH, Münster, Germany)

Schema: wgMLST_c_jejuni,1595

NGSmanager (https://github.com/genpat-it)

Schema “wgmlst_c_jejuni_2795





IZS.IT

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

- Interpreting isolate relatedness is highly organism-specific, but attempts to define species specific cutoffs is really important
- Variation seen within single strains was comparable to the known published cutoffs
- Variation between our sequencing and reference assemblies was moderate highlighting difference between reference assemblies and actual reference strains in the lab
- Maintenance of software and version testing must be always tested using the validation data and the WGS proficiency testing
- Repeatability and reproducibility and robustness were demonstrated
- The present study aimed to build evidence to recognize the accreditation status of WGS workflow for *C. jejuni* and *C. coli* sequencing

IZS TERAMO

LABORATORIO NAZIONALE DI RIFERIMENTO PER *CAMPYLOBACTER*

Thank you

Ministero della Salute

Katiuscia Zilli
Francesca Marotta
Anna Janowicz
Romina Romantini
Lisa Di Marcantonio
Federica Di Timoteo
Teresa Romualdi
Anno Abass
Giovanni Foschi
Roberta Di Romualdo
Eugenio Felicioni