



APPLICATIONS - DAY 9 AND 10

# Analysis of data in groups

05/06/2024

# Intended Learning Objectives

Specific objectives of this session:

1. Work with a real-life scenario and data
2. Apply the knowledge gained on bioinformatic tools, epidemiology, microbiology and surveillance
3. Work in interdisciplinary country teams
4. Work with members of other country teams

# Outline

This module consists of the following elements

1. Thursday morning and early afternoon: you will work in country groups
  - You will receive one of three bacterial genomic datasets to work on
  - You will analyze the data with the tools you have learned during the course (or with other tools of your knowledge!)
2. Second half of the afternoon: you will be pooled in larger groups, based on the dataset you have in common
  - You will work with other country teams and compare results
  - You will prepare some **slides to present** the scenario you received and the results of your work **on Friday**

# Scenario 1

A sheep with listeriosis was detected in a Spanish farm in region X. A *Listeria monocytogenes* cgMLST type L1-SL1-ST1-CT8169 (clonal complex CC1, serogroup IVb) was isolated from the sheep brain.

Following this report, you and your colleagues decide to investigate the occurrence of *Listeria* spp. in the affected farm as well as nearby farms, as studies have shown that **CC1** is the most prevalent clinical clone in industrialized countries and the most pathogenic, causing **severe listeriosis** cases in humans.

A total of **425 *Listeria* isolates** are obtained across the farms sampled and sequenced. You need to report the findings to the authorities.

## Scenario 2

In your country, **Shiga-toxin** producing ***Escherichia coli*** belonging to serotype **O157:H7** are one of the major causes of haemolytic-uremic syndrome (HUS). Strains collected from human cases are routinely sequenced and submitted to NCBI.

You have been asked to analyse the genomic data obtained between **2016-2020** (n = 432) and describe their main genomic features and the evolution of the population of this serotype.

**Escherichia coli**  
**O157:H7**

## Scenario 3

In your country, less common serotypes of *Salmonella* spp. are not routinely sequenced.

However, while preparing your yearly surveillance report for **2020**, you realise that the number of *S. enterica* identified as belonging to serotype **Bovismorbificans** by sero-agglutination were particularly high in that year (n=76) and you decide to sequence them all to determine if there were any genomic clusters among them.

# Ideas and tools

- Galaxy
- Enterobase
- BIGSdb-Pasteur
- Pathogenwatch
- Microreact
- GrapeTree
- Proksee
- ResFinder
- CARD
- PlasmidFinder
- PLSDB
- VFDB
- Any other tool of your choice!

You will be given questions to guide your analyses, but feel free to go further!

Based on your knowledge (microbiology, epidemiology, bioinformatics), what else would you do?

Based on your results, what would you do next? Which surveillance systems could you use to alert neighbouring other countries?

# Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Chiara Crestani, Carla Rodrigues, Alexandra Moura, and Carolina Nodari