Guided discussion using interactive polls

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

Outline

This session consists of the following elements:

1. Recapping the homework exercise
2. Taxonomic analysis and typing
3. Creation of epidemiological curves and hypothesis generation
4. Phylogenetic tree construction and interpretation
5. Identification of outbreak clusters
6. Correlation of epidemiological data with phylogenetic trees

Let’s recap the homework exercise

Simulated Exercise scenario

- In January 2026, public health authorities in **several regions** reported **an increase in suspected human brucellosis cases.**
- Patients presented with prolonged fever, fatigue, night sweats, and joint pain.
- Cases were identified across multiple regions, and early **investigations did not reveal a shared workplace, household, or occupational exposure**.
- During the same period, veterinary services reported **several livestock miscarriage events in rural areas**.
- These events raised concerns about a **possible *Brucella*-associated problem in livestock**, although the species involved was initially unknown.
- The temporal proximity of these events to the rise in human cases added complexity to the investigation.

Simulated Exercise scenario continued

- As part of routine surveillance and outbreak response activities, a **collection of bacterial isolates from human, animal, food, and environmental sources was sequenced**.
- The isolates are suspected to belong to the **genus *Brucella*, but species-level identification has not yet been confirmed**.
- During case interviews**, several patients reported travel during December**, including visits to **Christmas markets in a rural region**.
- **Not all patients reported travel themselves**; however**, some reported consuming food items brought back by relatives or friends** who had travelled during the holiday period.
- For a subset of cases, exposure information was incomplete or uncertain.

Data for the SimEx

Available isolates:

- 30 *Brucella* spp. genomes
- **Data format:**
- Illumina sequences
- Provided as FASTA files and raw FASTQ files
- Metadata and quality control of isolates in excel file

Access:

- All files are available on ScienceData via this link:
- https://sciencedata.dk/shared/e9cba091ad1d946970bfdd67e8491b2a

Metadata

Strain_ID	Host_type	Sample_sour	Sample_mat	Country	Region	Symptom_on	Collection_d	Reported_ex	Travel_histor	Notes
Animal_143	Animal	Farm	Goat milk	Country A	South		5/10/25	Routine surve	Unknown	Small farm
Animal_217	Animal	Farm	Sheep tissue	Country B	West		6/18/25	Routine surve	Unknown	No human cases
Animal_770	Animal	Wildlife	Wild boar tiss	Country A	Forest		7/15/25	Wildlife surveil	Unknown	Wildlife
Animal_920	Animal	Farm	Goat tissue	Country C	Central		9/2/25	Routine surve	Unknown	Different genotype
Animal_965	Animal	Slaughterhou	Pig tissue	Country B	South		6/30/25	Routine surve	Unknown	Domestic pig
Environment	Environment	Farm	Soil	Country B	Rural		7/1/25	Environmenta	Unknown	No link to cases
Environment	Environment	Forest	Soil	Country A	Forest		7/20/25	Wildlife surveil	Unknown	No human cases
Environment	Environment	Farm	Soil	Country A	Rural		11/15/25	Post-miscarria	Unknown	Cattle farm
Environment	Environment	Dairy	Surface swab	Country A	Rural		9/12/25	Environmenta	Unknown	Older isolate
Environment	Environment	Farm	Surface swab	Country B	West		11/18/25	Post-miscarria	Unknown	Abortions reported
Environment	Environment	Farm	Water	Country C	Central		11/22/25	Post-miscarria	Unknown	Abortions reported
Environment	Environment	Environment	Water	Country C	North		10/5/25	Environmenta	Unknown	Unrelated
Food_310	Food	Dairy	Raw milk	Country A	Rural		11/20/25	Routine samp	Unknown	Cattle herd
Food_585	Food	Retail	Game meat	Country B	East		8/2/25	Routine samp	Unknown	Wild source
Food_627	Food	Retail	Soft cheese (r	Country A	Rural		1/15/26	Market sampli	Unknown	Unpasteurized
Food_659	Food	Dairy	Raw milk	Country B	West		11/25/25	Routine samp	Unknown	Cattle herd
Food_668	Food	Dairy	Raw milk	Country C	Central		12/3/25	Routine samp	Unknown	Cattle herd
Food_978	Food	Retail	Cheese	Country C	South		8/15/25	Routine samp	Unknown	Pasteurized
Human_182	Human	Clinical	Blood	Country A	North	1/12/26	1/22/26	Dairy consum	No	Consumed food from relative
Human_264	Human	Clinical	Blood	Country C	Central	1/20/26	1/30/26	Dairy consum	Yes	Market visit reported
Human_359	Human	Clinical	Blood	Country A	Capital	1/8/26	1/18/26	Dairy consum	Yes	Visited Christmas market
Human_412	Human	Clinical	Blood	Country C	North	1/27/26	1/6/26	Dairy consum	Yes	Holiday travel
Human_430	Human	Clinical	Blood	Country A	West	1/18/26	1/28/26	Dairy consum	Unknown	Incomplete interview
Human_454	Human	Clinical	Blood	Country A	Capital	1/25/26	1/4/26	Dairy consum	Unknown	No occupational exposure
Human_488	Human	Clinical	Blood	Country B	South	1/23/26	1/2/26	Dairy consum	No	Cheese shared at home
Human_563	Human	Clinical	Blood	Country A	Rural	11/10/25	11/20/25	Animal contac	No	Veterinarian
Human_671	Human	Clinical	Blood	Country B	East	1/29/26	1/8/26	Dairy consum	No	Gifted food item
Human_694	Human	Clinical	Blood	Country A	Rural	11/20/25	12/1/25	Farm exposur	No	Cattle contact
Human_793	Human	Clinical	Blood	Country B	East	1/15/26	1/25/26	Dairy consum	Yes	Holiday travel
Human_829	Human	Clinical	Blood	Country A	Urban	12/5/25	12/15/25	Unknown	No	Sporadic human case; no known link to outbreak

Quality of the data

Sample	Bases (MB)	Qual Bases (	Qual bases %	Reads	Qual reads	Qual reads %	Most commo	2. Most com	Other adapte	insert size	N50	No ctgs	longest	total bps	grad
Animal_143	197	195	0.9882	1317510	1317364	0.9999	9	6	19		194857	33	420193	3273752	A
Animal_217	197	195	0.9882	1317454	1317272	0.9999	8	6	26		221848	33	419979	3275140	A
Animal_770	199	197	0.9881	1331498	1331326	0.9999	10	7	27		163604	41	371032	3302424	A
Animal_920	197	194	0.9883	1314480	1314340	0.9999	9	6	17		201422	28	393571	3287685	A
Animal_965	197	195	0.9882	1318680	1318490	0.9999	10	8	23		170357	36	519425	3279458	A
Environment_320	190	188	0.9882	1269824	1269714	0.9999	8	6	20		11712	429	48838	3248572	C
Environment_735	199	196	0.9882	1326974	1326814	0.9999	12	9	31		140595	37	519265	3299603	A
Environment_757	197	194	0.9883	1314420	1314272	0.9999	13	8	25		294096	32	518913	3264718	A
Environment_759	198	196	0.9882	1324892	1324730	0.9999	11	10	23		249753	30	454373	3289334	A
Environment_848	196	194	0.9882	1311802	1311664	0.9999	7	6	12		190237	34	382195	3264395	A
Environment_902	199	197	0.9883	1331460	1331294	0.9999	8	5	19		228720	33	364405	3262852	A
Environment_907	197	195	0.9882	1319084	1318912	0.9999	9	7	27		278268	24	609266	3287393	A
Food_310	197	194	0.9883	1314408	1314276	0.9999	8	7	16		251704	34	401035	3264092	A
Food_585	199	197	0.9882	1329878	1329714	0.9999	9	8	19		166918	39	320046	3301878	A
Food_627	198	196	0.9882	1324094	1323924	0.9999	12	5	18		242288	28	513252	3288924	A
Food_659	202	199	0.9882	1347408	1347232	0.9999	9	5	20		135549	48	299117	3257640	A
Food_668	197	194	0.9882	1314396	1314236	0.9999	7	6	17		294089	30	518914	3263903	A
Food_978	198	196	0.9883	1324380	1324240	0.9999	7	7	27		249788	29	511997	3287448	A
Human_182	198	196	0.9882	1324110	1323942	0.9999	10	7	39		242288	27	518613	3289389	A
Human_264	198	196	0.9882	1324084	1323940	0.9999	7	7	26		242288	27	518613	3289722	A
Human_359	198	196	0.9882	1324060	1323932	0.9999	9	6	31		242288	27	518613	3289392	A
Human_412	198	196	0.9881	1324126	1323938	0.9999	6	6	17		242288	27	518613	3289420	A
Human_430	198	196	0.9883	1324096	1323934	0.9999	7	7	20		242288	27	518613	3289419	A
Human_454	198	196	0.9881	1324124	1323944	0.9999	7	6	18		242288	27	518613	3289486	A
Human_488	198	196	0.9881	1324110	1323932	0.9999	7	6	24		242288	27	518613	3289412	A
Human_563	197	195	0.9882	1317344	1317160	0.9999	12	6	24		198280	31	353716	3274832	A
Human_671	198	196	0.9882	1324106	1323918	0.9999	13	10	25		242288	27	518613	3289403	A
Human_694	197	194	0.9882	1314360	1314154	0.9998	7	6	25		294081	31	518914	3263192	A
Human_793	198	196	0.9882	1324104	1323906	0.9999	12	5	25		242288	27	518613	3289762	A
Human_829	196	193	0.9882	1307822	1307652	0.9999	9	6	30		260526	25	584564	3254971	A

Data for the SimEx

https://sciencedata.dk/shared/e9cba091ad1d946970bfdd67e8491b2a

Your homework for next week (Session 2)

- Using tools of your choice:

- Identify the bacterial species present in the dataset
- Identify SNP differences between isolates
- Construct and interpret a phylogenetic tree
- Assess whether a genomic outbreak cluster is present, and define the outbreak cluster if so
- Infer the most likely source and route(s) of transmission, and discuss alternative explanations where relevant
- Assess whether the reported livestock miscarriage events are epidemiologically linked to the human cases
- Create basic epidemiological curves from the metadata using Microsoft Excel

- Critical reflection: Notes on any limitations or uncertainties in your analysis.

Did you attempt the exercise?

Taxonomic analysis and Typing

Could you identify the *Brucella* spp?

- 30 *Brucella* spp. genomes

Which *Brucella* species did you identify?

Could you identify the *Brucella* spp?

- 30 *Brucella* spp. Genomes

- 18 *B. melitensis*
- 8 *B. abortus*
- 4 *B. suis*

Did you find the same number of species in this dataset?

Which tools did you use to identfy the species?

Did you perform any other analyses? E.g., MLST?

If you did run MLST, did the species have the same STs?

MLST

- Based on 7 ‘housekeeping’ genes
- Each gene variant is assigned a variant number (allele)
- The combination of allele numbers -> MLST number
- Just one nucleotide change in one gene -> new MLST number
- One allele variant can still be closely related

MLST-2.0 Server - Results

mlst Profile: *kpneumoniae*

Organism: *Klebsiella pneumoniae*

Sequence Type: *147* **Input Files:** *kp1_1.fa*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
gapA	100	100	450	450	0	gapA_3
infB	100	100	318	318	0	infB_4
mdh	100	100	477	477	0	mdh_6
pgi	100	100	432	432	0	pgi_1
phoE	100	100	420	420	0	phoE_7
rpoB	100	100	501	501	0	rpoB_4
tonB	100	100	414	414	0	tonB_38

extended output

Input Files: *kp1_1.fa*

Sequence Type: *4843*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
gapA	100	100	450	450	0	gapA_3
infB	100	100	318	318	0	infB_4
mdh	100	100	477	477	0	mdh_6
pgi	100	100	432	432	0	pgi_1
phoE	100	100	420	420	0	phoE_7
rpoB	100	100	501	501	0	rpoB_273
tonB	100	100	414	414	0	tonB_38

extended output

Input Files: *kp11_1.fa*

MLST – one allele difference

In this case, there is only one SNP between the rpoB alleles

One allele variant can still be closely related!

```
# Length: 501
# Identity:     500/501 (99.8%)
# Similarity:   500/501 (99.8%)
# Gaps:           0/501 ( 0.0%)
# Score: 2496.0
#
#
#=======================================
```

```
  1 ACCCACTACGGTCGCGTATGTCCGATCGAAACGCCTGAAGGTCCGAACAT     50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 ACCCACTACGGTCGCGTATGTCCGATCGAAACGCCTGAAGGTCCGAACAT     50

 51 CGGTCTGATTAACTCCCTGTCCGTGTACGCGCAGACCAACGAATATGGCT    100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 CGGTCTGATTAACTCCCTGTCCGTGTACGCGCAGACCAACGAATATGGCT    100

101 TCCTTGAGACGCCGTATCGTAAAGTGACCGACGGTGTGGTTACTGACGAA    150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 TCCTTGAGACGCCGTATCGTAAAGTGACCGACGGTGTGGTTACTGACGAA    150

151 ATTCACTACCTGTCTGCTATCGAAGAAGGCAACTACGTTATCGCTCAGGC    200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 ATTCACTACCTGTCTGCTATCGAAGAAGGCAACTACGTTATCGCTCAGGC    200

201 GAACTCCAACCTGGATGAAAACGGCCACTTCGTAGAAGATCTGGTTACCT    250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 GAACTCCAACCTGGATGAAAACGGCCACTTCGTAGAAGATCTGGTTACCT    250

251 GCCGTAGCAAAGGCGAATCCAGCTTGTTCAGCCGCGACCAGGTTGACTAC    300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 GCCGTAGCAAAGGCGAATCCAGCTTGTTCAGCCGCGACCAGGTTGACTAC    300

301 ATGGACGTATCCACCCAGCAGGTGGTATCCGTCGGTGCGTCCCTGATCCC    350
    ||||||||||||||||||||.||||||||||||||||||||||||||||
301 ATGGACGTATCCACCCAGCAGGCGGTATCCGTCGGTGCGTCCCTGATCCC    350

351 GTTCCTGGAACACGATGACGCCAACCGTGCATTGATGGGTGCGAACATGC    400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 GTTCCTGGAACACGATGACGCCAACCGTGCATTGATGGGTGCGAACATGC    400

401 AACGTCAGGCGGTTCCGACTCTGCGCGCTGATAAGCCGCTGGTTGGTACC    450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 AACGTCAGGCGGTTCCGACTCTGCGCGCTGATAAGCCGCTGGTTGGTACC    450

451 GGTATGGAACGTGCTGTTGCCGTTGACTCCGGTGTTACTGCCGTGGCTAA    500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 GGTATGGAACGTGCTGTTGCCGTTGACTCCGGTGTTACTGCCGTGGCTAA    500

501 A    501
    |
501 A    501
```

Identification and typing of *Brucella* spp. [5-7]

MLST for *Brucella* spp. → https://pubmlst.org/organisms/brucella-spp

NOTE

STs that seem close to each other (e.g. ST-6 and ST-7) are **not** more similar than STs that are far away (e.g. ST-1 and ST-28).

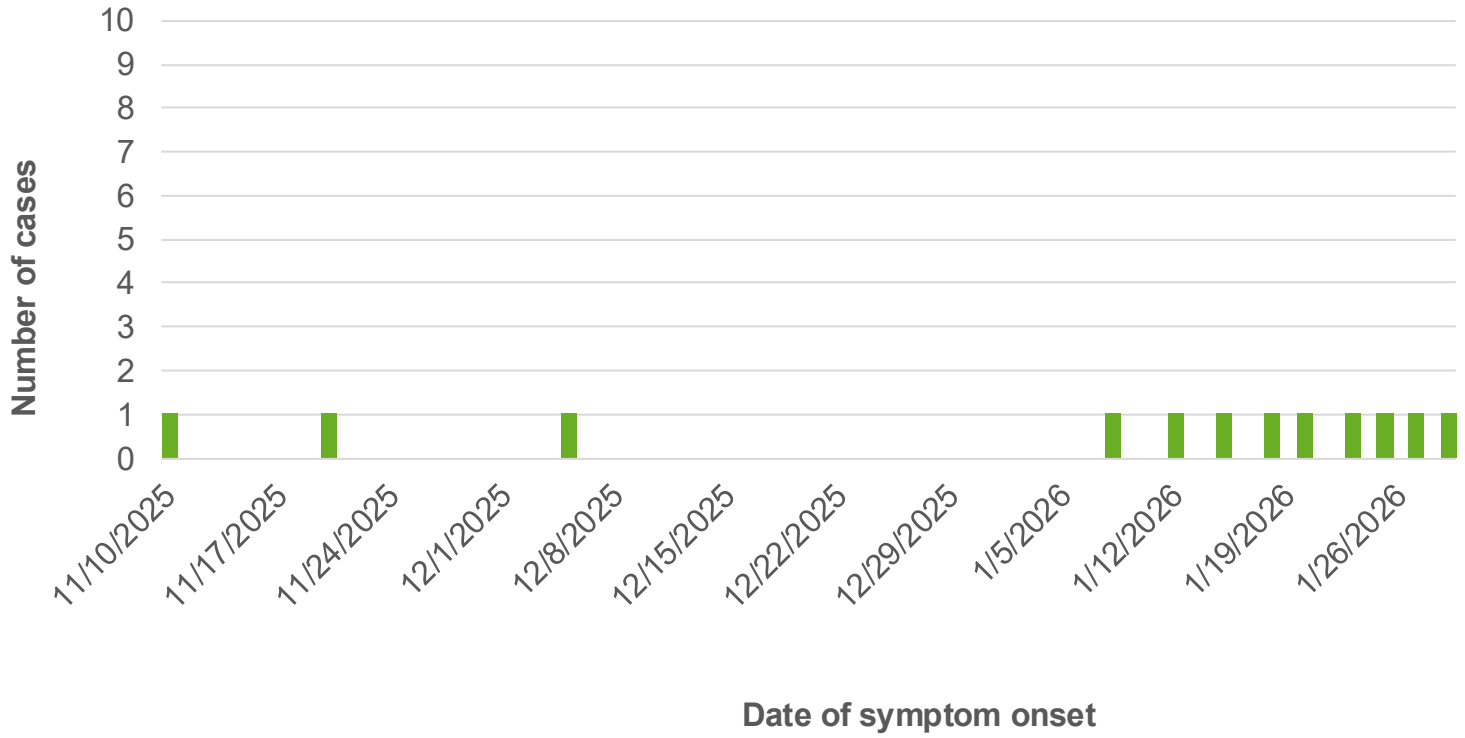
BruMLSA9

ST	gap	aroA	glk	dnaK	gyrB	trpE	cobQ	int_hyp	omp25
6	5	7	10	7	6	3	7	1	1
7	3	5	3	2	1	5	2	2	10
1	2	1	1	2	1	3	1	1	1
28	2	1	1	2	1	3	12	1	1

Epidemiological curves and hypothesis generation

Did you try making an epidemiological curve?

Epi curve


10
9
8
7
6
5
4
3
2
1
0
Number of cases
11/10/2025
11/17/2025
11/24/2025
12/1/2025
12/8/2025
12/15/2025
12/22/2025
12/29/2025
1/5/2026
1/12/2026
1/19/2026
1/26/2026
Date of symptom onset

What do we know from the scenario?

- Suspected human brucellosis cases
- Several livestock miscarriage events in rural areas
- Several patients reported travel during December, including visits to Christmas markets in a rural region
- Some patients reported consuming food items brought back by relatives or friends who had travelled during the holiday period.

What do we know from the metadata?

	A	B	C	D	E	F	G	H	I	J	K	L
1	**Strain_ID**	**Host_type**	**Species**	**Sample_source**	**Sample_material**	**Country**	**Region**	**Symptom_onset_date**	**Collection_date**	**Reported_exposure**	**Travel_history**	**Notes**
2	**Animal_143**	Animal	B. melitensis	Farm	Goat milk	Country A	South		2025-05-10	Routine surveillance	Unknown	Small farm
3	**Animal_217**	Animal	B. melitensis	Farm	Sheep tissue	Country B	West		2025-06-18	Routine surveillance	Unknown	No human cases
4	**Animal_770**	Animal	B. suis	Wildlife	Wild boar tissue	Country A	Forest		2025-07-15	Wildlife surveillance	Unknown	Wildlife
5	**Animal_920**	Animal	B. melitensis	Farm	Goat tissue	Country C	Central		2025-09-02	Routine surveillance	Unknown	Different genotype
6	**Animal_965**	Animal	B. suis	Slaughterhouse	Pig tissue	Country B	South		2025-06-30	Routine surveillance	Unknown	Domestic pig
7	**Environment**	Environment	B. melitensis	Farm	Soil	Country B	Rural		2025-07-01	Environmental surveillance	Unknown	No link to cases
8	**Environment**	Environment	B. suis	Forest	Soil	Country A	Forest		2025-07-20	Wildlife surveillance	Unknown	No human cases
9	**Environment**	Environment	B. abortus	Farm	Soil	Country A	Rural		2025-11-15	Post-miscarriage sampling	Unknown	Cattle farm
10	**Environment**	Environment	B. melitensis	Dairy	Surface swab	Country A	Rural		2025-09-12	Environmental surveillance	Unknown	Older isolate
11	**Environment**	Environment	B. abortus	Farm	Surface swab	Country B	West		2025-11-18	Post-miscarriage sampling	Unknown	Abortions reported
12	**Environment**	Environment	B. abortus	Farm	Water	Country C	Central		2025-11-22	Post-miscarriage sampling	Unknown	Abortions reported
13	**Environment**	Environment	B. melitensis	Environment	Water	Country C	North		2025-10-05	Environmental surveillance	Unknown	Unrelated
14	**Food_310**	Food	B. abortus	Dairy	Raw milk	Country A	Rural		2025-11-20	Routine sampling	Unknown	Cattle herd
15	**Food_585**	Food	B. suis	Retail	Game meat	Country B	East		2025-08-02	Routine sampling	Unknown	Wild source
16	**Food_627**	Food	B. melitensis	Retail	Soft cheese (raw milk)	Country A	Rural		2026-01-15	Market sampling	Unknown	Unpasteurized
17	**Food_659**	Food	B. abortus	Dairy	Raw milk	Country B	West		2025-11-25	Routine sampling	Unknown	Cattle herd
18	**Food_668**	Food	B. abortus	Dairy	Raw milk	Country C	Central		2025-12-03	Routine sampling	Unknown	Cattle herd
19	**Food_978**	Food	B. melitensis	Retail	Cheese	Country C	South		2025-08-15	Routine sampling	Unknown	Pasteurized
20	**Human_182**	Human	B. melitensis	Clinical	Blood	Country A	North	2026-01-12	2026-01-22	Dairy consumption	No	Consumed food from relative
21	**Human_264**	Human	B. melitensis	Clinical	Blood	Country C	Central	2026-01-20	2026-01-30	Dairy consumption	Yes	Market visit reported
22	**Human_359**	Human	B. melitensis	Clinical	Blood	Country A	Capital	2026-01-08	2026-01-18	Dairy consumption	Yes	Visited Christmas market
23	**Human_412**	Human	B. melitensis	Clinical	Blood	Country C	North	2026-01-27	2026-01-06	Dairy consumption	Yes	Holiday travel
24	**Human_430**	Human	B. melitensis	Clinical	Blood	Country A	West	2026-01-18	2026-01-28	Dairy consumption	Unknown	Incomplete interview
25	**Human_454**	Human	B. melitensis	Clinical	Blood	Country A	Capital	2026-01-25	2026-01-04	Dairy consumption	Unknown	No occupational exposure
26	**Human_488**	Human	B. melitensis	Clinical	Blood	Country B	South	2026-01-23	2026-01-02	Dairy consumption	No	Cheese shared at home
27	**Human_563**	Human	B. melitensis	Clinical	Blood	Country A	Rural	2025-11-10	2025-11-20	Animal contact	No	Veterinarian
28	**Human_671**	Human	B. melitensis	Clinical	Blood	Country B	East	2026-01-29	2026-01-08	Dairy consumption	No	Gifted food item
29	**Human_694**	Human	B. abortus	Clinical	Blood	Country A	Rural	2025-11-20	2025-12-01	Farm exposure	No	Cattle contact
30	**Human_793**	Human	B. melitensis	Clinical	Blood	Country B	East	2026-01-15	2026-01-25	Dairy consumption	Yes	Holiday travel
31	**Human_829**	Human	B. abortus	Clinical	Blood	Country A	Urban	2025-12-05	2025-12-15	Unknown	No	Sporadic human case; no known link to outbreak

Do you have a hypothesis on the outbreak source and transmission?

Phylogenetic tree construction and interpretation

SNP phylogeny workflow overview

- Data → Reference → Snippy → Gubbins → IQ-TREE → iTol

- I did this process for all the 30 genomes, then all for all of the *B. melitensus, B. abortus, and B. suis* genomes separately, and then finally the outbreak group.
- So - I did this analysis 5 times – this is what we mean we say it is an iterative process! ☺
- Reference genomes:
  - Took the highest quality genome in dataset for all genomes and each species:
    - All 30 genomes: Environment_907
    - *B. melitensis*: Environment_907
    - *B. abortus*: Human_829
    - *B. suis*: Animal_965

Reviewing QC for finding the reference

Sample	Bases (MB)	Qual Bases (MB)	Qual bases %	Reads	Qual reads	Qual reads %	Most common adapter (count)	2. Most common adapter (count)	Other adapters (count)	N50	No ctgs	longest	total bps	grade
Animal_143	197	195	0.9882	1317510	1317364	0.9999	9	6	19	194857	33	420193	3273752	A
Animal_217	197	195	0.9882	1317454	1317272	0.9999	8	6	26	221848	33	419979	3275140	A
Animal_770	199	197	0.9881	1331498	1331326	0.9999	10	7	27	163604	41	371032	3302424	A
Animal_920	197	194	0.9883	1314480	1314340	0.9999	9	6	17	201422	28	393571	3287685	A
Animal_965	197	195	0.9882	1318680	1318490	0.9999	10	8	23	170357	36	519425	3279458	A
Environment_320	190	188	0.9882	1269824	1269714	0.9999	8	6	20	11712	429	48838	3248572	C
Environment_735	199	196	0.9882	1326974	1326814	0.9999	12	9	31	140595	37	519265	3299603	A
Environment_757	197	194	0.9883	1314420	1314272	0.9999	13	8	25	294096	32	518913	3264718	A
Environment_759	198	196	0.9882	1324892	1324730	0.9999	11	10	23	249753	30	454373	3289334	A
Environment_848	196	194	0.9882	1311802	1311664	0.9999	7	6	12	190237	34	382195	3264395	A
Environment_902	199	197	0.9883	1331460	1331294	0.9999	8	5	19	228720	33	364405	3262852	A
Environment_907	197	195	0.9882	1319084	1318912	0.9999	9	7	27	278268	24	609266	3287393	A
Food_310	197	194	0.9883	1314408	1314276	0.9999	8	7	16	251704	34	401035	3264092	A
Food_585	199	197	0.9882	1329878	1329714	0.9999	9	8	19	166918	39	320046	3301878	A
Food_627	198	196	0.9882	1324094	1323924	0.9999	12	5	18	242288	28	513252	3288924	A
Food_659	202	199	0.9882	1347408	1347232	0.9999	9	5	20	135549	48	299117	3257640	A
Food_668	197	194	0.9882	1314396	1314236	0.9999	7	6	17	294089	30	518914	3263903	A
Food_978	198	196	0.9883	1324380	1324240	0.9999	7	7	27	249788	29	511997	3287448	A
Human_182	198	196	0.9882	1324110	1323942	0.9999	10	7	39	242288	27	518613	3289389	A
Human_264	198	196	0.9882	1324084	1323940	0.9999	7	7	26	242288	27	518613	3289722	A
Human_359	198	196	0.9882	1324060	1323932	0.9999	9	6	31	242288	27	518613	3289392	A
Human_412	198	196	0.9881	1324126	1323938	0.9999	6	6	17	242288	27	518613	3289420	A
Human_430	198	196	0.9883	1324096	1323934	0.9999	7	7	20	242288	27	518613	3289419	A
Human_454	198	196	0.9881	1324124	1323944	0.9999	7	6	18	242288	27	518613	3289486	A
Human_488	198	196	0.9881	1324110	1323932	0.9999	7	6	24	242288	27	518613	3289412	A
Human_563	197	195	0.9882	1317344	1317160	0.9999	12	6	24	198280	31	353716	3274832	A
Human_671	198	196	0.9882	1324106	1323918	0.9999	13	10	25	242288	27	518613	3289403	A
Human_694	197	194	0.9882	1314360	1314154	0.9998	7	6	25	294081	31	518914	3263192	A
Human_793	198	196	0.9882	1324104	1323906	0.9999	12	5	25	242288	27	518613	3289762	A
Human_829	196	193	0.9882	1307822	1307652	0.9999	9	6	30	260526	25	584564	3254971	A

Did you notice any low quality isolates? Did you remove any from the analysis?

Reviewing QC for finding the reference

Sample	Bases (MB)	Qual Bases (MB)	Qual bases %	Reads	Qual reads	Qual reads %	Most common adapter (count)	2. Most common adapter (count)	Other adapters (count)	N50	No ctgs	longest	total bps	grade
Animal_143	197	195	0.9882	1317510	1317364	0.9999	9	6	19	194857	33	420193	3273752	A
Animal_217	197	195	0.9882	1317454	1317272	0.9999	8	6	26	221848	33	419979	3275140	A
Animal_770	199	197	0.9881	1331498	1331326	0.9999	10	7	27	163604	41	371032	3302424	A
Animal_920	197	194	0.9883	1314480	1314340	0.9999	9	6	17	201422	28	393571	3287685	A
Animal_965	197	195	0.9882	1318680	1318490	0.9999	10	8	23	170357	36	519425	3279458	A
Environment_320	190	188	0.9882	1269824	1269714	0.9999	8	6	20	11712	429	48838	3248572	C
Environment_735	199	196	0.9882	1326974	1326814	0.9999	12	9	31	140595	37	519265	3299603	A
Environment_757	197	194	0.9883	1314420	1314272	0.9999	13	8	25	294096	32	518913	3264718	A
Environment_759	198	196	0.9882	1324892	1324730	0.9999	11	10	23	249753	30	454373	3289334	A
Environment_848	196	194	0.9882	1311802	1311664	0.9999	7	6	12	190237	34	382195	3264395	A
Environment_902	199	197	0.9883	1331460	1331294	0.9999	8	5	19	228720	33	364405	3262852	A
Environment_907	197	195	0.9882	1319084	1318912	0.9999	9	7	27	278268	24	609266	3287393	A
Food_310	197	194	0.9883	1314408	1314276	0.9999	8	7	16	251704	34	401035	3264092	A
Food_585	199	197	0.9882	1329878	1329714	0.9999	9	8	19	166918	39	320046	3301878	A
Food_627	198	196	0.9882	1324094	1323924	0.9999	12	5	18	242288	28	513252	3288924	A
Food_659	202	199	0.9882	1347408	1347232	0.9999	9	5	20	135549	48	299117	3257640	A
Food_668	197	194	0.9882	1314396	1314236	0.9999	7	6	17	294089	30	518914	3263903	A
Food_978	198	196	0.9883	1324380	1324240	0.9999	7	7	27	249788	29	511997	3287448	A
Human_182	198	196	0.9882	1324110	1323942	0.9999	10	7	39	242288	27	518613	3289389	A
Human_264	198	196	0.9882	1324084	1323940	0.9999	7	7	26	242288	27	518613	3289722	A
Human_359	198	196	0.9882	1324060	1323932	0.9999	9	6	31	242288	27	518613	3289392	A
Human_412	198	196	0.9881	1324126	1323938	0.9999	6	6	17	242288	27	518613	3289420	A
Human_430	198	196	0.9883	1324096	1323934	0.9999	7	7	20	242288	27	518613	3289419	A
Human_454	198	196	0.9881	1324124	1323944	0.9999	7	6	18	242288	27	518613	3289486	A
Human_488	198	196	0.9881	1324110	1323932	0.9999	7	6	24	242288	27	518613	3289412	A
Human_563	197	195	0.9882	1317344	1317160	0.9999	12	6	24	198280	31	353716	3274832	A
Human_671	198	196	0.9882	1324106	1323918	0.9999	13	10	25	242288	27	518613	3289403	A
Human_694	197	194	0.9882	1314360	1314154	0.9998	7	6	25	294081	31	518914	3263192	A
Human_793	198	196	0.9882	1324104	1323906	0.9999	12	5	25	242288	27	518613	3289762	A
Human_829	196	193	0.9882	1307822	1307652	0.9999	9	6	30	260526	25	584564	3254971	A

Did you use isolates from the dataset as the reference? Or did you use e.g NCBI?

Snippy

- Line up each isolate to the same reference
- Mark SNPs (single-letter differences) and indels (insertions/deletions)
- Result per isolate: SNP calls, a consensus sequence, and QC logs
- Goal: make differences comparable across samples
- https://github.com/tseemann/snippy

Core SNP phylogeny

If you call SNPs for multiple isolates from the same reference, you can produce an alignment of "core SNPs" which can be used to build a high-resolution phylogeny (ignoring possible recombination). A "core site" is a genomic position that is present in *all* the samples. A core site can have the same nucleotide in every sample ("monomorphic") or some samples can be different ("polymorphic" or "variant"). If we ignore the complications of "ins", "del" variant types, and just use variant sites, these are the "core SNP genome".

Input Requirements

- a set of Snippy folders which used the same `--ref` sequence.

Using `snippy-multi`

To simplify running a set of isolate sequences (reads or contigs) against the same reference, you can use the `snippy-multi` script. This script requires a *tab separated* input file as follows, and can handle paired-end reads, single-end reads, and assembled contigs.

```
# input.tab = ID R1 [R2]
Isolate1    /path/to/R1.fq.gz   /path/to/R2.fq.gz
Isolate1b   /path/to/R1.fastq.gz    /path/to/R2.fastq.gz
Isolate1c   /path/to/R1.fa      /path/to/R2.fa
# single end reads supported too
Isolate2    /path/to/SE.fq.gz
Isolate2b   /path/to/iontorrent.fastq
# or already assembled contigs if you don't have reads
Isolate3    /path/to/contigs.fa
Isolate3b   /path/to/reference.fna.gz
```

Snippy-core

- snippy-core keeps only sites present in all isolates → the core
- Two outputs:
  - core.full.aln = invariant + variant sites (full core)
  - core.aln = SNP-only columns (polymorphic sites)
- snippy-multi (concept): batch many isolates with one table; runs Snippy for each, then snippy-core
- https://github.com/tseemann/snippy

Then one would run this to generate the output script. The first parameter should be the `input.tab` file. The remaining parameters should be any remaining shared `snippy` parameters. The `ID` will be used for each isolate's `--outdir`.

```
% snippy-multi input.tab --ref Reference.gbk --cpus 16 > runme.sh

% less runme.sh   # check the script makes sense

% sh ./runme.sh   # leave it running over lunch
```

It will also run `snippy-core` at the end to generate the core genome SNP alignment files `core.*`.

Output Files

Extension	Description
.aln	A core SNP alignment in the `--aformat` format (default FASTA)
.full.aln	A whole genome SNP alignment (includes invariant sites)
.tab	Tab-separated columnar list of **core** SNP sites with alleles but NO annotations
.vcf	Multi-sample VCF file with genotype `GT` tags for all discovered alleles
.txt	Tab-separated columnar list of alignment/core-size statistics
.ref.fa	FASTA version/copy of the `--ref`
.self_mask.bed	BED file generated if `--mask auto` is used.

Gubbins: remove recombination before building the tree

- We don’t want SNPs introduced by recombination to drive the tree.
- After snippy-core, some SNPs may occur in dense clusters from recombination.
- Gubbins detects those recombinant regions and masks them → we keep the putative “true” SNPs from “vertical inheritance”.
- This is similar in spirit to CSI Phylogeny’s SNP pruning (e.g., pruning = 100 bp), but Gubbins is phylogeny-aware rather than a fixed-distance rule.
- https://github.com/nickjcroucher/gubbins

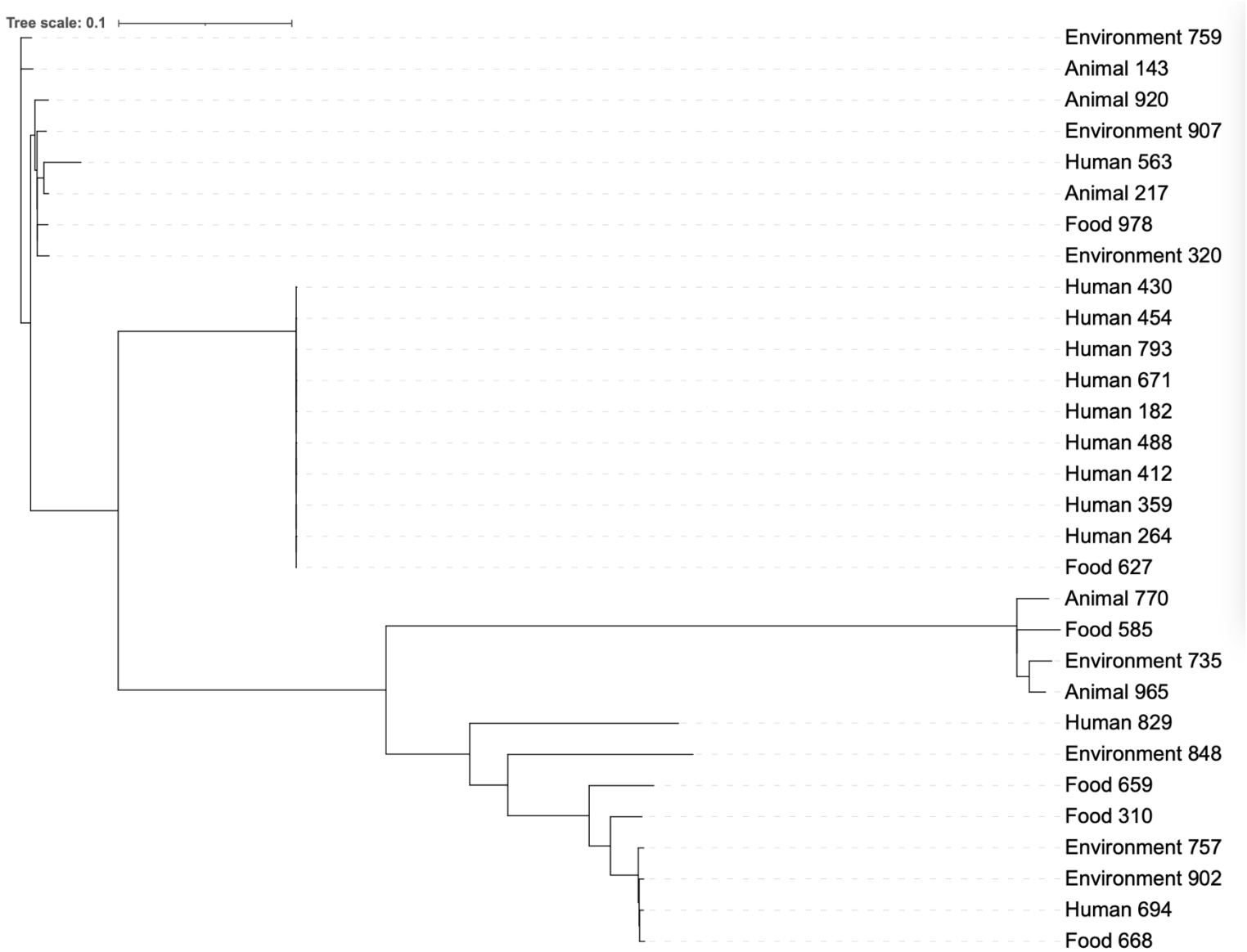
IQ-TREE: Maximum-likelihood tree

- Method: Maximum-likelihood (ML) phylogeny in IQ-TREE 2
- Input: Gubbins-filtered SNP-only core alignment
- Outputs:
  - snp_tree.treefile → best ML tree (this is the same as a **.newick** file)
  - snp_tree.contree → bootstrap consensus (supports on nodes)
  - snp_tree.iqtree → run log & best-fit model details
- Uploaded the .treefile to iTOL
- https://github.com/iqtree/iqtree3
- Also available online! https://iqtree.github.io

Which tools did you use for the phylogenetic analysis?

Let’s look at the trees and SNP matrixes – identification of the outbreak cluster

All 30 isolates – tree in iTol



All 30 isolates – SNP distance matrix in excel

94.12% of reference genome covered

	Animal_143	Animal_217	Animal_770	Animal_920	Animal_965	ivironment_320	ivironment_735	ivironment_757	ironment_759	ironment_848	ironment_902	ironment_907	Food_310	Food_585	Food_627	Food_659	Food_668	Food_978	Human_182	Human_264	Human_359	Human_412	Human_430	Human_454	Human_488	Human_563	Human_671	Human_694	Human_793	Human_829
Animal_143	0	368	7643	416	7642	420	7713	5644	239	6016	5641	392	5623	7794	2706	5720	5654	412	2711	2711	2706	2709	2710	2709	2710	682	2711	5639	2708	5741
Animal_217	368	0	7614	286	7613	248	7684	5615	405	5986	5612	218	5592	7765	2676	5691	5625	236	2681	2681	2676	2679	2680	2679	2680	438	2681	5610	2678	5714
Animal_770	7643	7614	0	7630	620	7634	689	7301	7650	7679	7300	7606	7280	770	8112	7378	7313	7621	8117	8117	8112	8115	8116	8115	8116	7642	8117	7298	8114	7406
Animal_920	416	286	7630	0	7629	290	7700	5633	401	6002	5630	258	5612	7781	2691	5708	5643	278	2696	2696	2691	2694	2695	2694	2695	610	2696	5628	2693	5730
Animal_965	7642	7613	620	7629	0	7633	401	7303	7649	7681	7302	7605	7280	740	8113	7380	7315	7620	8118	8118	8113	8116	8117	8116	8117	7643	8118	7300	8115	7410
Environment_320	420	248	7634	290	7633	0	7704	5637	407	6004	5634	224	5616	7785	2696	5713	5647	240	2701	2701	2696	2699	2700	2699	2700	584	2701	5632	2698	5734
Environment_735	7713	7684	689	7700	401	7704	0	7372	7720	7750	7371	7676	7351	787	8182	7447	7384	7691	8187	8187	8182	8185	8186	8185	8186	7712	8187	7369	8184	7479
Environment_757	5644	5615	7301	5633	7303	5637	7372	0	5651	3059	115	5609	673	7453	6117	1222	126	5628	6122	6122	6117	6120	6121	6120	6121	5641	6122	111	6119	3637
Environment_759	239	405	7650	401	7649	407	7720	5651	0	6021	5648	377	5630	7801	2713	5727	5661	397	2718	2718	2713	2716	2717	2716	2717	717	2718	5646	2715	5746
Environment_848	6016	5986	7679	6002	7681	6004	7750	3059	6021	0	3060	5978	3038	7831	6484	3131	3071	5997	6489	6489	6484	6487	6488	6487	6488	6014	6489	3054	6486	4013
Environment_902	5641	5612	7300	5630	7302	5634	7371	115	5648	3060	0	5606	674	7452	6114	1223	109	5625	6119	6119	6114	6117	6118	6117	6118	5638	6119	94	6116	3636
Environment_907	392	218	7606	258	7605	224	7676	5609	377	5978	5606	0	5586	7757	2668	5685	5619	212	2673	2673	2668	2671	2672	2671	2672	554	2673	5604	2670	5706
Food_310	5623	5592	7280	5612	7280	5616	7351	673	5630	3038	674	5586	0	7432	6096	1200	685	5607	6101	6101	6096	6099	6100	6099	6100	5618	6101	670	6098	3616
Food_585	7794	7765	770	7781	740	7785	787	7453	7801	7831	7452	7757	7432	0	8262	7530	7465	7772	8267	8267	8262	8265	8266	8265	8266	7792	8267	7450	8264	7558
Food_627	2706	2676	8112	2691	8113	2696	8182	6117	2713	6484	6114	2668	6096	8262	0	6192	6127	2686	5	5	0	3	4	3	4	2706	5	6112	2	6214
Food_659	5720	5691	7378	5708	7380	5713	7447	1222	5727	3131	1223	5685	1200	7530	6192	0	1234	5704	6196	6197	6192	6195	6196	6195	6196	5717	6197	1219	6194	3709
Food_668	5654	5625	7313	5643	7315	5647	7384	126	5661	3071	109	5619	685	7465	6127	1234	0	5638	6132	6132	6127	6130	6131	6130	6131	5651	6132	95	6129	3649
Food_978	412	236	7621	278	7620	240	7691	5628	397	5997	5625	212	5607	7772	2686	5704	5638	0	2691	2691	2686	2689	2690	2689	2690	572	2691	5623	2688	5725
Human_182	2711	2681	8117	2696	8118	2701	8187	6122	2718	6489	6119	2673	6101	8267	5	6196	6132	2691	0	10	5	8	9	8	9	2711	10	6117	7	6219
Human_264	2711	2681	8117	2696	8118	2701	8187	6122	2718	6489	6119	2673	6101	8267	5	6197	6132	2691	10	0	5	8	9	8	9	2711	10	6117	7	6219
Human_359	2706	2676	8112	2691	8113	2696	8182	6117	2713	6484	6114	2668	6096	8262	0	6192	6127	2686	5	5	0	3	4	3	4	2706	5	6112	2	6214
Human_412	2709	2679	8115	2694	8116	2699	8185	6120	2716	6487	6117	2671	6099	8265	3	6195	6130	2689	8	8	3	0	7	6	7	2709	8	6115	5	6217
Human_430	2710	2680	8116	2695	8117	2700	8186	6121	2717	6488	6118	2672	6100	8266	4	6196	6131	2690	9	9	4	7	0	7	8	2710	9	6116	6	6218
Human_454	2709	2679	8115	2694	8116	2699	8185	6120	2716	6487	6117	2671	6099	8265	3	6195	6130	2689	8	8	3	6	7	0	7	2709	8	6115	5	6217
Human_488	2710	2680	8116	2695	8117	2700	8186	6121	2717	6488	6118	2672	6100	8266	4	6196	6131	2690	9	9	4	7	8	7	0	2710	9	6116	6	6218
Human_563	682	438	7642	610	7643	584	7712	5641	717	6014	5638	554	5618	7792	2706	5717	5651	572	2711	2711	2706	2709	2710	2709	2710	0	2711	5636	2708	5744
Human_671	2711	2681	8117	2696	8118	2701	8187	6122	2718	6489	6119	2673	6101	8267	5	6197	6132	2691	10	10	5	8	9	8	9	2711	0	6117	7	6219
Human_694	5639	5610	7298	5628	7300	5632	7369	111	5646	3054	94	5604	670	7450	6112	1219	95	5623	6117	6117	6112	6115	6116	6115	6116	5636	6117	0	6114	3634
Human_793	2708	2678	8114	2693	8115	2698	8184	6119	2715	6486	6116	2670	6098	8264	2	6194	6129	2688	7	7	2	5	6	5	6	2708	7	6114	0	6216
Human_829	5741	5714	7406	5730	7410	5734	7479	3637	5746	4013	3636	5706	3616	7558	6214	3709	3649	5725	6219	6219	6214	6217	6218	6217	6218	5744	6219	3634	6216	0

- Looks like we already have an outbreak cluster emerging of 10 isolates!
- But some isolates also have less than 100 SNP differences – if we check the metadata – they are all *B. abortus*
- But for now - let's remove the other species and repeat the analysis just for *B. melitensis*

All 18 *B. melitensis* isolates – tree in iTol

All 18 *B. melitensis* isolates – SNP distance matrix in excel

97.31% of reference genome covered

	Animal_143	Animal_217	Animal_920	Environment_320	Environment_759	Environment_907	Food_627	Food_978	Human_182	Human_264	Human_359	Human_412	Human_430	Human_454	Human_488	Human_563	Human_671	Human_793
Animal_143	0	381	437	438	255	414	2805	433	2810	2810	2805	2808	2810	2808	2810	713	2810	2807
Animal_217	381	0	298	257	422	231	2772	248	2777	2777	2772	2775	2777	2775	2777	460	2777	2774
Animal_920	437	298	0	297	416	269	2787	288	2792	2792	2787	2790	2792	2790	2792	638	2792	2789
Environment_320	438	257	297	0	419	232	2789	247	2794	2794	2789	2792	2794	2792	2794	609	2794	2791
Environment_759	255	422	416	419	0	393	2808	412	2813	2813	2808	2811	2813	2811	2813	750	2813	2810
Environment_907	414	231	269	232	393	0	2765	223	2770	2770	2765	2768	2770	2768	2770	583	2770	2767
Food_627	2805	2772	2787	2789	2808	2765	0	2782	5	5	0	3	5	3	5	2808	5	2
Food_978	433	248	288	247	412	223	2782	0	2787	2787	2782	2785	2787	2785	2787	600	2787	2784
Human_182	2810	2777	2792	2794	2813	2770	5	2787	0	10	5	8	10	8	10	2813	10	7
Human_264	2810	2777	2792	2794	2813	2770	5	2787	10	0	5	8	10	8	10	2813	10	7
Human_359	2805	2772	2787	2789	2808	2765	0	2782	5	5	0	3	5	3	5	2808	5	2
Human_412	2808	2775	2790	2792	2811	2768	3	2785	8	8	3	0	8	6	8	2811	8	5
Human_430	2810	2777	2792	2794	2813	2770	5	2787	10	10	5	8	0	8	10	2813	10	7
Human_454	2808	2775	2790	2792	2811	2768	3	2785	8	8	3	6	8	0	8	2811	8	5
Human_488	2810	2777	2792	2794	2813	2770	5	2787	10	10	5	8	10	8	0	2813	10	7
Human_563	713	460	638	609	750	583	2808	600	2813	2813	2808	2811	2813	2811	2813	0	2813	2810
Human_671	2810	2777	2792	2794	2813	2770	5	2787	10	10	5	8	10	8	10	2813	0	7
Human_793	2807	2774	2789	2791	2810	2767	2	2784	7	7	2	5	7	5	7	2810	7	0

- It looks like we have our outbreak cluster of 10 or less SNP differences
- Let's repeat the analysis again with just the suspected outbreak isolates

B. melitensis outbreak isolates – SNP distance matrix in excel

99.8% of reference genome covered

	Food_627	Human_18	Human_26	Human_35	Human_412	Human_43	Human_45	Human_48	Human_67	Human_793
Food_627	0	5	5	0	4	5	3	5	5	2
Human_182	5	0	10	5	9	10	8	10	10	7
Human_264	5	10	0	5	9	10	8	10	10	7
Human_359	0	5	5	0	4	5	3	5	5	2
Human_412	4	9	9	4	0	9	7	9	9	6
Human_430	5	10	10	5	9	0	8	10	10	7
Human_454	3	8	8	3	7	8	0	8	8	5
Human_488	5	10	10	5	9	10	8	0	10	7
Human_671	5	10	10	5	9	10	8	10	0	7
Human_793	2	7	7	2	6	7	5	7	7	0

- Looks we have our outbreak cluster!
- What about the other species in our dataset?

B. abortus isolates - SNP matrix in excel

99.14% of reference genome covered

	Environment_	Environment_	Environment_	Food_310	Food_659	Food_668	Human_694	Human_829
Environment_757	0	3211	121	696	1278	135	115	3819
Environment_848	3211	0	3212	3188	3295	3226	3204	4230
Environment_902	121	3212	0	697	1279	118	98	3818
Food_310	696	3188	697	0	1254	711	691	3795
Food_659	1278	3295	1279	1254	0	1293	1273	3902
Food_668	135	3226	118	711	1293	0	102	3832
Human_694	115	3204	98	691	1273	102	0	3814
Human_829	3819	4230	3818	3795	3902	3832	3814	0

B. suis isolates - SNP matrix in excel

99.63% of reference genome covered

	Animal_770	Animal_965	Environment_	Food_585
Animal_770	0	673	760	932
Animal_965	673	0	449	895
Environment_735	760	449	0	956
Food_585	932	895	956	0

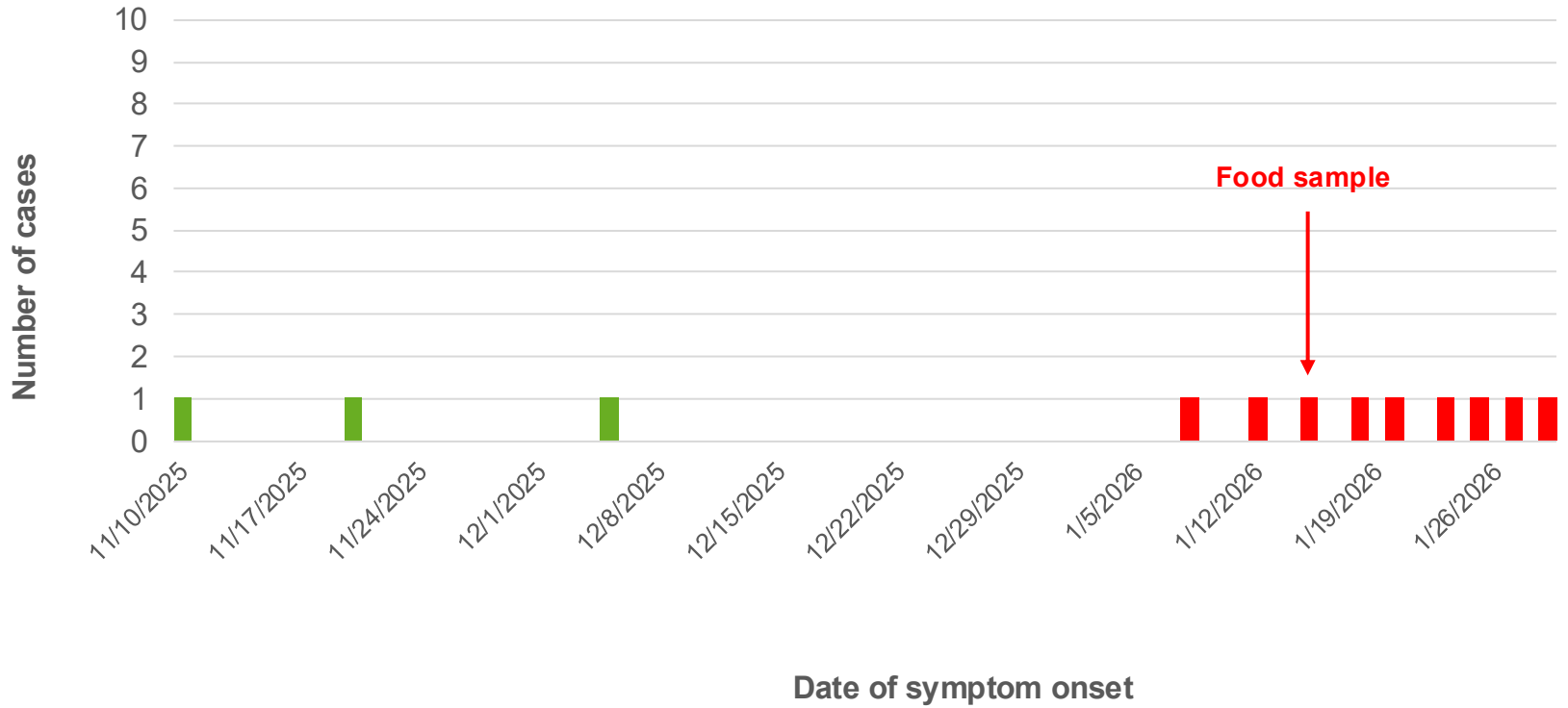
Do you agree with the outbreak cluster?

Correlation of epidemiological data with phylogenetic trees

Comparing the results with the metadata

Strain_ID	Host_type	Species	Sample_source	Sample_material	Country	Region	Symptom_onset_date	Collection_date	Reported_exposure	Travel_history	Notes	Outbreak?
Animal_143	Animal	B. melitensis	Farm	Goat milk	Country A	South		2025-05-10	Routine surveillance	Unknown	Small farm	No
Animal_217	Animal	B. melitensis	Farm	Sheep tissue	Country B	West		2025-06-18	Routine surveillance	Unknown	No human cases	No
Animal_770	Animal	B. suis	Wildlife	Wild boar tissue	Country A	Forest		2025-07-15	Wildlife surveillance	Unknown	Wildlife	No
Animal_920	Animal	B. melitensis	Farm	Goat tissue	Country C	Central		2025-09-02	Routine surveillance	Unknown	Different genotype	No
Animal_965	Animal	B. suis	Slaughterhouse	Pig tissue	Country B	South		2025-06-30	Routine surveillance	Unknown	Domestic pig	No
Environment_320	Environment	B. melitensis	Farm	Soil	Country B	Rural		2025-07-01	Environmental surveillance	Unknown	No link to cases	No
Environment_735	Environment	B. suis	Forest	Soil	Country A	Forest		2025-07-20	Wildlife surveillance	Unknown	No human cases	No
Environment_757	Environment	B. abortus	Farm	Soil	Country A	Rural		2025-11-15	Post-miscarriage sampling	Unknown	Cattle farm	No
Environment_759	Environment	B. melitensis	Dairy	Surface swab	Country A	Rural		2025-09-12	Environmental surveillance	Unknown	Older isolate	No
Environment_848	Environment	B. abortus	Farm	Surface swab	Country B	West		2025-11-18	Post-miscarriage sampling	Unknown	Abortions reported	No
Environment_902	Environment	B. abortus	Farm	Water	Country C	Central		2025-11-22	Post-miscarriage sampling	Unknown	Abortions reported	No
Environment_907	Environment	B. melitensis	Environment	Water	Country C	North		2025-10-05	Environmental surveillance	Unknown	Unrelated	No
Food_310	Food	B. abortus	Dairy	Raw milk	Country A	Rural		2025-11-20	Routine sampling	Unknown	Cattle herd	No
Food_585	Food	B. suis	Retail	Game meat	Country B	East		2025-08-02	Routine sampling	Unknown	Wild source	No
Food_659	Food	B. abortus	Dairy	Raw milk	Country B	West		2025-11-25	Routine sampling	Unknown	Cattle herd	No
Food_668	Food	B. abortus	Dairy	Raw milk	Country C	Central		2025-12-03	Routine sampling	Unknown	Cattle herd	No
Food_978	Food	B. melitensis	Retail	Cheese	Country C	South		2025-08-15	Routine sampling	Unknown	Pasteurized	No
Human_563	Human	B. melitensis	Clinical	Blood	Country A	Rural	2025-11-10	2025-11-20	Animal contact	No	Veterinarian	No
Human_694	Human	B. abortus	Clinical	Blood	Country A	Rural	2025-11-20	2025-12-01	Farm exposure	No	Cattle contact	No
Human_829	Human	B. abortus	Clinical	Blood	Country A	Urban	2025-12-05	2025-12-15	Unknown	No	Sporadic human case; no known link to outbreak	No
Food_627	Food	B. melitensis	Retail	Soft cheese (raw milk)	Country A	Rural		2026-01-15	Market sampling	Unknown	Unpasteurized	Yes
Human_182	Human	B. melitensis	Clinical	Blood	Country A	North	2026-01-12	2026-01-22	Dairy consumption	No	Consumed food from relative	Yes
Human_264	Human	B. melitensis	Clinical	Blood	Country C	Central	2026-01-20	2026-01-30	Dairy consumption	Yes	Market visit reported	Yes
Human_359	Human	B. melitensis	Clinical	Blood	Country A	Capital	2026-01-08	2026-01-18	Dairy consumption	Yes	Visited Christmas market	Yes
Human_412	Human	B. melitensis	Clinical	Blood	Country C	North	2026-01-27	2026-01-06	Dairy consumption	Yes	Holiday travel	Yes
Human_430	Human	B. melitensis	Clinical	Blood	Country A	West	2026-01-18	2026-01-28	Dairy consumption	Unknown	Incomplete interview	Yes
Human_454	Human	B. melitensis	Clinical	Blood	Country A	Capital	2026-01-25	2026-01-04	Dairy consumption	Unknown	No occupational exposure	Yes
Human_488	Human	B. melitensis	Clinical	Blood	Country B	South	2026-01-23	2026-01-02	Dairy consumption	No	Cheese shared at home	Yes
Human_671	Human	B. melitensis	Clinical	Blood	Country B	East	2026-01-29	2026-01-08	Dairy consumption	No	Gifted food item	Yes
Human_793	Human	B. melitensis	Clinical	Blood	Country B	East	2026-01-15	2026-01-25	Dairy consumption	Yes	Holiday travel	Yes

Epi curve


10
9
8
7
6
5
4
3
2
1
0
Number of cases
Food sample
11/10/2025
11/17/2025
11/24/2025
12/1/2025
12/8/2025
12/15/2025
12/22/2025
12/29/2025
1/5/2026
1/12/2026
1/19/2026
1/26/2026
Date of symptom onset

Has your hypothesis on the outbreak source and transmission changed after results of the phylogenetic analysis?

Outbreak source and transmission

- The dataset represents a foodborne outbreak of *Brucella melitensis* occurring shortly after the Christmas period (December 2025 → January 2026).
- The most likely transmission route is ingestion of unpasteurized sheep/goat cheese sold at a Christmas market, with cases arising through both direct consumption by market visitors and later exposure via food sharing after return from travel.
- The outbreak primarily affects human cases, with **a single food isolate (Food_627)** acting as the epidemiological and genomic bridge between cases, consistent with a point-source exposure. No sustained human-to-human transmission is included in this scenario.
- Reports of livestock miscarriages and associated *Brucella abortus* detections reflect realistic background surveillance signals that are temporally related but epidemiologically and genomically unrelated to the human outbreak.

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