

Troubleshooting

December 26, 2026, Niamh Lacy-Roberts and Research Assistant at DTU, nlac@food.dtu.dk,
Natasia Rebekka Thornval, PhD student at DTU, nareth@food.dtu.dk

Outline

This session consists of the following elements:

1. Discussion of common challenges encountered during the analysis, including ambiguities and tool limitations.
2. Strategies for dealing with conflicting molecular and epidemiological data.

A reality check – perfect outbreaks are rare

- Real datasets are messy
- Different tools answer different questions
- Some uncertainty is normal — and expected
- Genomics supports decision-making, it does not deliver absolute truth
- SNP-based analysis is not static and hard to compare between laboratories
  - cgMLST has nomenclature that is standardized and is easier to compare between laboratories

Common challenge 1: borderline clusters

What it might look like:
- Isolates sit just above or below a SNP / allele threshold
- “Almost in” vs “almost out” of the outbreak

Why it happens:
- Arbitrary or strict thresholds
- Natural diversity within circulating strains
- Sampling gaps

Key takeaway
- Thresholds are guidelines, not rules

Common challenge 2: unstable trees

What it might look like:

Tree topology changes after:

- Different reference genome
- Slight filtering changes
- Subsetting samples

Why it happens

- Limited variation
- Reference bias
- Low-quality or uneven data

Key takeaway

- Trees are models, not facts

Common challenge 3: tool disagreement

What it might look like:

- cgMLST suggests clustering
- SNP analysis suggests separation (or vice versa)
- Different ST but close clustering

Why it happens

- Different underlying assumptions
- Different resolution
- Different handling of missing data

Key takeaway

- Disagreement ≠ error
- Ask what question each tool is answering

Importance of data quality

Small issues → big interpretive effects

- Low or uneven depth
- Assemblies vs reads
- Mixed or contaminated samples
- Missing regions in reference (low coverage)

- Do QC first!
- If results are confusing, check QC again before interpretation

When genomics and epidemiology disagree

This can be common in outbreaks

Examples:

- Genetically close, epidemiologically unrelated
- Epidemiologically linked, genetically distant
- This does not mean one is “wrong”
- Last session will include examples

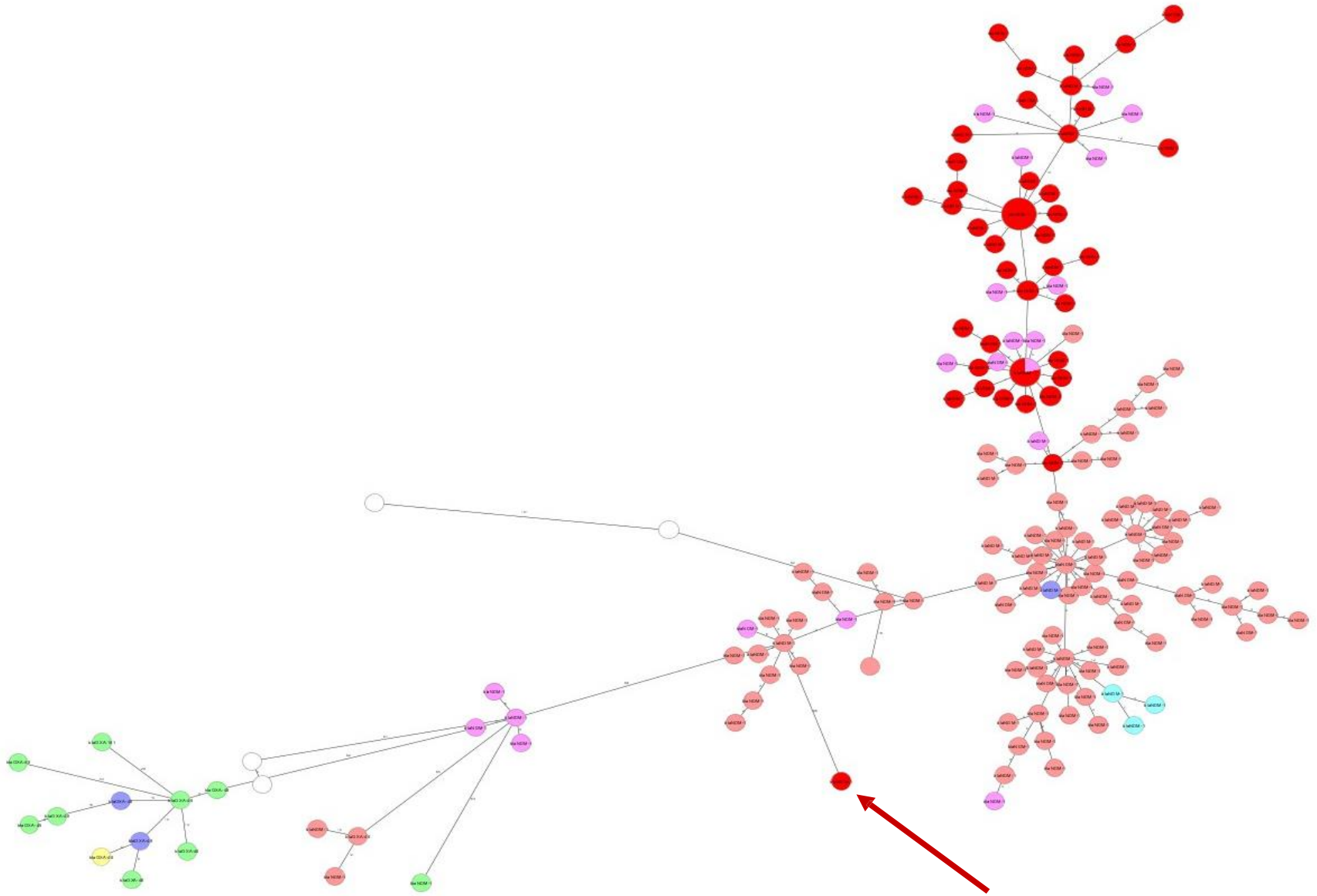
When results conflict or are vague

Ask these questions:

- Are the assumptions correct? (timeframe, reference, filtering)
- Is there unsampled diversity?
- Could this reflect background circulation?
- Are we forcing a binary “in/out” decision?

- Disagreement is part of outbreak analysis
- Tools inform, they don’t decide
- Uncertainty should be stated, not hidden
- Interpretation improves with context
- Sometimes you have to keep digging :D

cgMLST tree – hospital outbreak



Adapted from Henrik Hasman, Statens Serum Institut, Denmark, slides from EURL multidisciplinary training workshop on simulated exercise, 2022.

What is going on with that isolate?

Targets of Distance Columns (CPO C. freundii ST18)

Right-click on the allele type columns to jump to the according contig position in the Sample

arget	Begin	End	GenBank gene	GenBank product	GenBank note	GenBank protein_id	200117_A19_...	AMA003417	AMA003565	CPO20190159	AMA00338
.322_RS02285	465,622	467,868		phosphoenolpyruvate--protein phosphotransferase PtsP	member of a ...	WP_003033984.1	? (failed)	1	1	1	1
.322_RS03425	716,674	721,020		autotransporter domain-containing protein	Derived by a...	WP_071684359.1	? (failed)	1	1	1	? (not found
.322_RS03765	802,371	803,735		PTS sugar transporter subunit IIC	Derived by a...	WP_054528657.1	? (failed)	1	? (not found)	1	1
.322_RS04195	912,678	914,693		tRNA(Met) cytidine acetyltransferase TmcA	cetylates the...	WP_054528641.1	? (failed)	1	1	1	1
.322_RS05765	1,252,228	1,254,714		fimbrial assembly protein	Derived by a...	WP_054528576.1	? (failed)	1	1	1	1
.322_RS06635	1,433,991	1,435,370		cobyrinic acid a,c-diamide synthase	Derived by a...	WP_044701540.1	? (failed)	1	1	1	1
.322_RS06975	1,490,987	1,491,643		DNA-binding response regulator	Derived by a...	WP_003030486.1	? (failed)	1	1	1	1
.322_RS09275	1,966,895	1,967,473		TetR family transcriptional regulator	Derived by a...	WP_046670695.1	? (failed)	1	1	1	1
.322_RS09820	2,083,766	2,084,500		DNA-binding response regulator	Derived by a...	WP_003836390.1	? (failed)	1	1	1	1
.322_RS11920	2,514,178	2,514,801		DSBA oxidoreductase	Derived by a...	WP_003035975.1	? (failed)	1	1	1	1
.322_RS12760	2,702,258	2,703,679		2-oxoglutarate/malate translocator	Derived by a...	WP_003837022.1	? (failed)	1	1	1	1
.322_RS13805	2,920,181	2,921,314		LPS O-antigen length regulator	Derived by a...	WP_054528176.1	? (failed)	1	1	1	1
.322_RS14935	3,176,896	3,177,909		4-hydroxy-2-oxovalerate aldolase	Derived by a...	WP_003021379.1	? (failed)	1	1	1	1
.322_RS15475	3,301,194	3,301,901		flagellar basal body L-ring protein	Derived by a...	WP_042270212.1	? (failed)	1	1	1	1
.322_RS15765	3,357,582	3,359,564		type IV secretion protein Rhs	Derived by a...	WP_072143931.1	? (failed)	1	1	1	1
.322_RS19670	4,225,537	4,226,988		potassium transporter	Derived by a...	WP_003017848.1	? (failed)	1	1	1	1
.322_RS21515	4,624,296	4,625,513		MFS transporter	Derived by a...	WP_054528867.1	? (failed)	1	1	1	1
.322_RS07405	1,575,702	1,576,793		enterohemolysin	Derived by a...	WP_054528497.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS07605	1,603,957	1,604,262		hypothetical protein	Derived by a...	WP_057101149.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS08700	1,830,199	1,830,762		hypothetical protein	Derived by a...	WP_003843940.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS17560	3,773,522	3,773,764		transcriptional regulator	Qin prophag...	WP_003839576.1	? (not found)	1	? (not found)	? (not found)	1
.322_RS22180	4,766,146	4,767,330		elongation factor Tu	Derived by a...	WP_003031109.1	? (not found)	? (not found)	? (not found)	? (not found)	? (not found
.322_RS06380	1,382,478	1,383,593		amino acid oxidase	Derived by a...	WP_054528547.1	? (not found)	1	? (not found)	? (not found)	? (not found
.322_RS17175	3,686,035	3,686,850		AraC family transcriptional regulator	Derived by a...	WP_054528023.1	? (not found)	1	1	1	1
.322_RS17930	3,844,410	3,846,275		DNA mismatch repair protein MutL	Derived by a...	WP_054527983.1	? (not found)	1	1	1	1
.322_RS20890	4,482,716	4,483,960		O-antigen polymerase	Derived by a...	WP_046671022.1	? (not found)	1	1	? (not found)	? (not found
.322_RS22035	4,737,170	4,739,713		nitrite reductase large subunit	Derived by a...	WP_003023592.1	? (not found)	1	1	1	1
.322_RS23080	211,671	211,868		hypothetical protein						1	1
.322_RS23130	545,890	546,069		hypothetical protein						? (not found)	1
.322_RS02845	588,933	589,343		formate hydrogenlyase maturation protein Hyc						1	2
.322_RS03030	623,789	624,124		L-valine transporter subunit YgaH					(not found)	? (not found)	1
.322_RS03050	627,414	628,478		proline/betaine ABC transporter permease ProW	Derived by a...	WP_003846040.1	1	1	? (not found)	? (not found)	1
.322_RS03070	633,115	633,525	nrdI	ribonucleotide reductase assembly protein NrdI	in Salmonella...	WP_003037273.1	1	1	? (not found)	? (not found)	1
.322_RS03235	667,222	668,508		capsular polysaccharide biosynthesis protein	Derived by a...	WP_003839728.1	1	1	1	1	1

Hypermutator!!

Adapted from Henrik Hasman, Statens Serum Institut, Denmark, slides from EURL multidisciplinary training workshop on simulated exercise, 2022.

Do not edit
How to change the design



Did you encounter a result you weren’t sure how to interpret during the exercise?

ⓘ The Slido app must be installed on every computer you’re presenting from

slido



Thank you for listening! ☺

Acknowledgements

The creation of this training material was commissioned by ECDC with the direct involvement of Niamh Lacy-Roberts, Research Assistant at DTU, nlac@food.dtu.dk