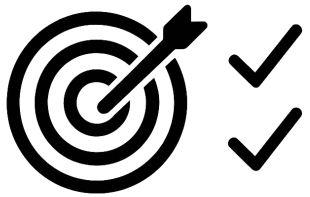
Genomic Tools in Outbreak Investigations - Case Study Cholera Outbreak

Typing Tools: From PFGE to WGS

August 27, 2025

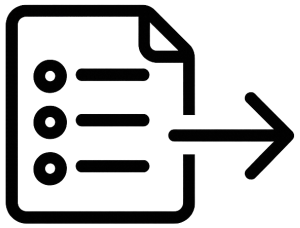
Intended Learning Objectives



Specific objectives of this session:

1. Explain why strain-level typing matters for public health and food safety (surveillance, outbreaks, source tracking).
2. Compare legacy methods (serotyping, PFGE, MLST) with WGS-based typing.
3. Differentiate SNP-based phylogenetics and gene-by-gene (cg/wgMLST), including core strengths, limitations, and when to use each.
4. Interpret SNP and allelic distances and apply context-dependent thresholds together with epidemiology.
5. Learn commonly used tools/databases for *in silico* serotyping/MLST and basic SNP/cgMLST comparisons.

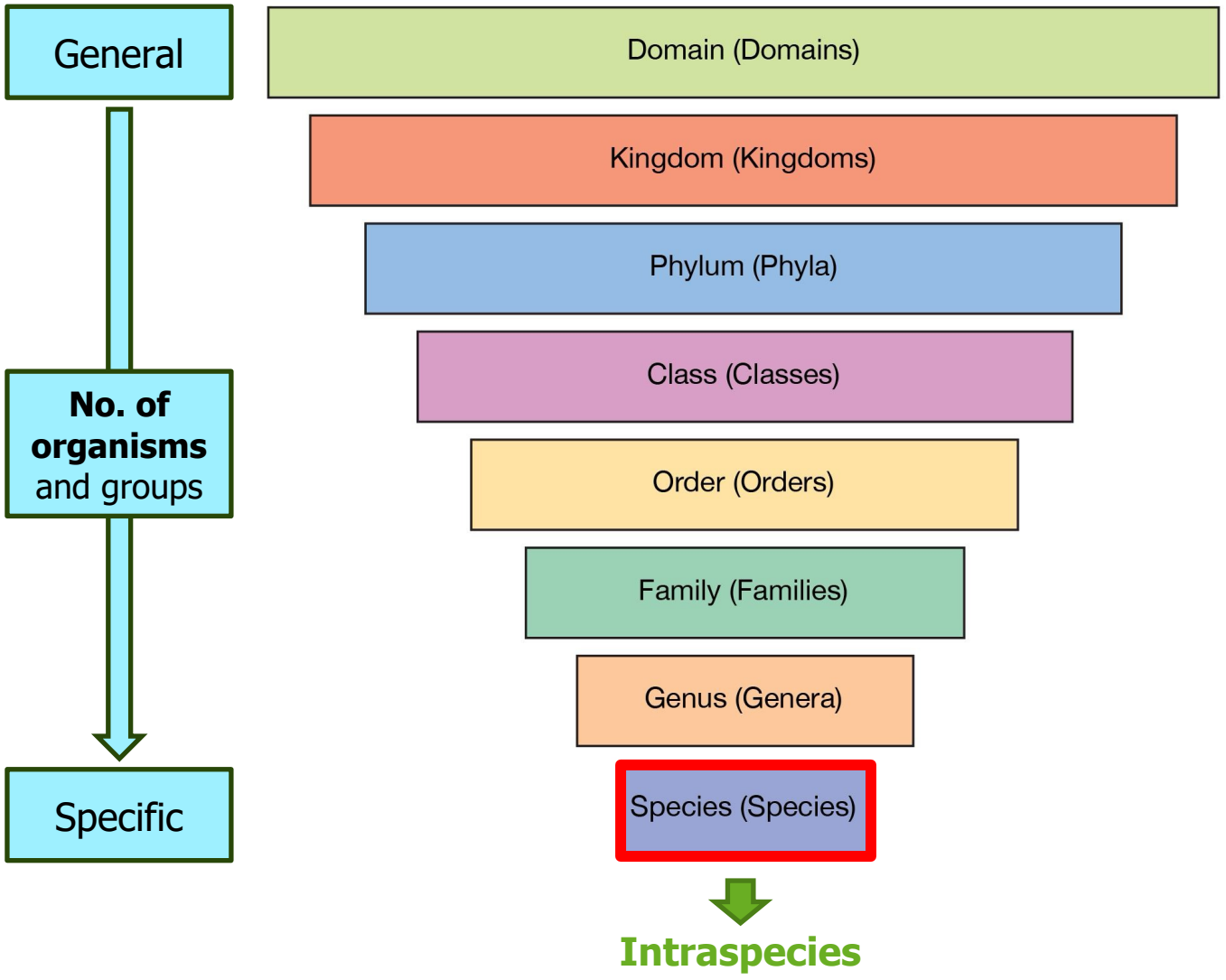
Outline



This session consists of the following elements:

1. Why species names aren't enough — what typing adds.
2. From phenotypic & legacy molecular methods to *in silico* serotyping/MLST.
3. WGS-based typing: value and the EU mandate for outbreak investigations
   - Two approaches: SNP-based phylogenetics vs gene-by-gene (cg/wgMLST)
   - When to use which: strengths, limits, thresholds (with epidemiology) + key databases/tools
4. Data analysis & thresholds

Why Species Names aren't enough



Variability

- Phenotypic features (e.g., capsule production)
- Antimicrobial resistance patterns
- Virulence properties
- Gene content (e.g., plasmids/phages)

Ferreira Wanda F. Canas 340; Microbiologia. ISBN: 978-972-757-515-2

Same Species, different Strains, different Risk

The bacterial risk to humans in most cases **resides** at the **strain level**, not the species

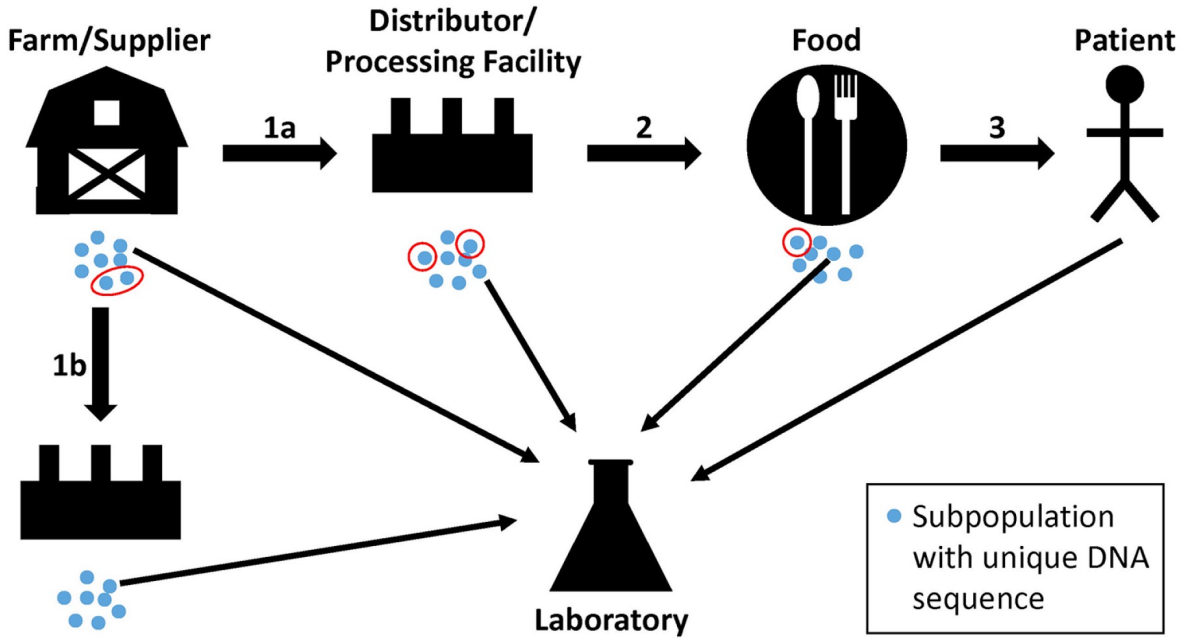


FIGURE 1 | Hypothetical pathway of bacterial pathogens along the farm-to-fork continuum. Genetically unique subpopulations of bacterial pathogens (blue dots) can grow in farms or other suppliers from reservoir populations. Representatives of one or more subpopulations can be transmitted to distributors and processing facilities, food, and to patients (red circles; steps 1a, 2, and 3). At each step, the founding bacteria may grow into multiple subpopulations. Product, environmental, or clinical samples may yield representative bacteria from closely related subpopulations that are analyzed by the food safety laboratory (thin arrows). Farms and suppliers may have more than one customer, resulting in the dissemination of genetically identical bacteria to multiple distributors or food processing facilities (step 1b).

Compare microorganisms of the **same species** from **different samples** (clinical + food + environment + others)

Guinane et al. 2019. Genomics of Foodborne Microorganisms. Chapter 35, Pightling et al. 2018. Front. Microbiol. 9

What Bacterial Typing Adds

Typing allow us to **discriminate**/**differentiate** microorganisms within the same species (intraspecies level) by evaluating the relationship between microorganisms

Typing

Antimicrobial Resistance Monitoring

- Antimicrobial stewardship
- Support patient management
- Detection of emerging antibiotic resistance mechanisms

Outbreak Investigation

- Identification of outbreaks
- Infection **Prevention**
- Infection **Control**

Epidemiological Surveillance

- Surveillance data
- Detection/notification of emerging pathogens

van Belkum et al. 2007. Clin Microbiol Infect. 3:1-46; van Rossum et al. 2020. Nat Ver Microbiol. 18(9):491-506

Typing Methods

Phenotypic

01

- **Serotyping**
- Phagotyping
- Biotyping
- Antimicrobial Susceptibility (AST)

Genotypic/ Molecular

02

- **PFGE**
- **MLST**
- MLVA
- RFLP
- FT-IR (spectrometric)

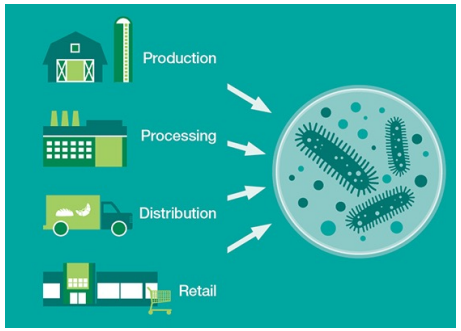
Whole Genome Sequencing

03

WGS

New "Golden Age" of Microbiology Metagenomic

04



Novais et al. 2019. Eur J Clin Microbiol Infect Dis. 38(3):427-448

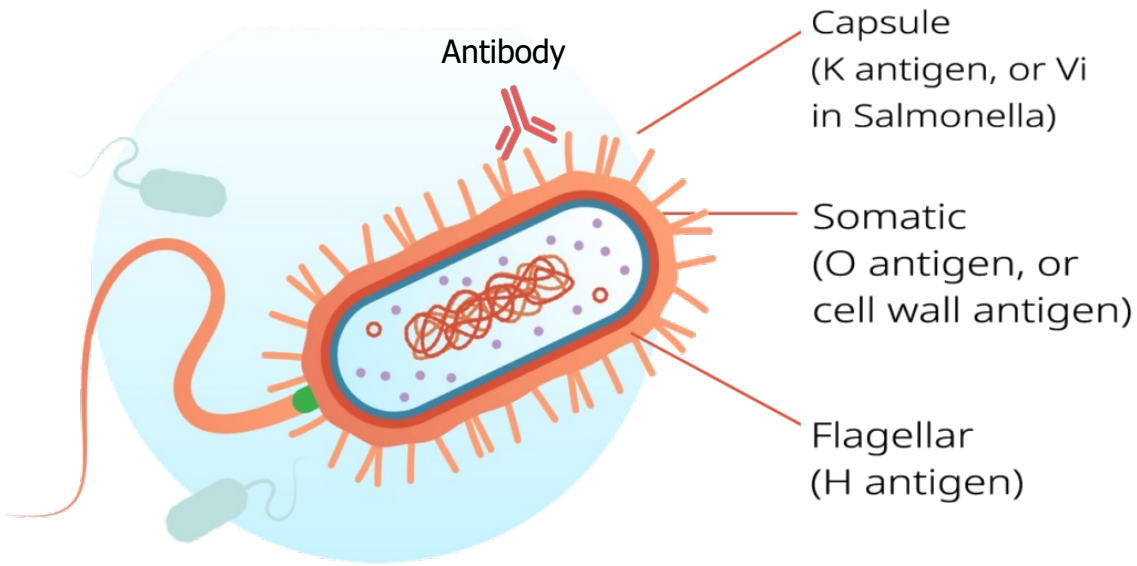
Which typing technique demonstrates the greatest ability to discriminate between strains?

Typing Methods

Phenotypic

Serotyping

Carried out by **Reference Institutes** and applied only to pathogens that are organized into **serotypes** (e.g., *Salmonella*, *E. coli* STEC, *Yersinia enterocolitica*, *Listeria monocytogenes*, some species of *Vibrio*)



Based on **antigen-antibody** agglutination reactions ➔ serotype represented by an **antigenic formula** (O:H:K)

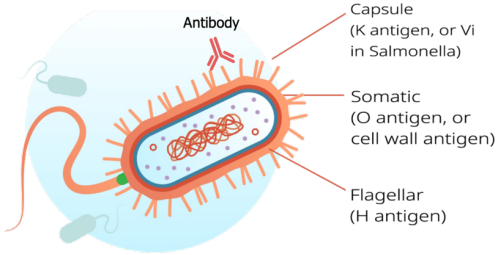
- Labor- and time-consuming
- Expensive (several sera needed)
- Restricted application

Mughini-Gras et al. 2020. Front Microbiol. 10:2578

Why *in silico* serotyping?

Addresses wet-lab serotyping limits by using WGS data to **analyze antigen loci** (O-somatic, H-flagellar, K/capsular) **directly** → portable, reproducible

WGS of the bacterial isolate



INPUT

WGS raw reads (FASTQ)

Genome assembly (FASTA)

QC

- Coverage
- Contamination
- Species check

Antigen loci detection

- O somatic genes
- H flagellar genes
- K/capsular genes

PREDICTED SEROTYPE

Scalable: works with genome raw reads or assemblies; analysis of thousands of genomes at the same time

Uelze et al. 2020. One Health Outlook. 2(3); Mughini-Gras et al. 2020. Front Microbiol. 10:2578; Image created with BioRender.com

Why *in silico* serotyping?

Table - Correlation between *in silico* serotype prediction compared to those determined by the traditional phenotypic method according to the White–Kauffmann–Le Minor scheme (*Salmonella*) or by slide agglutination using antisera to somatic O and flagella H antigens (STEC)

Reference	Pathogen	Tested Isolates	SeqSero/SeqSero2	SISTR	SeroTypeFinder
Joensen et al. (2015)	STEC	682 isolates (569 tested for O antigens and 508 for H antigens)	NA	NA	98.4% agreement (O antigens) 99.2% agreement (H antigens)
Chattaway et al. (2016)	STEC no O157	102 isolates	NA	NA	96.1% agreement (O antigens) 100% agreement (H antigens)
Zhang et al. (2015)	*Salmonella*	3,061 isolates (GenomeTrakr)	92.6% agreement	-	NA
	Salmonella	304 isolates (CDC)	98.7% agreement	-	NA
	Salmonella	324 isolates (public genomes)	91.5% agreement	-	NA
Yoshida et al. (2016)	*Salmonella*	4,188 isolates (public genomes)	-	94.6% agreement	NA
Ibrahim et al. (2018)	*Salmonella*	1,041 isolates (clinically-relevant)	86.4% agreement	-	NA
Robertson et al. (2018)	*Salmonella*	42,400 isolates (public genomes)	-	91.9% agreement	NA
Hendriksen et al. (2018)	*Salmonella*	786 isolates (APHA, DTU, PHE)	65% agreement	88% agreement	NA
Zhang et al. (2019)	*Salmonella*	2,280 isolates (NARMS - CDC)	97.6% agreement	97.7% agreement	NA

Abbreviations: NA, Not Applicable; STEC, Shiga toxin-producing *Escherichia coli*

Agreement >90% in most cases!

Common tools & databases (examples)

Table 3 List of different tools for in silico serovar prediction

Tool (Reference)	Species	Method	Source
SeqSero [137]	*Salmonella* spp.	mapping of raw reads to database of O and H antigen alleles	command line: https://github.com/denglab/SeqSero Web tool: http://denglab.info/SeqSero
SeqSero2 [138]	*Salmonella* spp.	similar SeqSero, additional performs rapid serotype prediction based on unique k-mers of serotype determinants	command line: https://github.com/denglab/SeqSero2
Salmonella TypeFinder	*Salmonella* spp.	in addition to running SeqSero, determines 7 gene MLST and infers serovar from sequence type according to Enterobase database	Web tool: https://cge.cbs.dtu.dk/services/SalmonellaTypeFinder/
SISTR [139]	*Salmonella* spp.	mapping of database of O and H antigen alleles against genome assembly, additionally, determines cgMLST and infers serovar from cgMLST clustering result	command line: https://github.com/phac-nml/SISTR_cmd Web tool: https://lfz.corefacility.ca/sistr-app/
MOST [140]	*Salmonella* spp.	determines 7-gene MLST and reports number of respective serovar – ST matches from the PHE/Achtmann database	command line: https://github.com/phe-bioinformatics/MOST
Serotype Finder [141]	*E. coli*	mapping of raw reads to database of O and H antigen alleles, or mapping of antigen alleles against genome assembly	command line: https://github.com/Papos92/ecoli_serotyper Web tool: https://cge.cbs.dtu.dk/services/SerotypeFinder/
ECTyper	*E. coli*	Information not available	command line: https://github.com/phac-nml/ecoli_serotyping
EBEis from the Enterobase Tool Kit	*E. coli Shigella* spp.	mapping of database of O and H antigen alleles against genome assembly	command line: https://github.com/zheminzhou/EToKi#ebeis%2D%2D-in-silico-serotype-prediction-for-escherichia-coli%2D%2Dshigella-spp
Seq_typing	*E. coli*	mapping of raw reads to database of O and H antigen alleles, or mapping of antigen alleles against genome assembly	command line: https://github.com/B-UMMI/seq_typing
LisSero [134]	*Listeria monocytogenes*	mapping of database of 5 DNA regions (*lmo1118, lmo0737, ORF2110, ORF2819, prs*) against genome assembly	command line: https://github.com/MDU-PHL/LisSero

Other less common:

- ***Vibrio parahaemolyticus*** – O and K antigens prediction via Kaptive database and the VPsero tool (doi:10.3389/fmicb.2021.620224)
- ***Vibrio cholerae*** – O antigen prediction via the VicPred tool (doi:10.3389/fmicb.2021.691895)
- **Other vibrios** (e.g., *V. vulnificus*) – routine *in-silico* serotyping support is limited

Uelze et al. 2020. One Health Outlook. 2(3); Graaf-van Bloois et al. 2023. Microb Genom. 9(5):mgen001007

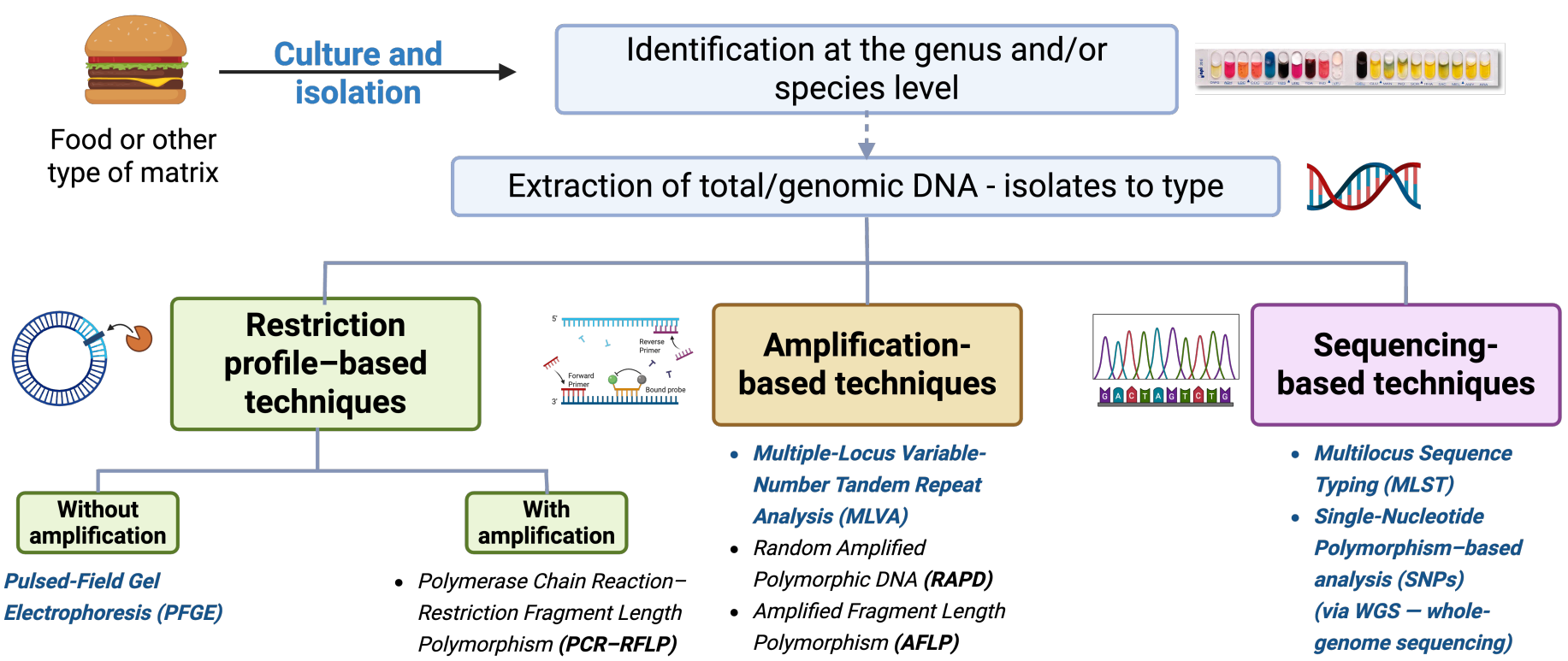
Typing Methods

Genotypic or Molecular

Molecular Typing Methods

Strain discrimination from **genetic content** — **DNA** is **unique** to each microorganism

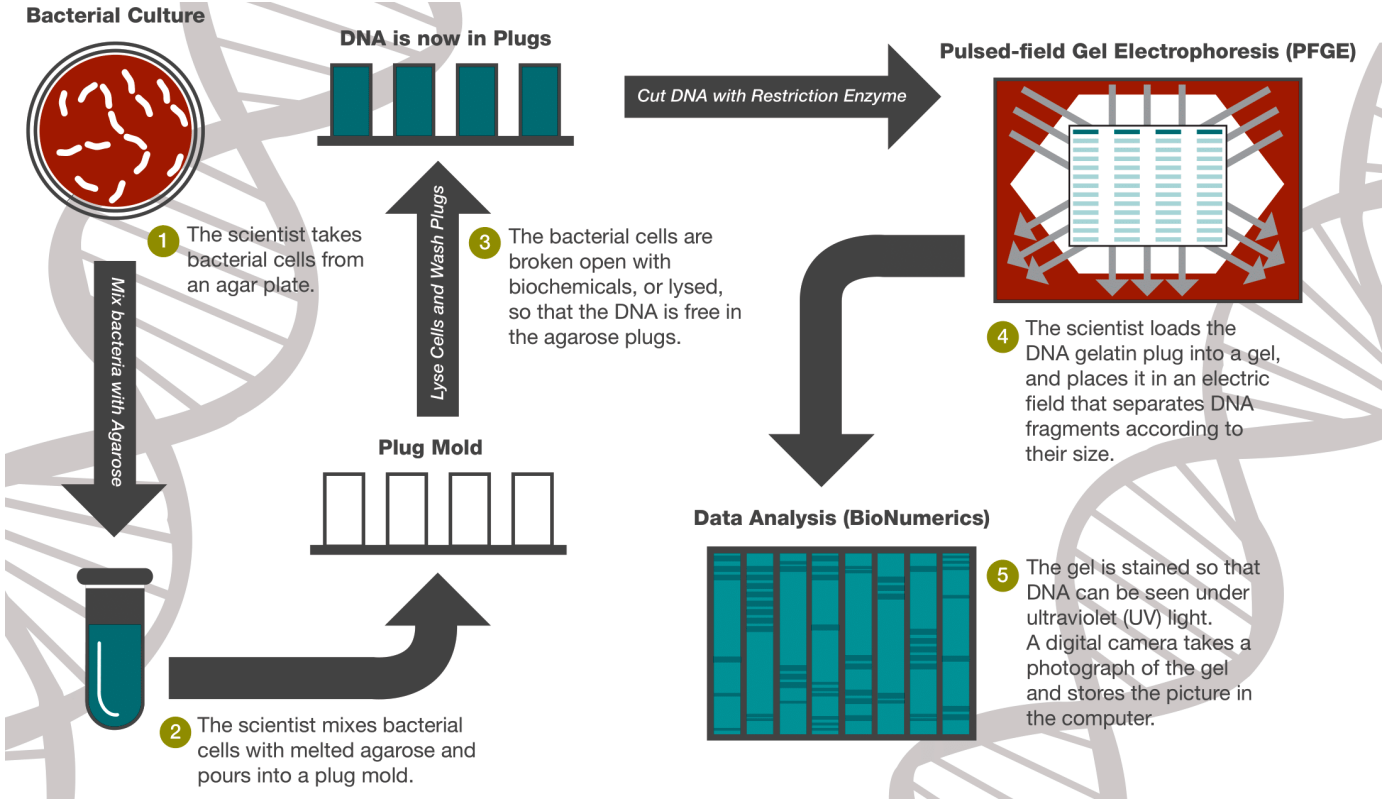
Except for WGS (SNPs – based analysis), discrimination of molecular methods is species/scheme-dependent



Sabat et al. 2013. Euro Surveill. 18(4):):pii=20380; Image created with BioRender.com

Pulsed-Field Gel Electrophoresis - PFGE

Divides a bacterial population into **pulsotypes**: macro-restriction of specific DNA sites with **restriction enzymes**, followed by separation of the fragments by pulsed-field gel electrophoresis.

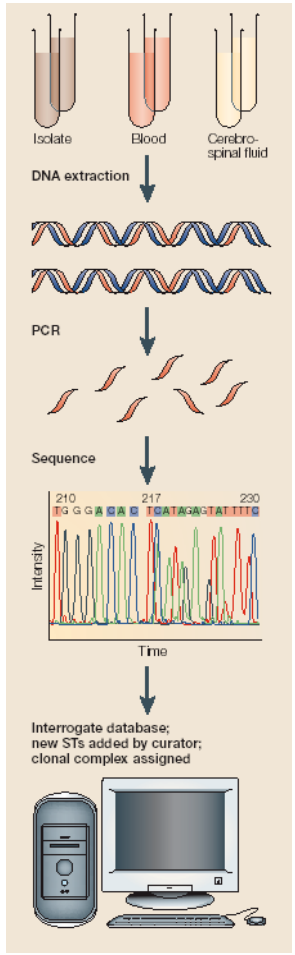


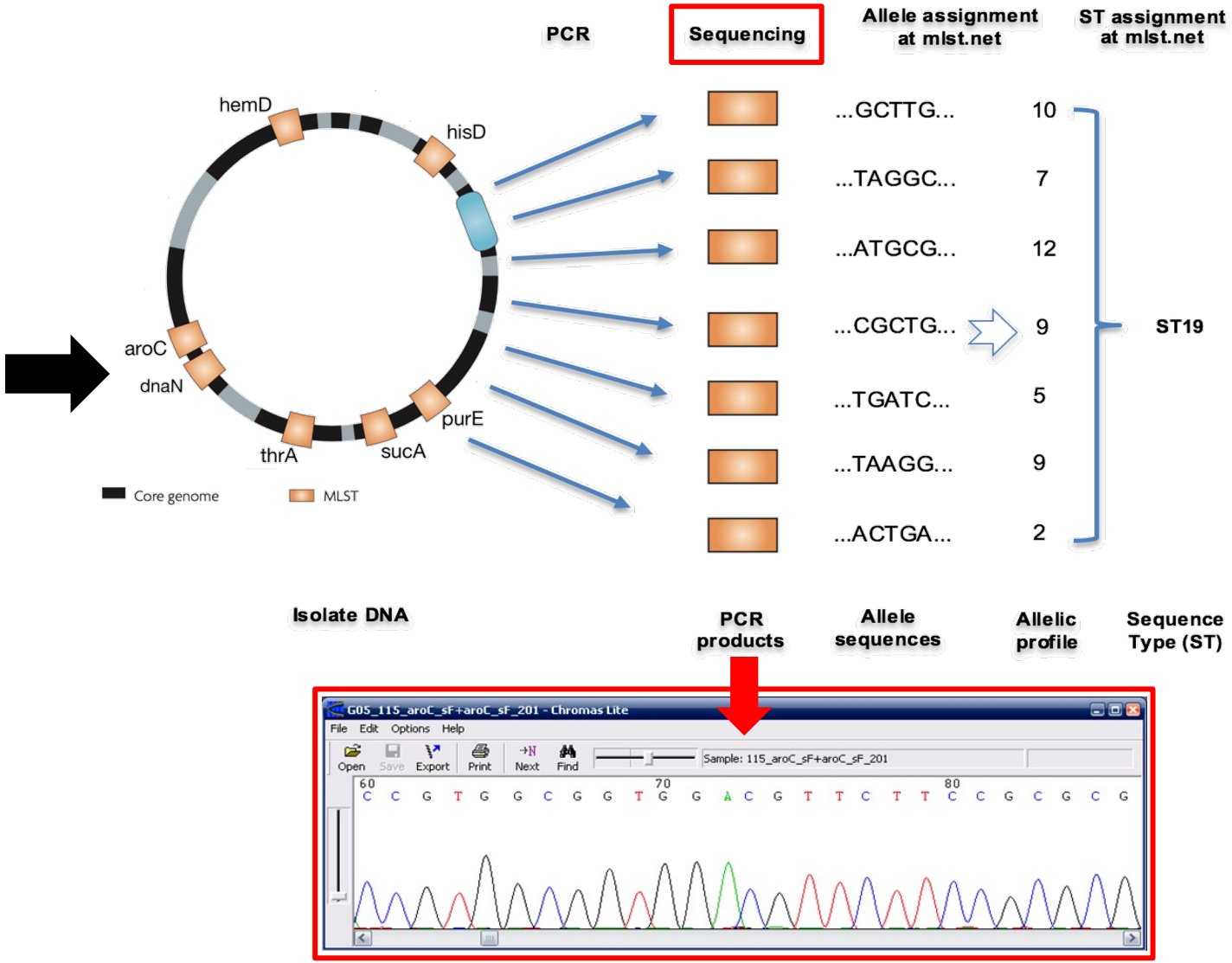
- Time-consuming assay (≈3–5 days)
- Limited inter-laboratory reproducibility
- Low throughput (limited number of strains per run)
- Difficult to interpret without software
- Profiles depend on enzyme and run conditions

Largely replaced by WGS for surveillance

Li et al. 2009. FEMS Microbiol Rev. 33(5):892-916, EFSA Journal. 2019. 17(12):5898, https://www.cdc.gov/pulsenet/pathogens/pfge.html

Multilocus Sequence Typing - MLST





- Detects sequence variation in ~7 housekeeping genes — conserved, but with some diversity
- Each unique allelic profile is assigned a **Sequence Type (ST)**

✅ Widely used, standardized nomenclature

❌ Limited resolution, labor-intensive, costly

Online Salmonella MLST minimum spanning tree (MST): https://lfz.corefacility.ca/sistr-app/ (select Phylogeny and MST)

In silico MLST

Image created with ChatGPT

Traditional MLST has largely been **replaced** by ***in silico* MLST**, where allele calls and sequence types are extracted directly from whole-genome sequencing (WGS) data

Genome-based *Salmonella* serotyping as the new gold standard

SCIENTIFIC REPORTS
nature research

Sangeeta Banerji[1,3], Sandra Simon[1,3], Andreas Tille[2], Angelika Fruth[1] & Antje Flieger[1*]

Serotype	Sequenced Isolates	Correlation			Prediction failure			Miscorrelation		
		Seq-Sero	MLST	Seq-Sero+MLST	Seq-Sero	MLST	Seq-Sero + MLST	Seq-Sero	MLST	Seq-Sero+MLST
Agona	3	3	3	3	0	0	0	0	0	0
Choleraesuis	33	0	33	33	3	0	0	0	0	0
Choleraesuis monophasic	3	0	0	0	0	0	0	3	3	3
Derby	55	55	55	55	0	0	0	0	0	0
11:z41:e,n,z15 (novel serovar)	10	10	10	10	0	0	0	0	0	0
Enteritidis	115	115	115	115	0	0	0	0	0	0
Infantis	50	49	50	50	1	0	0	0	0	0
Kentucky	7	7	7	7	0	0	0	0	0	0
Kintambo	3	0	3	3	0	0	0	0	0	0
Kottbus	12	0	12	12	0	0	0	0	0	0
Mbandaka	15	15	15	15	0	0	0	0	0	0
Mikawasima	10	10	10	10	0	0	0	0	0	0
Muenchen	25	0	25	25	0	0	0	0	0	0
Paratyphi B	6	6	6	6	0	0	0	0	0	0
Paratyphi B monophasic	1	1	1	1	0	0	0	0	0	0
Total (%)	100.0	84.0	95.2	99.4	1.0	0	0	1.0	4.8	0.6
Overall (%)	100.0	98.0	95.2	99.4	1.0	0	0	1.0	4.8	0.6
Strathcona	2	2	2	2	0	0	0	0	0	0
Stourbridge	14	14	14	14	0	0	0	0	0	0
Sundsvall	1	0	1	1	0	0	0	0	0	0
Typhi	74	74	74	74	0	0	0	0	0	0
Typhimurium biphasic	52	52	32	52	0	0	0	0	20	0
Typhimurium monophasic	19	17	17	19	0	0	0	2	2	0
Serologically rough	6	5	6	6	1	0	0	0	0	0
Total number	520	437	495	517	5	0	0	5	25	3
Total (%)	100.0	84.0	95.2	14.0	1.0	0	0	1.0	4.8	0.6
Overall (%)	100.0	98.0	95.2	99.4	1.0	0	0	1.0	4.8	0.6

Banerji et al. 2020. Sci Rep. 10:4333

MLST Databases

EnteroBase

Database (URL)	Pathogen (example)	Scheme
Enterobase (https://enterobase.warwick.ac.uk/)	*Escherichia/Shigella*	7 genes
	Salmonella	7 genes
	Yersinia	7 genes
	Moraxella	7 genes
BIGSdb (https://bigsdb.pasteur.fr/klebsiella/primers-used/)	*Klebsiella pneumoniae*	7 genes

PubMLST → currently supports MLST schemes for **>100 different bacterial species**, including of clinical relevance

PubMLST Public databases for molecular typing and microbial genome diversity

Database (URL)	Pathogen (example)	Scheme
PubMLST (https://pubmlst.org/organisms)	*Campylobacter jejuni/coli*	7 genes
	Clostridioides difficile	7 genes
	Clostridium perfringens	8 genes
	Cronobacter spp.	7 genes
	Enterococcus faecium/faecalis	7 genes
	Haemophilus influenzae	7 genes
	Pseudomonas aeruginosa	7 genes
	Staphylococcus aureus	7 genes
	Streptococcus pneumoniae	7 genes
	Vibrio spp. (other available)	4 genes

https://enterobase.warwick.ac.uk/; https://bigsdb.pasteur.fr/; https://pubmlst.org/

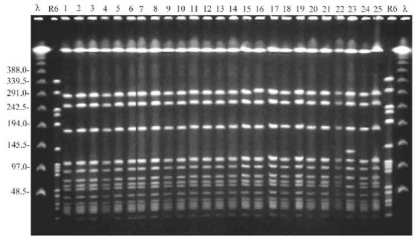
Typing Methods

Whole Genome Sequencing(WGS)-based

Whole Genome Sequencing-based Typing

Allow us to gain unprecedented understanding of **similarities** and **differences** between isolates of the same species

Traditional typing methods
(e.g., serotyping, PFGE, MLVA, MLST)



- Difficult to automate
- Limited digital integration
- Low/moderate power of discrimination
- Specific for each microorganism
- Time-consuming

Methods based on **(total) DNA sequencing**
(e.g., SNPs, cgMLST, wgMLST)

- **Compare "the whole" available genome**
- **High accuracy** and **discrimination** power
- Applicable **across organisms**
- **Data** can be **shared** between laboratories
- Cost and time – already competitive
- Enables global real-time surveillance

Guinane et al. 2019. Genomics of Foodborne Microorganisms. Chapter 35, Pightling et al. 2018. Front. Microbiol. 9

Whole Genome Sequencing-based Typing

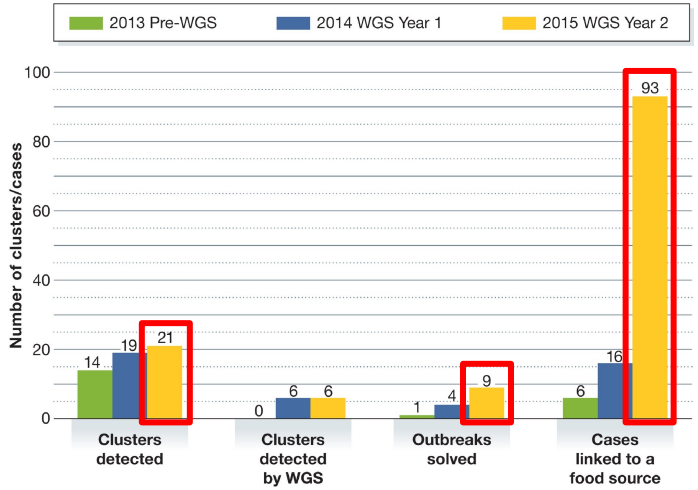
PulseNet International
Connects the Dots for FOOD SAFETY

WGS is **replacing traditional methods** to type foodborne pathogens, as recommended by multiple organizations for outbreak detection, surveillance, and source tracking

REVIEW ARTICLES

PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance

PulseNet International is a global network dedicated to laboratory-based surveillance for food-borne diseases. The network comprises the national and regional laboratory networks of Africa, Asia Pacific, Canada, Europe, Latin America and the Caribbean, the Middle East, and the United States. The PulseNet International vision is the standardised use of whole genome sequencing (WGS) to identify and subtype food-borne bacterial pathogens worldwide, replacing traditional methods to strengthen preparedness and response, reduce global social and economic disease burden, and save lives. To meet the needs of real-time surveillance, the PulseNet International network will standardise subtyping via WGS using whole genome multilocus sequence typing (wgMLST), which delivers sufficiently high resolution and epidemiological concordance, plus unambiguous nomenclature for the purposes of surveillance. Standardised protocols, validation studies, quality control programmes, database and nomenclature development, and training should support the implementation and decentralisation of WGS. Ideally, WGS data collected for surveillance purposes should be publicly available, in real time where possible, respecting data protection policies. WGS data are suitable for surveillance and outbreak purposes and for answering scientific questions pertaining to source attribution, antimicrobial resistance, transmission patterns, and virulence, which will further enable the protection and improvement of public health with respect to food-borne disease.



Listeria monocytogenes **surveillance pre and post WGS implementation**

It offers higher **accuracy** and **discrimination** than PFGE (old gold standard) — detecting more outbreaks, tracing contamination sources, and identifying sporadic cases

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202500179

WGS Mandated for EU Outbreak Investigations

Official Journal
of the European Union

EN
L series

2025/179

3.2.2025

COMMISSION IMPLEMENTING REGULATION (EU) 2025/179

of 31 January 2025

on the collection and transmission of molecular analytical data within the frame of epidemiological investigations of food-borne outbreaks in accordance with Directive 2003/99/EC of the European Parliament and of the Council

(3) Whole genome sequencing (WGS) is a modern molecular analytical technique for microbiological studies which facilitates greatly the swift identification of clusters of microorganisms, supporting the epidemiological investigations. It enables to establish links between isolates of food-borne pathogens recovered from humans, food, animals, feed and the related environment during the outbreak investigation.

(4) To substantially facilitate food-borne outbreak investigations and the timely detection of the sources of those outbreaks, Member States should be required to collect *Salmonella enterica*, *Listeria monocytogenes*, *Escherichia coli*, *Campylobacter jejuni* and *Campylobacter coli* isolates derived from food, animal, feed and related environmental samples from food and feed business operators and during official controls, where those isolates are associated or suspected to be associated with a food-borne outbreak. Member States should also be required to carry out WGS on those isolates.

2. The competent authority shall carry out WGS on at least one isolate of each serovar, biotype or molecular type of the collected isolates referred to in paragraph 1 of this Article in an official laboratory which is referred to in Article 37 of Regulation (EU) 2017/625 of the European Parliament and of the Council ([5]) and which is accredited for this method according to ISO 17025.

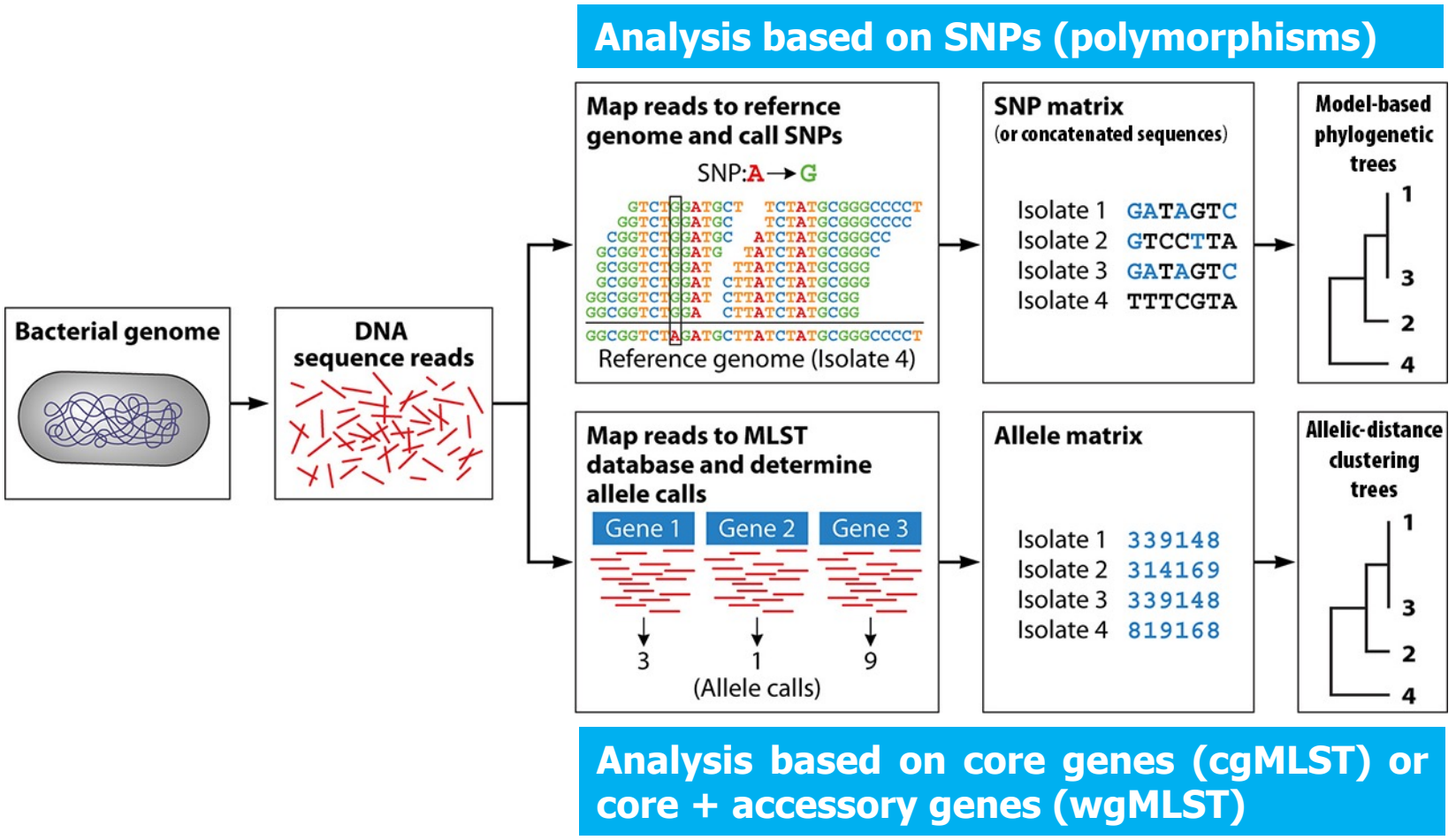
WGS now mandatory in EU

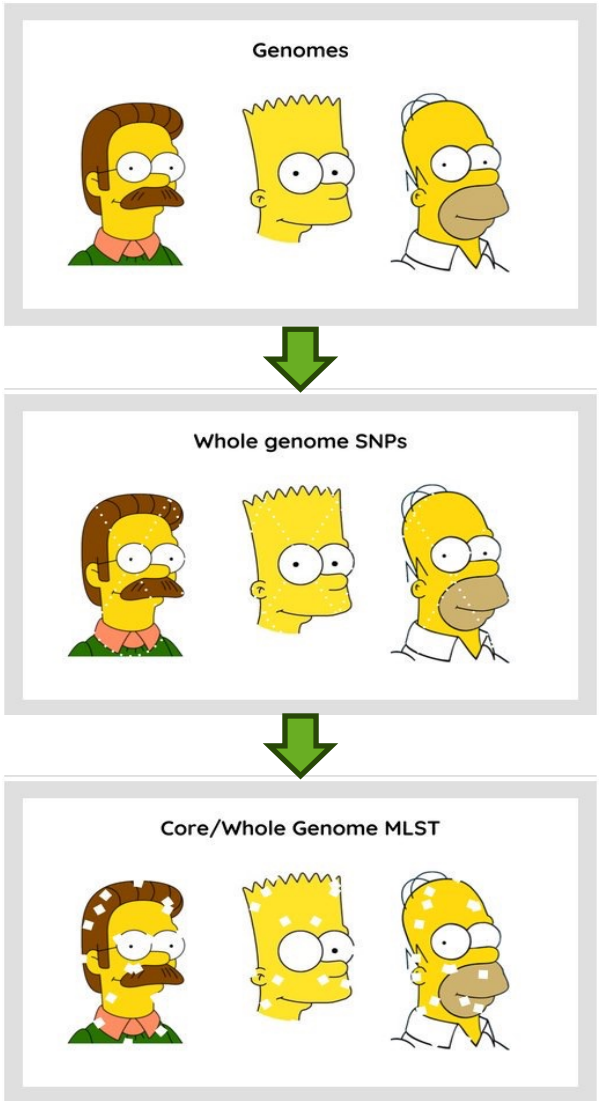
WGS-based typing has become **mandatory** in the **EU** for foodborne **outbreak investigations** of major foodborne pathogens

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202500179

WGS-based Typing Approaches

WGS enables strain comparison through two main approaches: **SNP**-based phylogenetics and **gene-by-gene** allele profiling





EFSA Journal 2019. 17(12):5898, Timme et al. 2019. J Clin Microbiol. 57(5):e01816 (adapted figure)

SNP-based Phylogenetics

Map reads to a close reference genome → call high-quality SNPs, optionally remove recombined regions, infer distances, and build phylogenetic/evolutionary trees

Strengths

- ✅ **Highest resolution** for very closely related isolates (e.g., outbreak delimitation)
- ✅ Supports **time-scaled phylogenies** using mature frameworks
- ✅ Can use raw reads (**no assembly needed**)

Limitations

- ⚠️ **Reference bias:** requires a suitable, close reference
- ⚠️ **Sensitive** to **HGT** and **recombination** (must be masked)
- ⚠️ Harder to standardize across labs and tools (**less portable**)

Use When…

- ➡️ You suspect a **recent outbreak** (e.g., hospital cluster) or rapid transmission
- ➡️ You have **high coverage data** and a good reference genome
- ➡️ You need **maximum resolution** and detailed phylogenetic context

Wick et al. 2025. Access Microbiol. 7(5):001025.v3; HGT, Horizontal Gene Transfer

Gene-by-gene Methods

Call alleles at predefined loci → compare isolates by allelic profiles/distances, build clustering trees

• **cgMLST**: core loci shared by most strains • **wgMLST**: core + accessory loci

Strengths

- ✅ **Standardized** and highly **portable** across centers and time
- ✅ **More tolerant** to **recombination**; doesn't require a reference genome
- ✅ Intuitive reporting (STs, allele differences); easy database integration

Limitations

- ⚠ Slightly lower maximum resolution than high-quality SNP analyses
- ⚠ Requires curated **scheme**; allele calling can suffer with low coverage/fragmented assemblies, missed loci
- ⚠ Different schemes aren't directly comparable

Use When…

- ➡ **Surveillance** & **early outbreak** detection (triage), multi-center studies, long timescales
- ➡ Species with **high recombination**
- ➡ Need easy, reproducible clustering and **cross-lab comparability**

Uelze et al. 2020. One Health Outlook. 2(3)

Available Databases for WGS-based Typing

Free platforms (e.g., EnteroBase, PubMLST/BIGSdb, INNUENDO) support **SNPs**, **cgMLST** or **wgMLST** analyses, but **schemes differ** (locus sets, allele definitions, thresholds) across databases and organisms

🚨 Report **reference** + **filters (SNP)** and **scheme** + **version** + **% loci called (cg/wgMLST)** for comparability

Table 4 Key characteristics of selected platforms

Platform	Reference	Typing approach	Central instance/ Local instance	Commercial (C)/ Academic (A)
BIGSdb	[151]	wg/cgMLST	Both possible	A
Bionumerics	http://www.applied-maths.com/applications/wgmlst	wg/cgMLST and SNP	Both possible	C
CGE	[152]	cgMLST and SNP	Cloud	A
COMPARE	[153]	cgMLST and SNP	Cloud	A
Enterobase	[16]	wg/cgMLST and SNP	Cloud	A
INNUENDO	[154]	wg/cgMLST	Local	A
IRIDA	[155]	wg/cgMLST and SNP	Local	A
NCBI Pathogens	[156]	wg/cgMLST and SNP	Cloud	A
PathogenWatch	https://pathogen.watch	cgMLST	Cloud	A
PATRIC	[68, 150]	SNP	Cloud	A
SeqSphere+	https://www.ridom.de/seqsphere/	wg/cgMLST	Local	C

EXAMPLES:

Campylobacter jejuni, C. coli, L. monocytogenes, Vibrio

Schemes for almost all bacteria

E. coli, Salmonella enterica, Yersinia, Vibrio, Clostridioides

Klebsiella

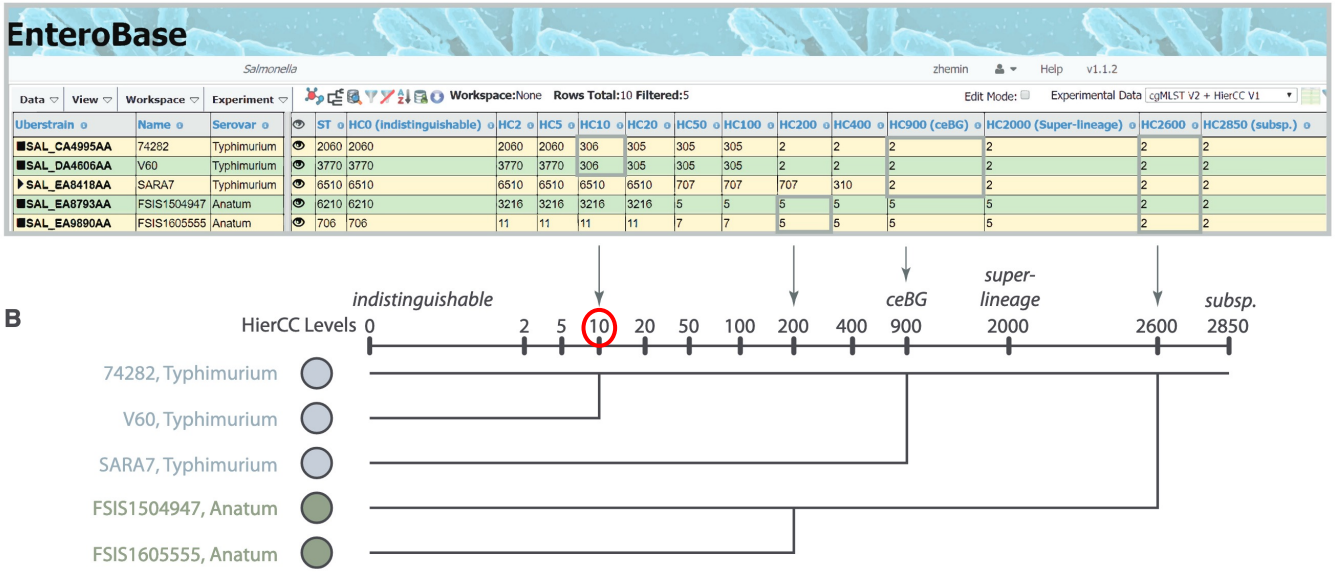
Schemes for almost all bacteria

Uelze et al. 2020. One Health Outlook. 2(3)

Data Analysis

Depending on the method used (**SNPs**, **cgMLST**, or **wgMLST**) and the purpose of the analysis, **different discrimination levels/cut-offs can be applied** to determine whether two or more isolates are related or not

A Example of cluster definition based on the ***Salmonella*** cgMLST scheme from Enterobase



B

Investigation of outbreaks (cgMLST) – frequently used up to 10 allelic differences ➔ **always consider epidemiological information**

TABLE 2 | Conditions used to determine whether whole-genome sequence analyses support a match between two or more genomes.

	Supports	Neutral	Does not support
SNP distance	<21	21–100	>100
Bootstrap support	>0.89	0.80–0.89	<0.80
Tree topology	Monophyletic	Paraphyletic	Polyphyletic

Investigation of foodborne disease outbreaks and contamination events (SNPs) – proposed use of up to 20 SNP differences ➔ **always consider epidemiological information**

Zhou et al. 2020. Genome Res. 30(1):138-152, Pightling et al. 2018. Front. Microbiol. 9:1482

Data Analysis

Review

Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches

A.C. Schürch [1], S. Arredondo-Alonso [1], R.J.L. Willems [1], R.V. Goering [2,*]

Clinical Microbiology and Infection 24 (2018) 350–354

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Table 1
Examples of relatedness criteria for wg/cgMLST and SNP typing schemes of representative clinically relevant bacteria

Organism	Relatedness threshold[a]		References
	wg/cgMLST	SNPs	
Acinetobacter baumannii	≤8	≤3	[25,26]
Brucella spp.	Epidemiologic validation in progress[b]		http://www.applied-maths.com/applications/wgmlst
Campylobacter coli, C. jejuni	≤14	≤15	[27,28]
Cronobacter spp.	Epidemiologic validation in progress[b]		http://www.applied-maths.com/applications/wgmlst
Clostridium difficile	Epidemiologic validation in progress[b]	≤4	[29], http://www.cgmlst.org/ncs, http://www.applied-maths.com/applications/wgmlst
Enterococcus faecium	≤20	≤16	[30]
Enterococcus raffinosus	Epidemiologic validation in progress[b]		http://www.applied-maths.com/applications/wgmlst
Escherichia coli	≤10	≤10	[31,32], https://enterobase.warwick.ac.uk/
Francisella tularensis	≤1	≤2	[33,34]
Klebsiella oxytoca	Epidemiologic validation in progress[b]		http://www.applied-maths.com/applications/wgmlst
Klebsiella pneumonia	≤10	≤18	[35,36]
Legionella pneumophila	≤4	≤15	[37]
Listeria monocytogenes	≤10	≤3	[38,39]
Mycobacterium abscessus		≤30	[40]
Mycobacterium tuberculosis	≤12	≤12	[41]
Neisseria gonorrhoeae	Epidemiologic validation in progress[b]	≤14	[42], http://www.applied-maths.com/applications/wgmlst
Neisseria meningitidis	Epidemiologic validation in progress[b]		http://www.cgmlst.org/ncs
Pseudomonas aeruginosa	≤14	≤37	[31,43]
Salmonella dublin	Epidemiologic validation in progress[b]	≤13	[44], https://enterobase.warwick.ac.uk/
Salmonella enterica	Epidemiologic validation in progress[b]	≤4	[45], http://www.cgmlst.org/ncs, http://www.applied-maths.com/applications/wgmlst, https://enterobase.warwick.ac.uk/
Salmonella typhimurium	Epidemiologic validation in progress[b]	≤2	[46], https://enterobase.warwick.ac.uk/
Staphylococcus aureus	≤24	≤15	[47,48]
Streptococcus suis		≤21	[49]
Vibrio parahaemolyticus	≤10		[50]
Yersinia spp.	0		[51]

cg, core genome; MLST, multilocus sequence typing; SNP, single nucleotide polymorphism; wg, whole genome.
[a] Data often represent single studies that can be used to begin formulation of species-specific interpretation criteria. Thus, these data should be coupled with newly published similar studies to ensure that resulting values are not atypical and can be generally applied.
[b] Proposed wg/cgMLST schemes are available online (http://www.cgmlst.org/ncs, http://www.applied-maths.com/applications/wgmlst, https://enterobase.warwick.ac.uk/) but as yet have not been epidemiologically validated.

In summary

List of learning points in this section:

1. Typing = intraspecies (strain-level) discrimination, essential for surveillance, source tracking, and outbreak investigation; WGS can also support AMR monitoring.
2. In the EU, WGS is mandatory for food-borne outbreak investigations (Regulation EU 2025/179)
3. Two complementary WGS approaches:
   - SNP-based phylogenetics → maximum resolution; requires a close reference and recombination handling.
   - Gene-by-gene (cg/wgMLST) → portable and standardized; scheme-dependent and slightly lower max resolution.
4. Always pair genomics with epidemiology — distance cut-offs are context-dependent.
5. Report methods clearly for comparability: reference + filters (SNP) and scheme + version + % loci called (cg/wgMLST).

Acknowledgements

The creation of this training material was commissioned by ECDC to Technical University of Denmark (DTU) with the direct involvement of Joana Mourão (Postdoc at DTU, PharmD, PhD – joam@food.dtu.dk)