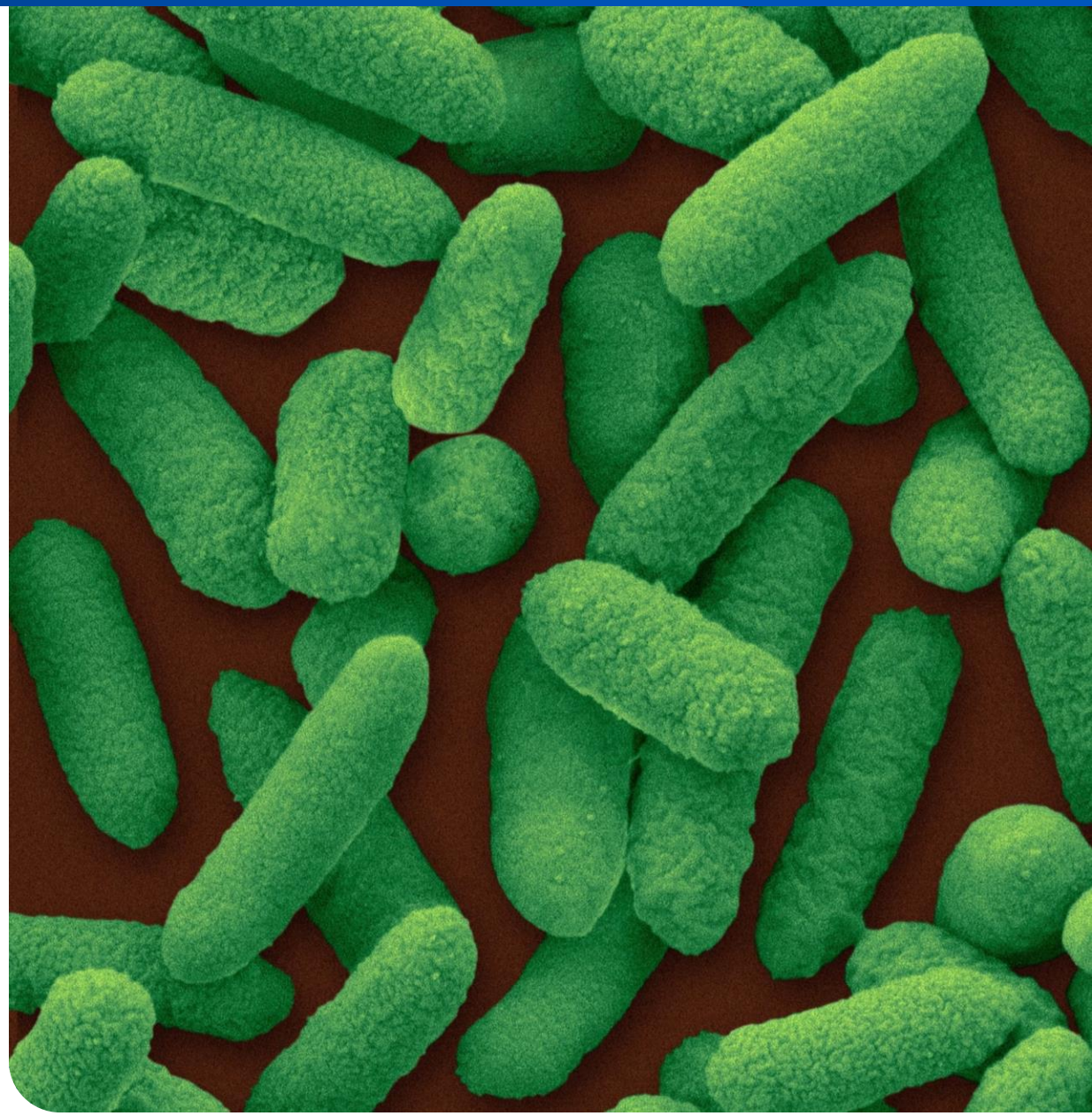
Core Genome and Whole Genome Multilocus Sequence Typing of *Cronobacter* Isolates

2026 GenomeTrakr Meeting
College Park, MD



Lavin A. Joseph, DrPH, MS
EDLB Validation and Verification Lead
Microbiologist | Campy and Cronobacter Database Coordinator
Division of Foodborne, Waterborne, and Environmental Diseases
National Center for Emerging and Zoonotic Infectious Diseases



Why *Cronobacter* surveillance matters

Practical question: Can public-health agencies use whole genome sequencing (WGS) data to rapidly detect clinical *Cronobacter* cases and identify potential associations with environmental isolates, enabling risk-mitigation measures to protect vulnerable infants?

2–10

voluntarily reported cases annually since 1979

44%

reported case-fatality rate up to

2024

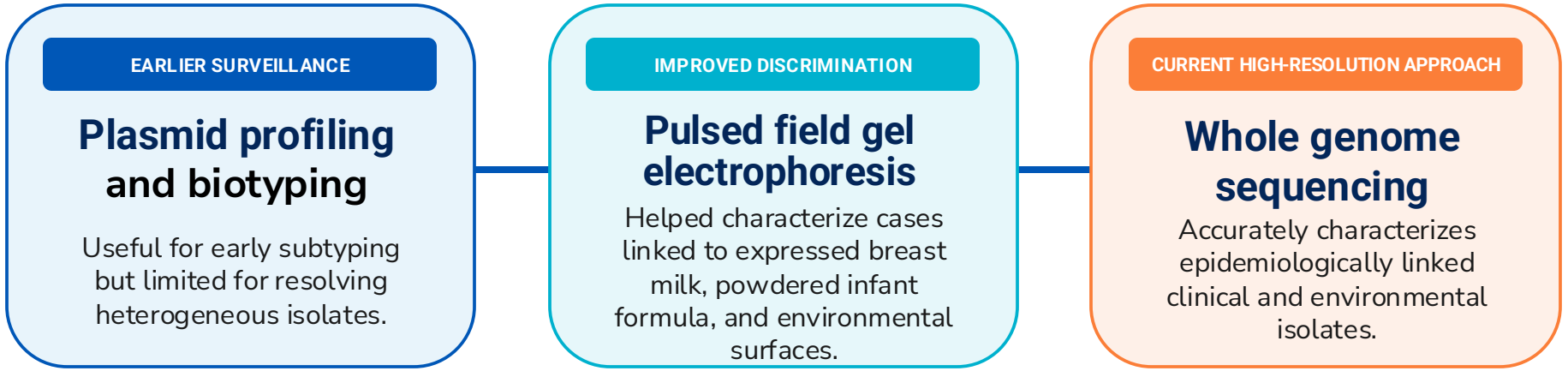
invasive infant cases became nationally notifiable by CSTE

- *Cronobacter* are gram-negative opportunistic pathogens; *C. sakazakii* and *C. malonaticus* cause most clinical infections.
- Severe disease includes meningitis and sepsis in infants <12 months, especially those with very low birth weight.
- Epidemiologically associated isolates have been recovered from powdered infant formula, breast milk/breast-pump equipment, and home environments.
- National notification plus earlier subtyping limitations created the need for standardized WGS surveillance through PulseNet.

Source: Joseph et al., Microbiology Spectrum, 2026

Evolution of *Cronobacter* subtyping

Surveillance methods moved from lower-resolution phenotypic/fragment methods to WGS-based comparisons.



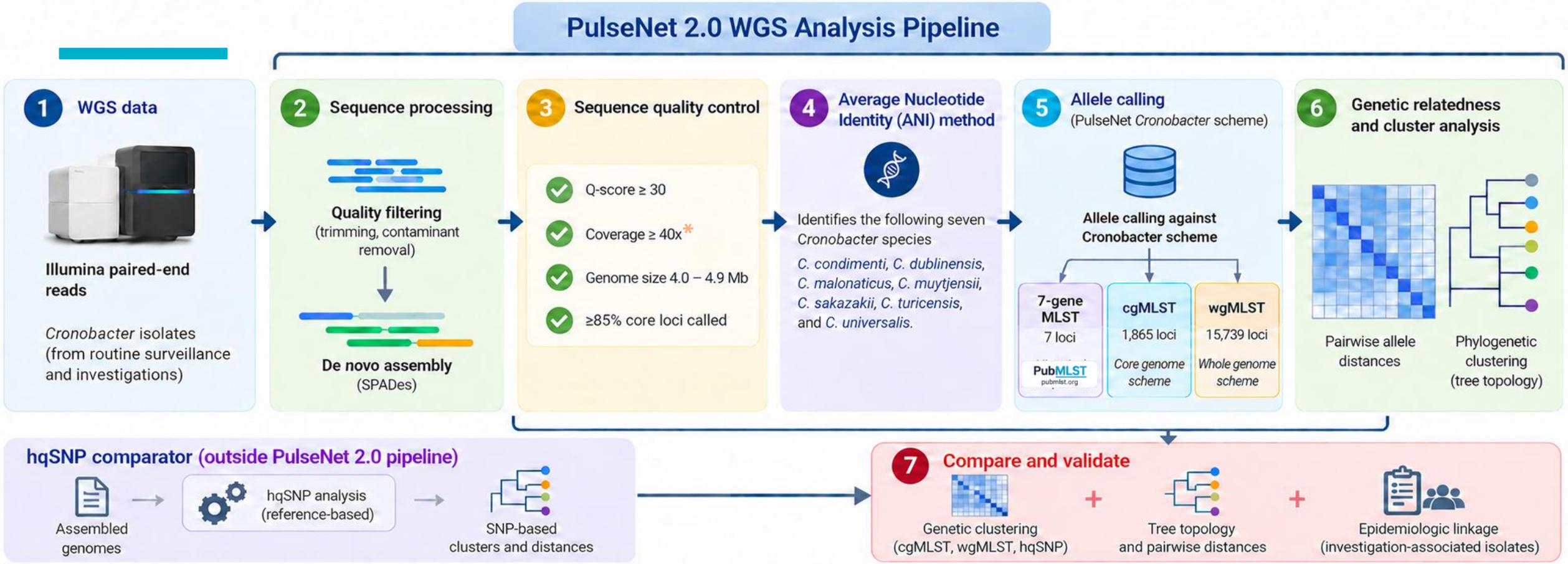
PulseNet — CDC's national laboratory network for foodborne disease surveillance — uses standardized molecular subtyping to detect and investigate clusters

Need: Validate standardized WGS approaches for routine PulseNet *Cronobacter* surveillance and outbreak detection

Study Question: Can core genome multilocus sequence typing (cgMLST) and whole genome multilocus sequence typing (wgMLST) cluster epidemiologically related isolates comparably to high-quality single nucleotide polymorphism (hqSNP) analysis?

Source: Joseph et al., Microbiology Spectrum, 2026

Whole genome sequencing analysis



***Why this matters**

Low coverage can reduce the number of confidently called loci. The study therefore used ≥40× coverage to optimize cgMLST allele calling and kept lower-coverage sequences out of cluster analysis.

Source: Joseph et al., Microbiology Spectrum, 2026

The three WGS analysis schemes answer different questions

1 7-gene MLST

Which established sequence type and clonal complex does this isolate resemble?

2 cgMLST

How many differences are present across 1,865 shared core loci?

3 wgMLST

Do accessory-genome loci add resolution when closely related isolates need to be separated?

Comparator hqSNP compares high-quality single-nucleotide differences against a closed reference genome.

The validation question was practical: can a standardized allele-based workflow preserve the epidemiologic signal that gold standard hqSNP analysis reveals?

Source: Joseph et al., Microbiology Spectrum, 2026

The validation dataset spans routine surveillance use cases

The study combined broad Cronobacter diversity with a focused epidemiologic test set.

All-isolate validation in PulseNet 2.0

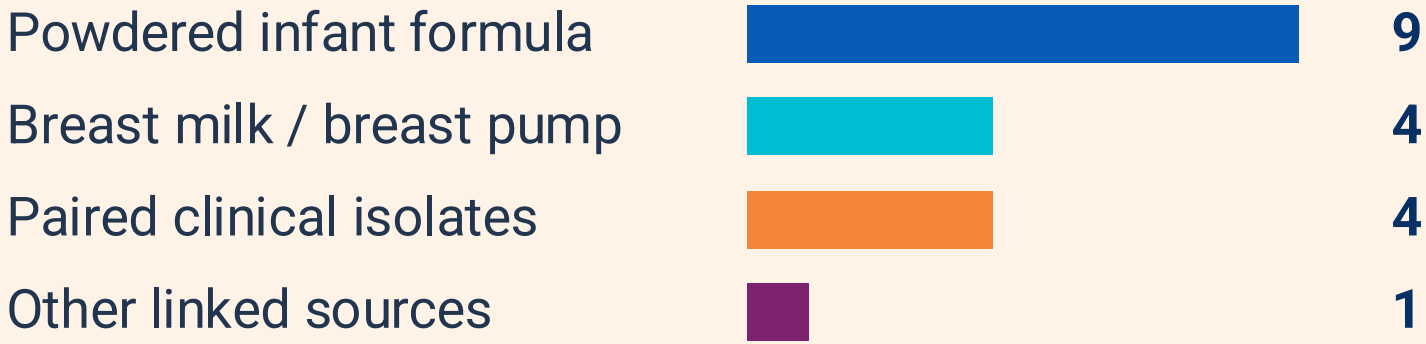
743 Sequences

- Clinical, food, and environmental sources
- Six *Cronobacter* species:
  - 639 *C. sakazakii*
  - 43 *C. turicensis*
  - 33 *C. dubliensis*
  - 19 *C. malonaticus*
  - 7 *C. muytjensii*
  - 2. *C. universalis*

Epidemiologic validation

73 *C. sakazakii* isolates from 18 investigations

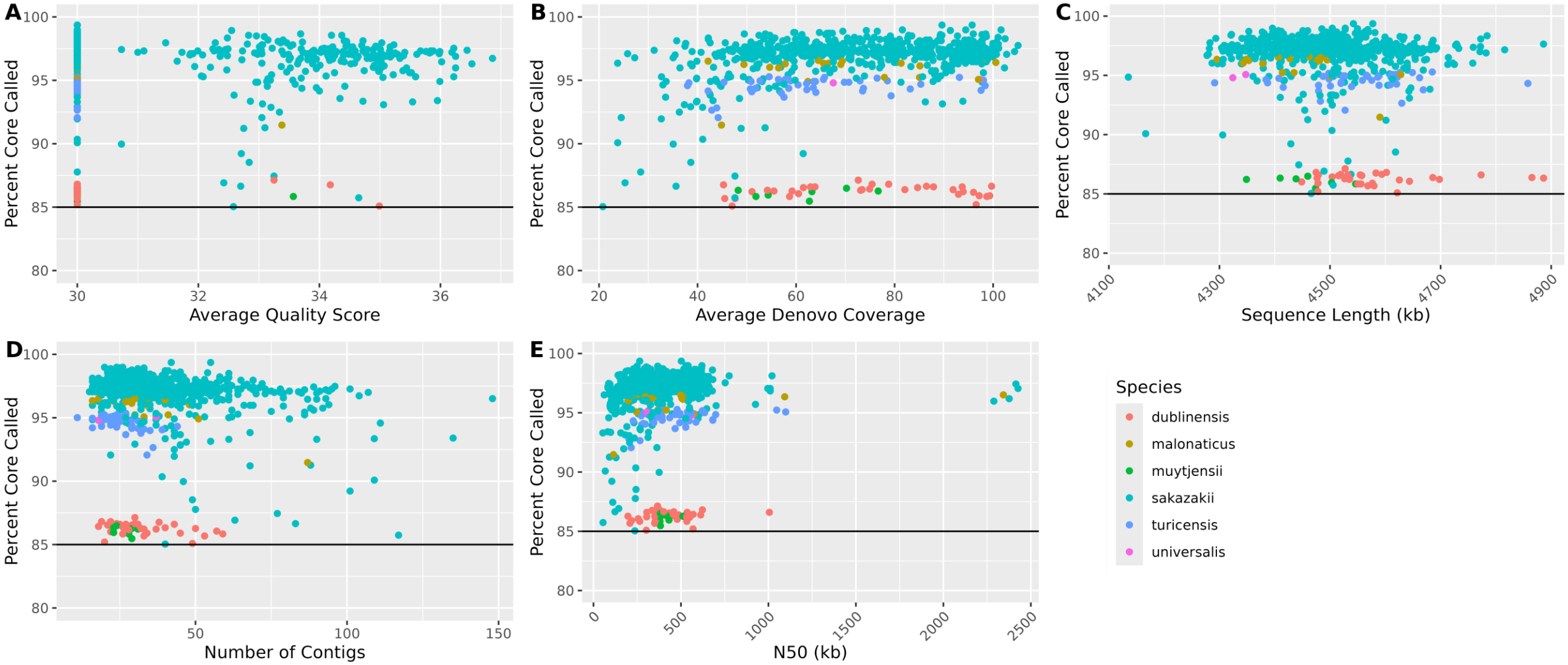
Sources represented in the investigation set



Source: Joseph et al., Microbiology Spectrum, 2026

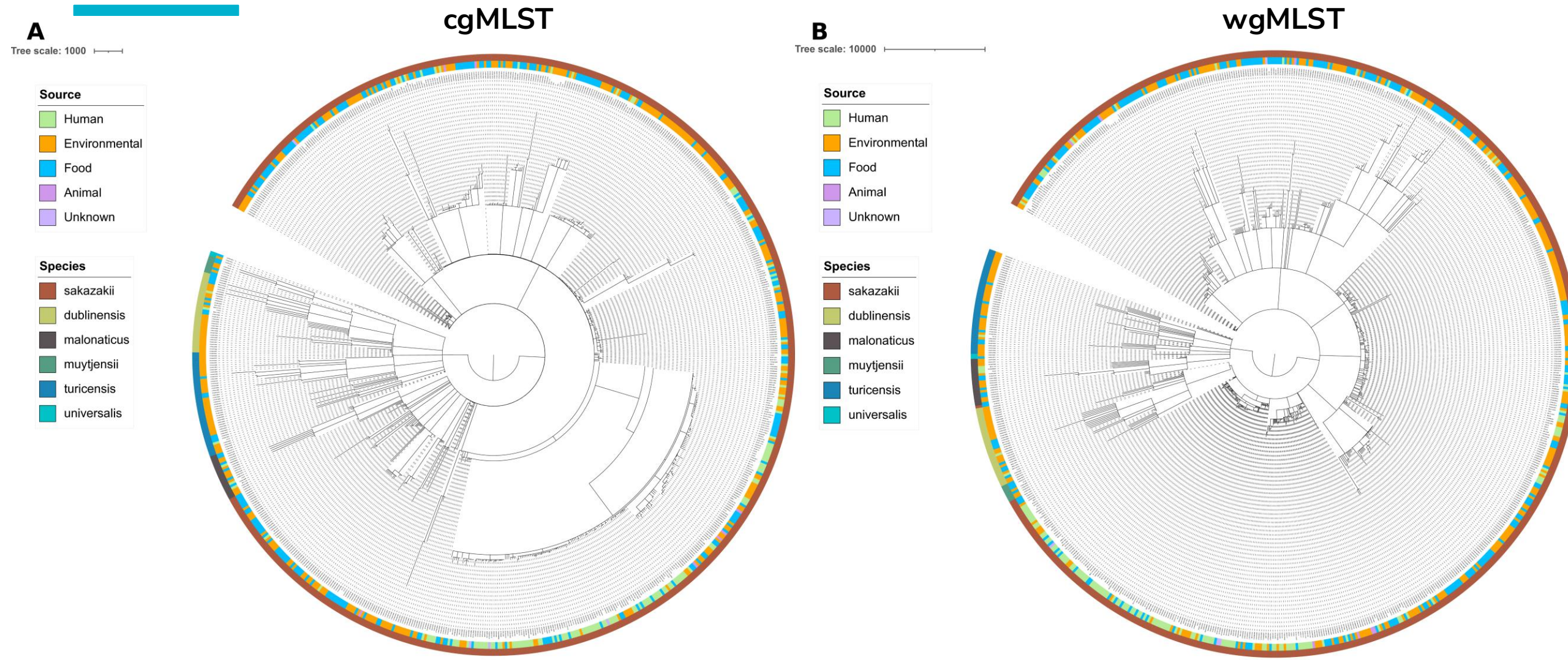
Quality control determines when allele calls are trustworthy

Routine surveillance depends on consistent sequence quality before clustering begins.



Source: Joseph et al., Microbiology Spectrum, 2026

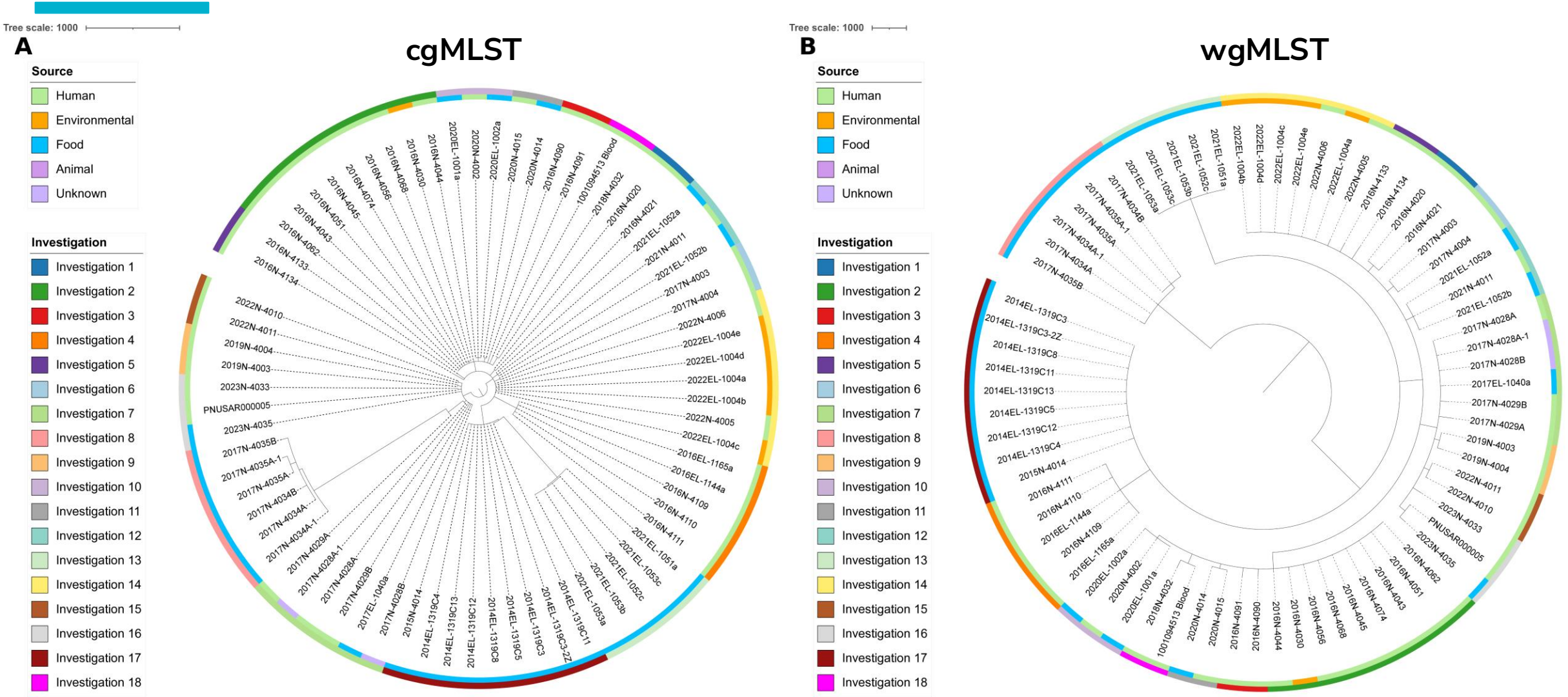
Allele-based typing separated *Cronobacter* species



Source: Joseph et al., Microbiology Spectrum, 2026

Allele-based typing separated related from unrelated isolates

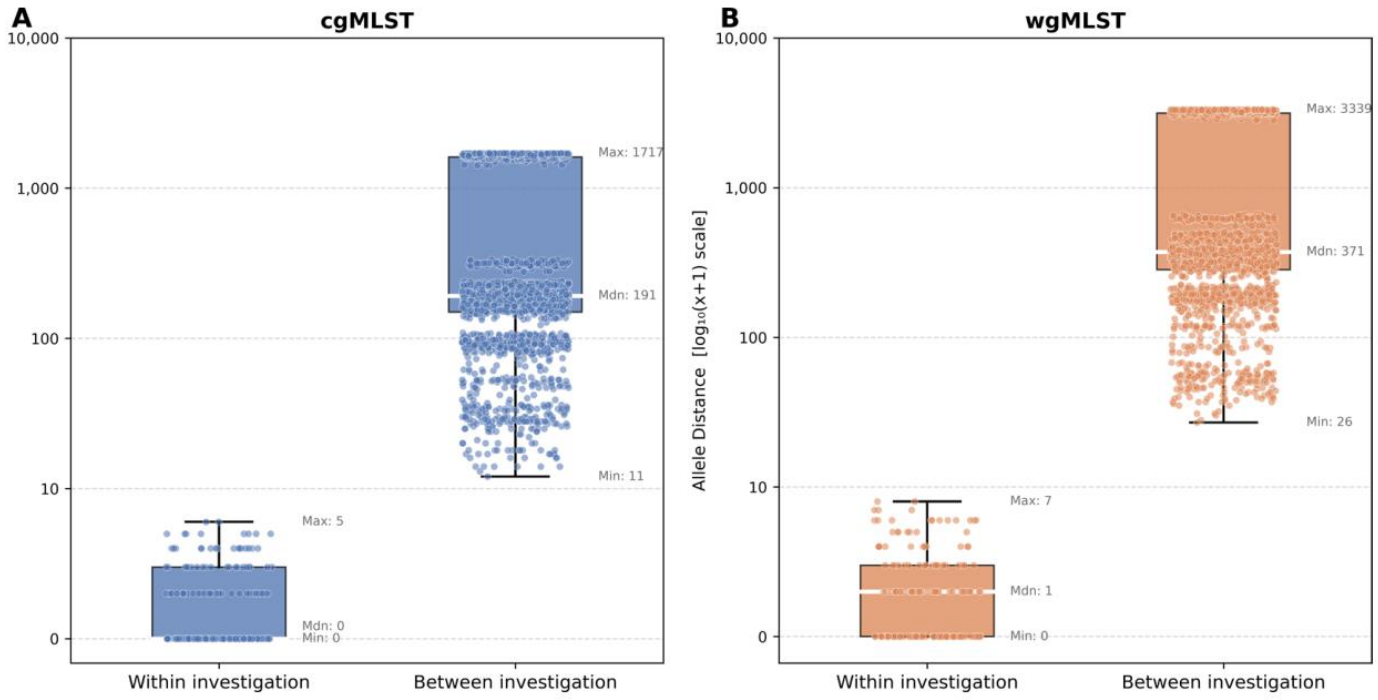
Within-investigation distances were far smaller than between-investigation distances.



Source: Joseph et al., Microbiology Spectrum, 2026

Allele-based typing separated related from unrelated isolates

Within-investigation distances were far smaller than between-investigation distances.



Summary

All 18 investigations clustered concordantly with epidemiologic information.
Within investigations: ≤8 hqSNPs, ≤5 cgMLST alleles, and ≤7 wgMLST alleles.
Between-investigation distances were much larger, with statistically significant separation ($P < 0.001$).

Practical interpretation: Isolates within an investigation and within 10 alleles are considered closely related in this data set.

Source: Joseph et al., Microbiology Spectrum, 2026

cgMLST and wgMLST were comparable with hqSNP

Investigations	No. Isolates	Sequence Type	Range of pairwise hqSNP and allele differences			Investigation Source
			hqSNP	cgMLST	wgMLST	
Investigation 1	2	ST4	0	0	0	Powdered Infant Formula
Investigation 2	9	ST4	0-7	0-5	0-6	Breast Milk
Investigation 3	2	ST4	1	1	1	Paired clinical isolates
Investigation 4	5	ST4	0-8	0-3	1-7	Breast Milk
Investigation 5	2	ST4	1	1	1	Powdered Infant Formula
Investigation 6	2	ST4	0	0	0	Paired clinical isolates
Investigation 7	7	ST4	0-1	0	0-1	Powdered Infant Formula
Investigation 8	6	ST1	0	0	0	Powdered Infant Formula
Investigation 9	2	ST4	0	0	1	Paired clinical isolates
Investigation 10	3	ST4	0-5	0-2	1-7	Powdered Infant Formula
Investigation 11	2	ST4	1	0	0	Powdered Infant Formula
Investigation 12	3	ST4	0	0	0	Powdered Infant Formula
Investigation 13	5	ST801	0-7	0-4	0-4	Powdered Infant Formula
Investigation 14	7	ST4	0-1	0-1	0-2	Breast Pump
Investigation 15	2	ST802	2	0	2	Paired clinical isolates
Investigation 16	3	ST4	0-4	0-3	0-5	Powdered Infant Formula
Investigation 17	9	ST4	0-3	0-2	0-2	Powdered Infant Formula
Investigation 18	2	ST4	1	1	1	Paired clinical isolates

Bottom line: cgMLST supports standardized routine surveillance; wgMLST and hqSNP add resolution for outbreak investigations and phylogenetic analysis.

Source: Joseph et al., Microbiology Spectrum, 2026

The 10-allele rule is a practical screen, not a stand-alone conclusion

Genetic proximity narrows the field; epidemiology determines whether the field makes sense.

1

Detect a close group

Use cgMLST to flag isolates within a practical allele threshold for review.

2

Check the context

Compare specimen source, timing, patient links, and other epidemiologic evidence.

3

Escalate resolution

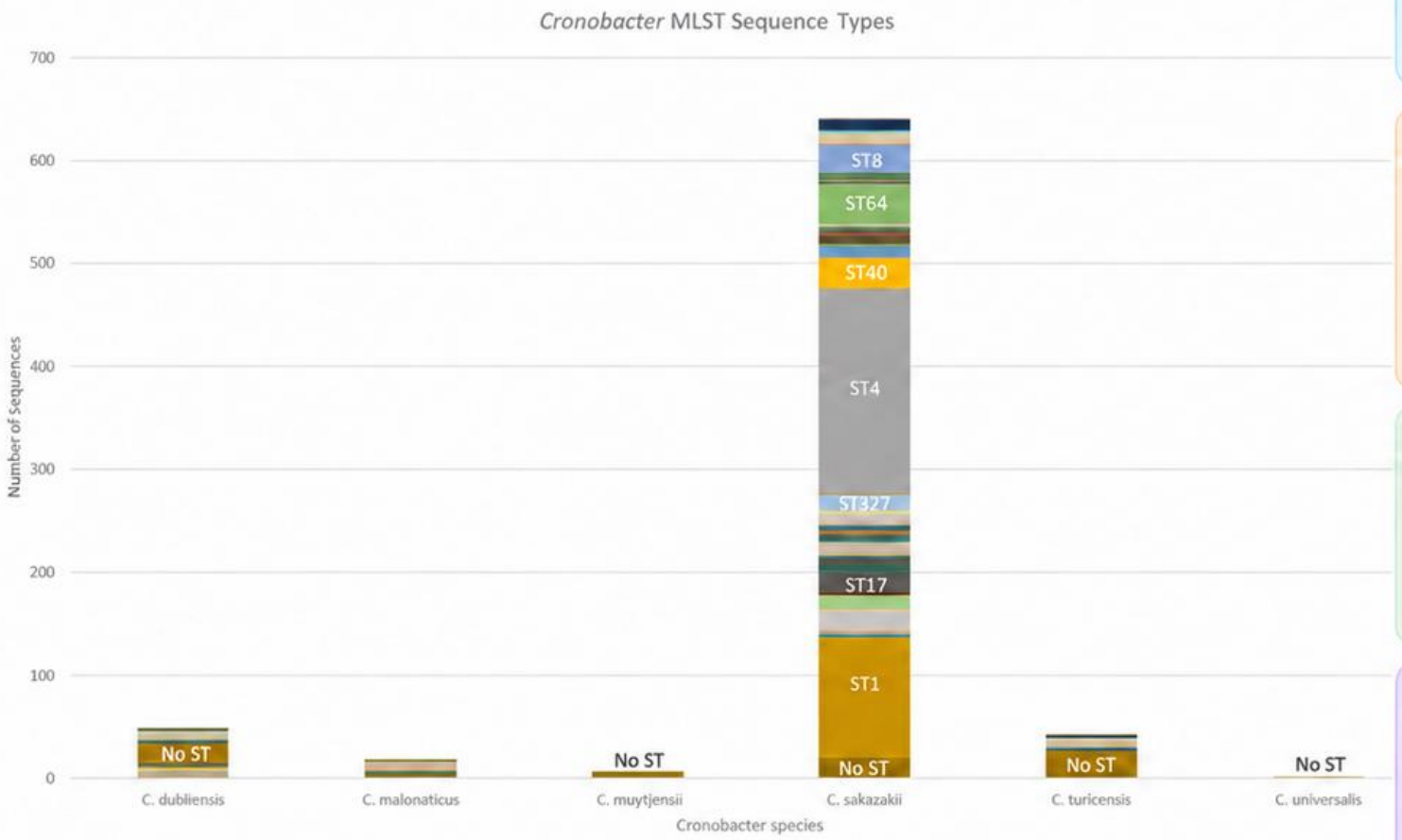
Use wgMLST or hqSNP when the investigation needs more discrimination or a phylogenetic view.

The paper notes that the exact cgMLST relatedness threshold for *C. sakazakii* remains to be established.

Source: Joseph et al., Microbiology Spectrum, 2026

Sequence typing adds historical context and diversity tracking

Classic MLST outputs remain useful alongside cgMLST and wgMLST.



Each color represents a distinct sequence type; selected STs are labeled.

Key contributions of 7-gene MLST

1 — *C. sakazakii* showed 66 different sequence types across 619 isolates.

- ST4 was the most common ST.
- ST4 represented 16 of 18 investigations.

2 — Sequence types varied by species.

- *C. sakazakii*: 66 STs / 619 isolates (majority ST4)
- *C. dublinensis*: 9 STs / 15 isolates
- *C. malonaticus*: 10 STs / 15 isolates
- *C. muytjensii*: 1 ST / 1 isolate
- *C. turicensis*: 12 STs / 16 isolates
- *C. universalis*: 0 STs / 0 isolates

3 — Clinically relevant sequence types.

- *C. sakazakii* ST4 has been frequently associated with invasive neonatal disease and has global distribution.
- Other STs (e.g., ST1, ST8, ST17, ST40, ST64, ST327) have also been reported in clinical isolates.

4 — MLST enables comparison to historical sequence type and clonal complex nomenclature.

- Sequence types were assigned from WGS data and linked to PubMLST (pubmlst.org).
- Provides a shared nomenclature for comparison with previous studies and global surveillance data.

Source: Joseph et al., Microbiology Spectrum, 2026

A practical PulseNet decision path emerges from the validation

The methods can be layered into routine surveillance and investigation response.

1

Sequence and QC

Automatically apply PulseNet quality thresholds before clustering.

2

Routine screen

Use cgMLST for standardized cross-laboratory cluster detection.

3

Investigate

Bring in epidemiology and source information to interpret a close group.

4

Resolve

Use wgMLST and hqSNP for higher-resolution or phylogenetic questions.

Result: a common routine screen with a clear path to deeper analysis when public health action requires it.

Source: Joseph et al., Microbiology Spectrum, 2026

Take-home messages

- **WGS-based cgMLST is ready for standardized routine PulseNet surveillance of *Cronobacter*, while wgMLST and hqSNP provide additional discriminatory power for investigations.**
- **7-gene MLST provides sequence type context to link current isolates with historical nomenclature and clinical relevance.**

Why it matters: These findings provide a foundation for standardized surveillance that can support earlier public-health action to protect vulnerable infants.

Thank you

Contact: Lavin A. Joseph | LJoseph1@cdc.gov

The findings and conclusions are those of the authors and do not necessarily represent the official position of CDC.

Acknowledgements

- Krittika Krishnan
- Cynney Walters
- Monica S. Im
- Dieter Ilse Marc De Coninck
- Sung B. Im
- Grant M. Williams
- Lee S. Katz
- Kelley B. Hise
- Heather A. Carleton
- Christine C. Lee

Publication QR Code