Introduction to the exercise

Overview of SimEx scenario and tasks

January 19, 2026, Niamh Lacy-Roberts, Research Assistant at DTU, nlac@food.dtu.dk

Session Overview

- Explanation of the exercise objectives, tasks, and deliverables
- Overview of the analysis's participants are expected to perform independently, along with examples of user-friendly and command-line tools they may choose to use for these analyses:
  - 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

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 comm	2. Most com	Other adapte	insert size	N50	No ctgs	longest	total bps	gra
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

Science Data

Shared by Niamh Emily Eleanor Lacy-Roberts

GenEpi-BioTrain_Virtual_training_22_Brucellosis

Name	Size	Modified
FASTA	95.2 MB	5 days ago
FASTQ	3.6 GB	5 days ago
SimEx_Brucella_metadata_and_quality_for_participants.xlsx	15 kB	seconds ago
2 folders and 1 file	3.7 GB	

Data for the SimEx

Link is also available on the learning portal:

Session 1: Introduction to molecular epidemiology and epidemiological analysis in outbreak Investigations

Session 1 includes six topics:

1. Introduction to *Brucella* spp. and brucellosis

- Epidemiological background of *Brucella* spp., diversity, overview of brucellosis

2. One Health perspective:

- One-Health concept.
- Incidence, prevalence, transmission and global context. Human, animal, and environmental relationships.

3. Taxonomic analysis and typing:

- Concept, application, WGS-based tools (alignment and gene-by-gene approaches) and interpretation.

4. Epidemiological investigations:

- Epidemiological data, epidemiological curves and hypothesis generation.

5. Phylogenetic analysis:

- Overview of phylogenetic analyses, brief comparison of classical and genomic approaches, explanation of SNP-based phylogeny.
- How SNP trees are used in epidemiology, and overview of tree construction and interpretation.

6. Introduction to the exercise:

You will receive a small set of approximately **30 *Brucella* spp. genomes** (assemblies as FASTA files and raw FASTQ files) and the respective metadata via ScienceData (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.

Some tool suggestions

Online / web-based:

- KmerFinder – species identification
- CSI Phylogeny – automated reference-based SNP phylogeny designed for Illumina data
- MINTyper – SNP-based outbreak typing designed for ONT data
- PathogenWatch – integrated surveillance, trees + metadata
- Microreact – interactive visualisation of trees and metadata
- iTOL – advanced phylogenetic tree visualisation
- EnteroBase – large-scale cgMLST & SNP analyses for selected species

Command-line

- Snippy – reference-based SNP calling
- IQ-TREE / RAxML – maximum likelihood tree building
- chewBBACA – cgMLST allele calling
- Parsnp – core-genome SNP alignment (assembly-based)
- Gubbins – recombination detection & masking

Thank you for listening! ☺

Acknowledgements

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