



ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data

Verónica Mixão

Genomics and Bioinformatics Unit, Department of Infectious Diseases

National Institute of Health Doutor Ricardo Jorge (INSA) – Lisbon, Portugal

“New Tools for Surveillance and Risk Assessment”

March 28th, 2023

Genomic surveillance

“Genomic surveillance is the process of constantly monitoring pathogens and analyzing their genetic similarities and differences.”

WHO, 2022



A pivotal outcome of genomics surveillance is the **identification of pathogen genetic clusters/lineages** and their **characterization** in terms of geotemporal spread or linkage to clinical and demographic data

Genomics surveillance – main typing workflows

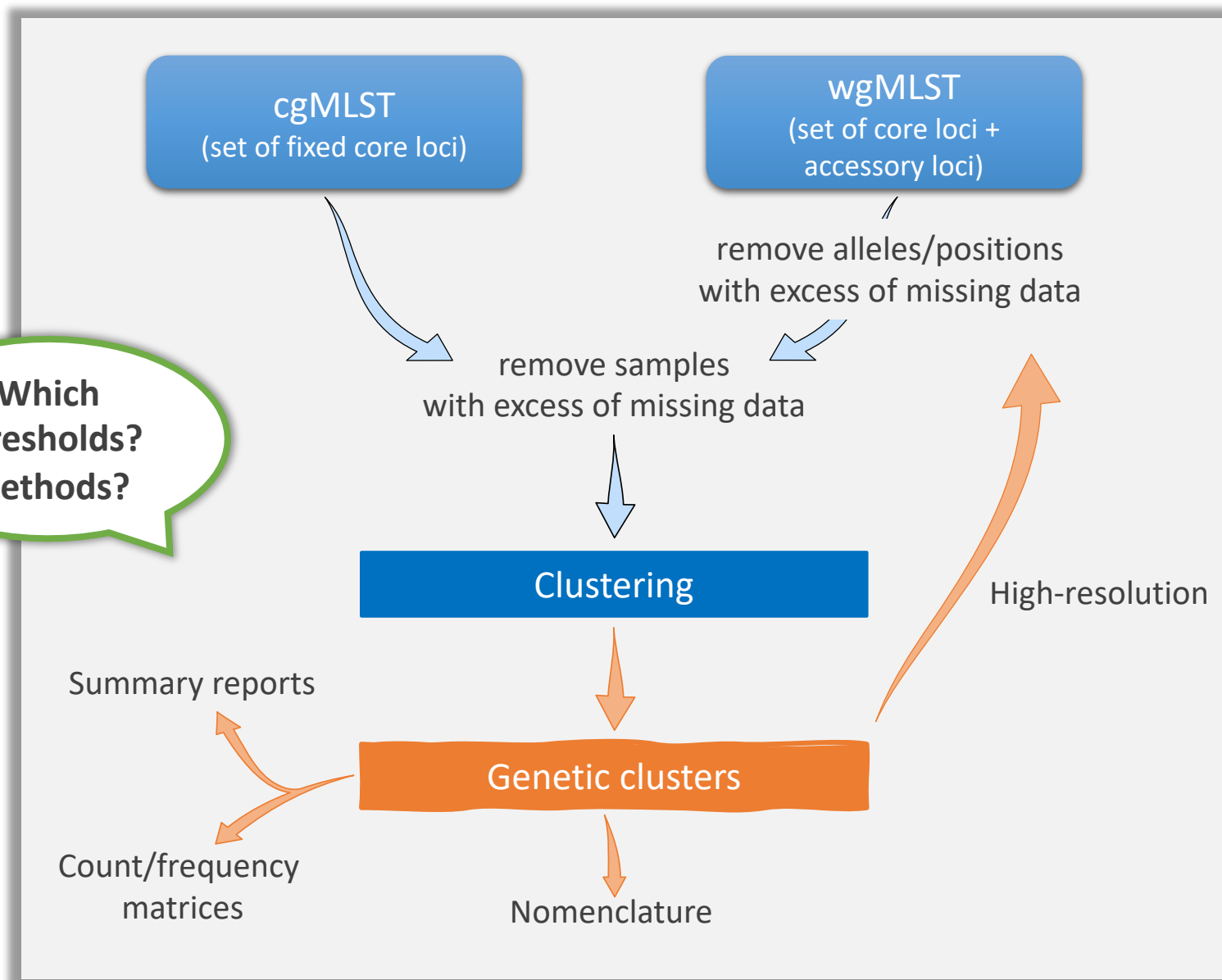


Genomic data
(SNPs or alleles)

1	2	5	7	8	9	10	11
1	2	5	7	8	9	10	12
0	3	5	7	8	9	10	12
1	4	6	7	8	9	10	12
0	4	6	0	8	9	0	12

filter the genomic data based on metadata variables?

Which thresholds?
Methods?



*Single-Nucleotide Polymorphisms

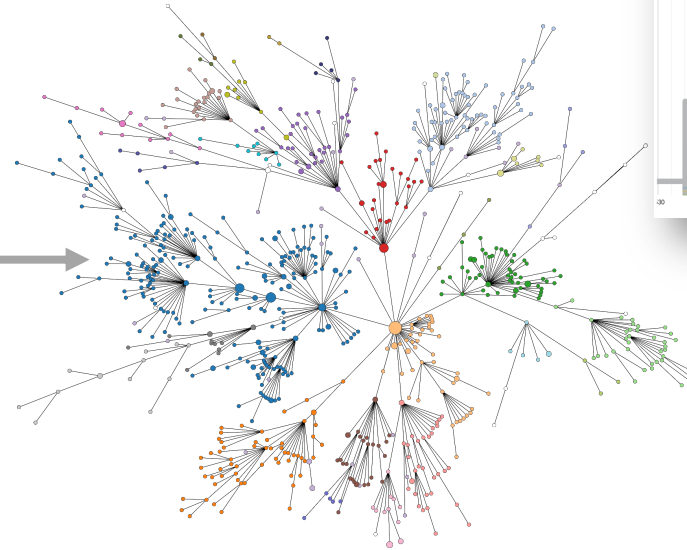
Genomics surveillance – main typing workflows

Multiple bioinformatics solutions, but similar steps and goals...

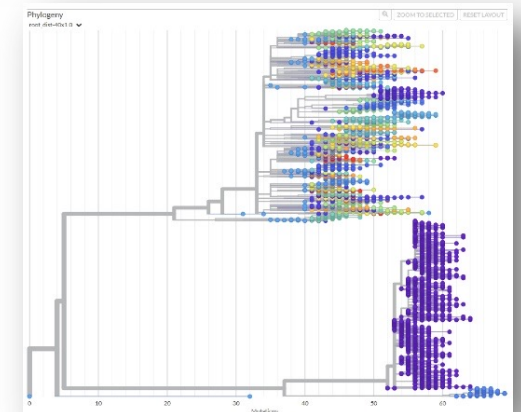
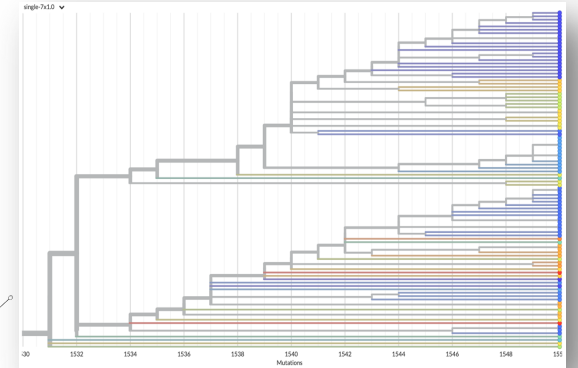
Genomic data
(SNPs or alleles)

1	2	5	7	8	9	10	11
1	2	5	7	8	9	10	12
0	3	5	7	8	9	10	12
1	4	6	7	8	9	10	12
0	4	6	0	8	9	0	12

Distance
between samples

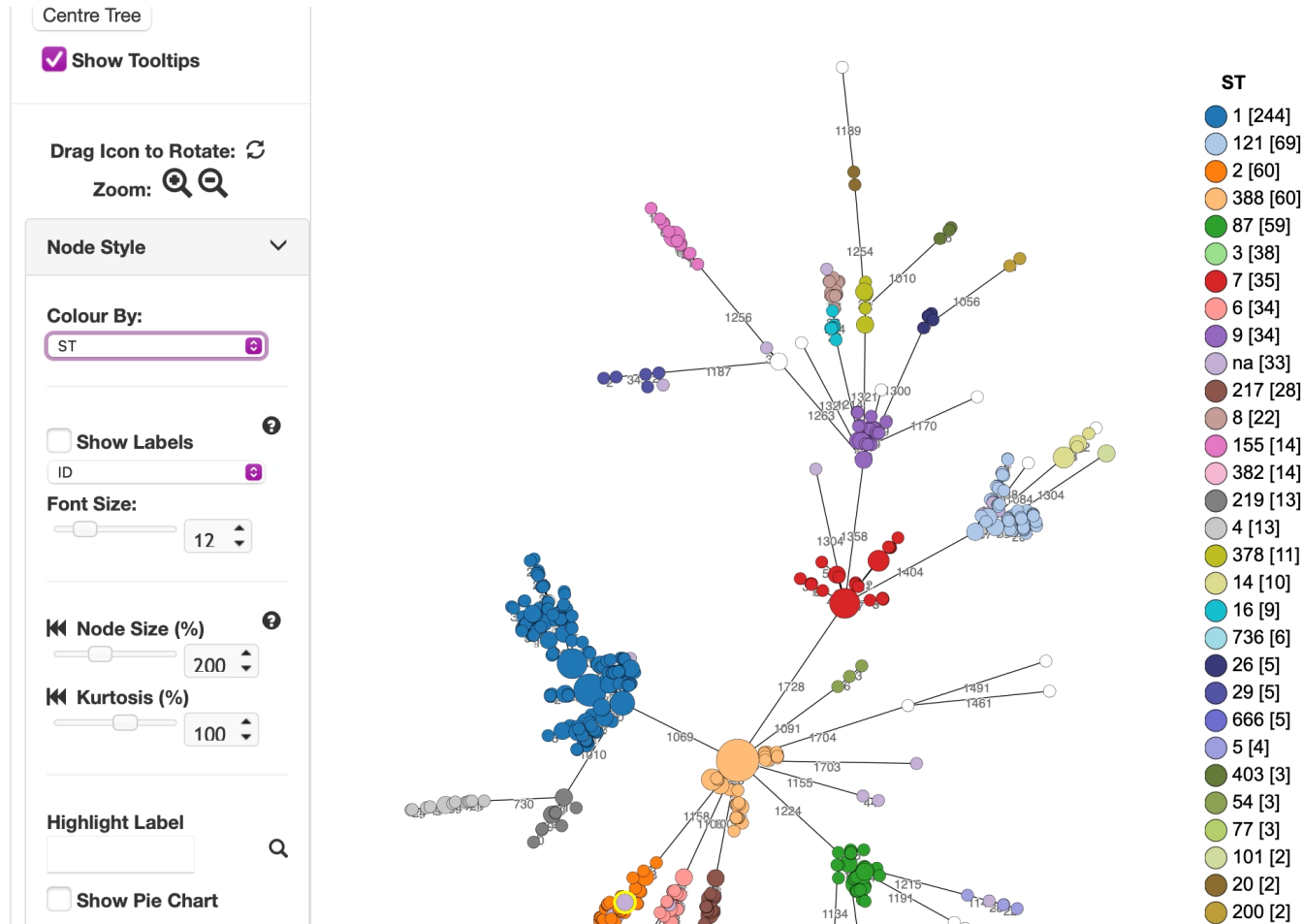


Tree or tree-like representations



Identification and characterization of genetic clusters

The example of *Listeria monocytogenes*...

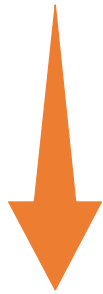


The **detection of genetic clusters** with possible public health relevance is still

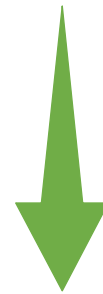
a challenging step that mostly relies on **non-automated** and **difficult to reproduce** approaches

The routine of genomic surveillance

There is a need for solutions that facilitate, automate and speed-up the detection of genetic clusters and their linkage to clinical/epidemiological data



REPRODUCIBILITY



TIME-EFFICIENCY



FLEXIBILITY

ReporTree

- Flexible solution to **automatically identify genetic clusters** at any (or all) distance thresholds (e.g., high resolution thresholds used for [outbreak detection](#))
- Generate **surveillance-oriented reports** based on the available metadata, such as timespan, geography or vaccination/clinical status
- Allows maintaining **cluster nomenclature** between runs and generates [hierarchical codes for nomenclature](#)



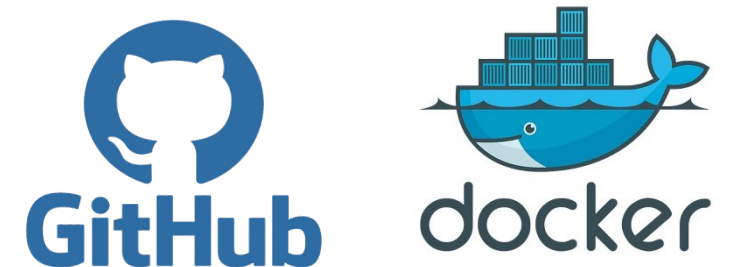
ReporTree: a surveillance-oriented tool to strengthen the linkage between pathogen genetic clusters and epidemiological data

Mixão V, Pinto M, Sobral D, Pasquale AD, Gomes JP, Borges V

[Author information](#) ▶

Preprint from Research Square, 30 Sep 2022

DOI: 10.21203/rs.3.rs-1404655/v2 PPR: PPR552299



<https://github.com/insapathogenomics/ReporTree>

Genomics surveillance – main typing workflows

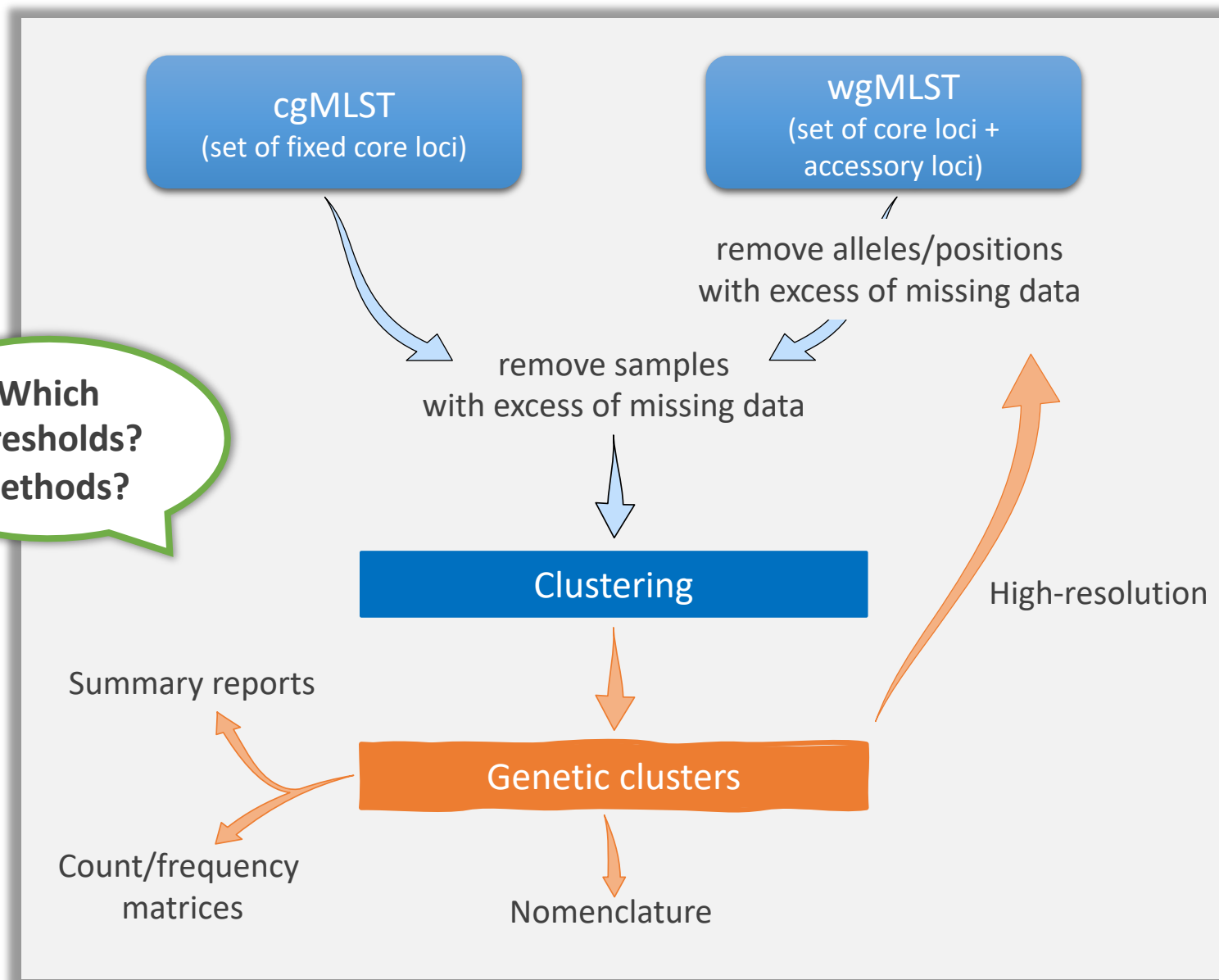


Genomic data
(SNPs or alleles)

1	2	5	7	8	9	10	11
1	2	5	7	8	9	10	12
0	3	5	7	8	9	10	12
1	4	6	7	8	9	10	12
0	4	6	0	8	9	0	12

filter the genomic data based on metadata variables?

Which thresholds?
Methods?



*Single-Nucleotide Polymorphisms

Genomics surveillance – main typing workflows

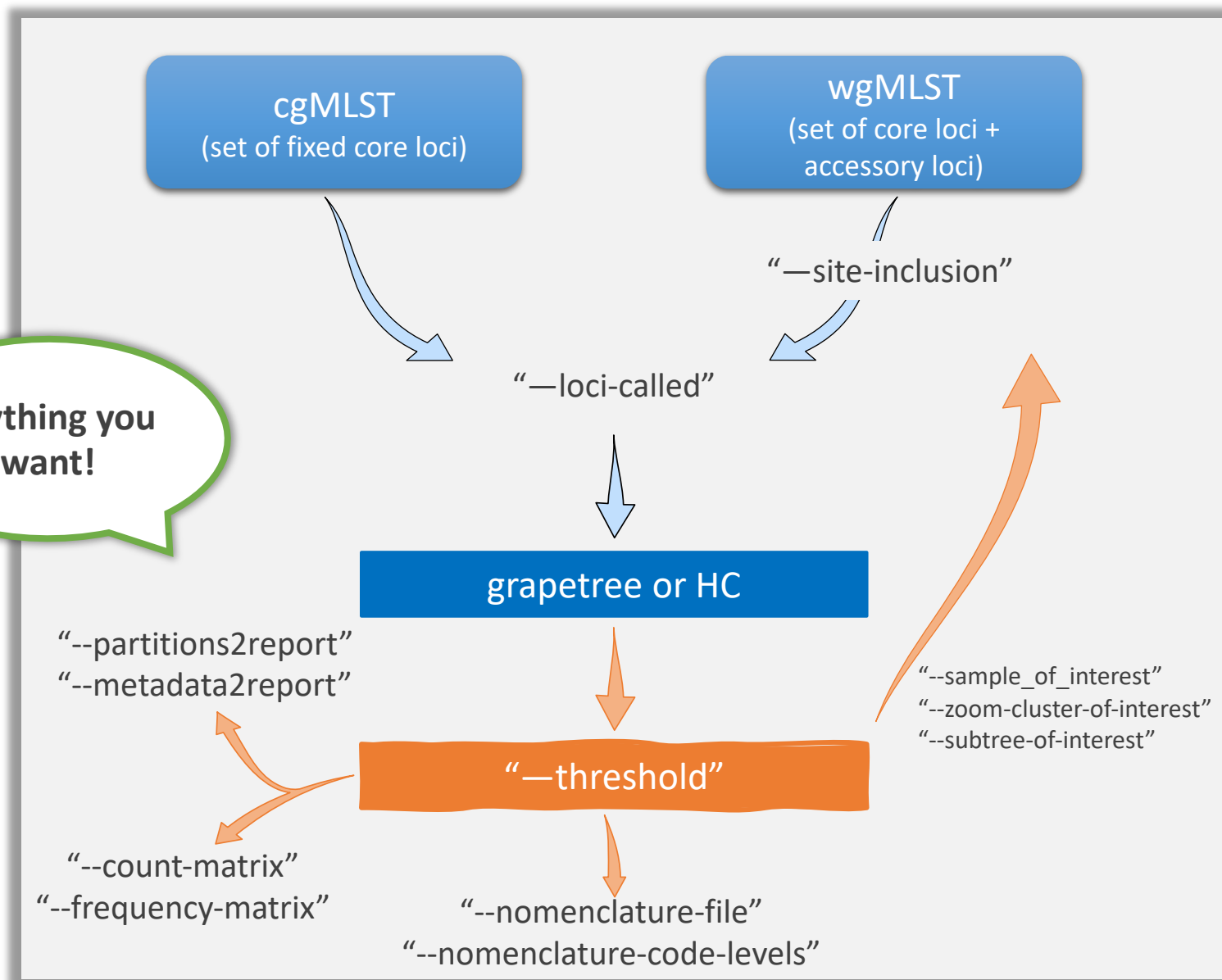


Genomic data
(SNPs or alleles)

1	2	5	7	8	9	10	11
1	2	5	7	8	9	10	12
0	3	5	7	8	9	10	12
1	4	6	7	8	9	10	12
0	4	6	0	8	9	0	12

“--subset” and “-f”

Anything you want!



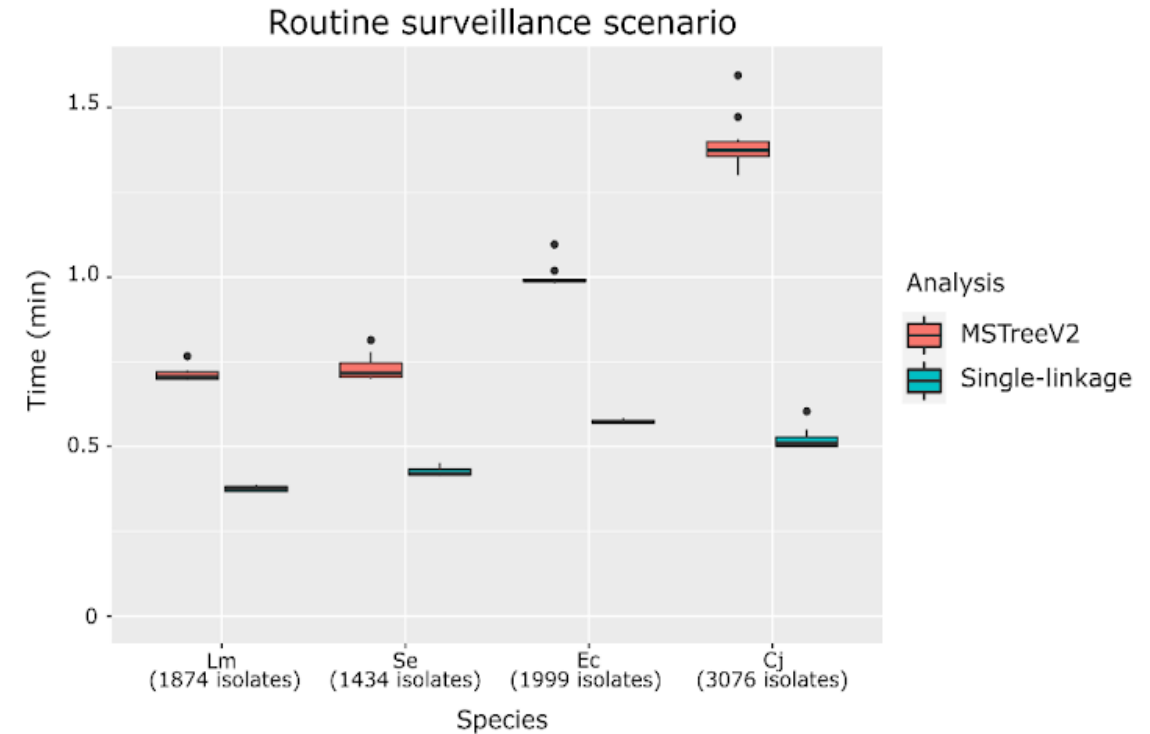
*Single-Nucleotide Polymorphisms

ReporTree validation and benchmarking



TIME-EFFICIENCY

- Using diverse datasets with >1,000 isolates of foodborne bacterial pathogens, ReporTree identifies genetic clusters at potential outbreak level and performs their characterization in less than [1min](#)

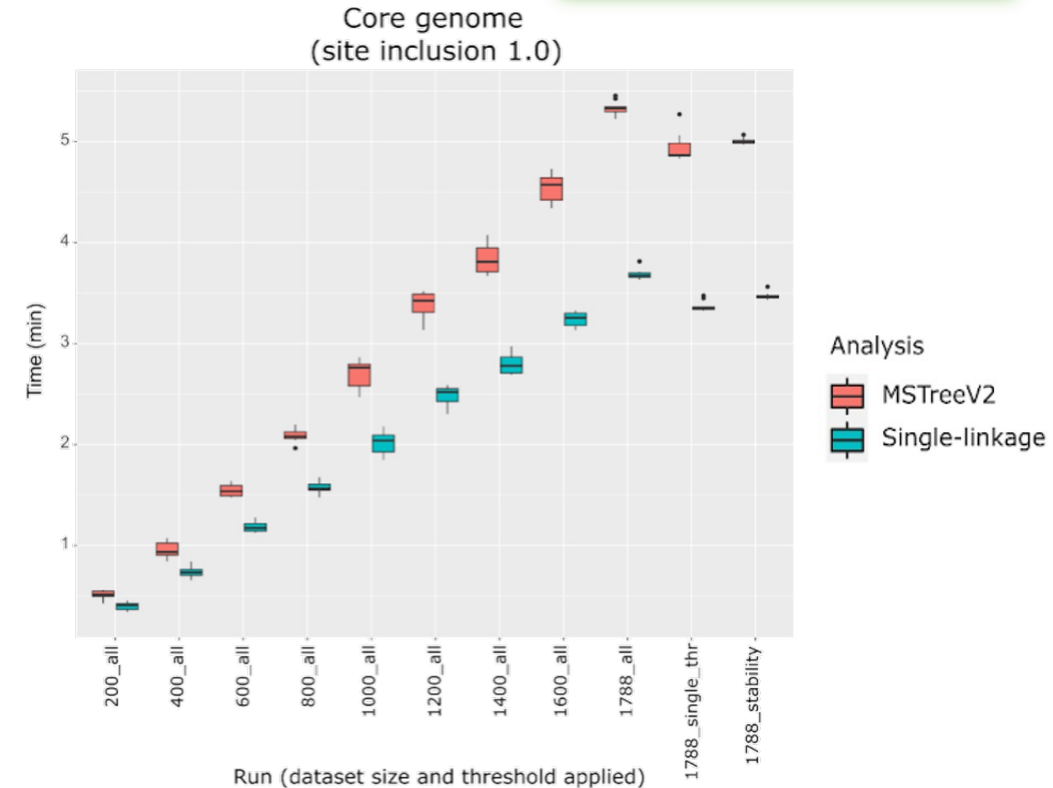


ReporTree validation and benchmarking



TIME-EFFICIENCY

- Using **diverse datasets with >1,000 isolates of foodborne bacterial pathogens**, ReporTree identifies genetic clusters at potential outbreak level and performs their characterization in less than **1min**
- For an alignment-based core SNP pipeline, using a diverse dataset with **1788 *Mycobacterium tuberculosis* isolates**, the running times is less than **6 min**

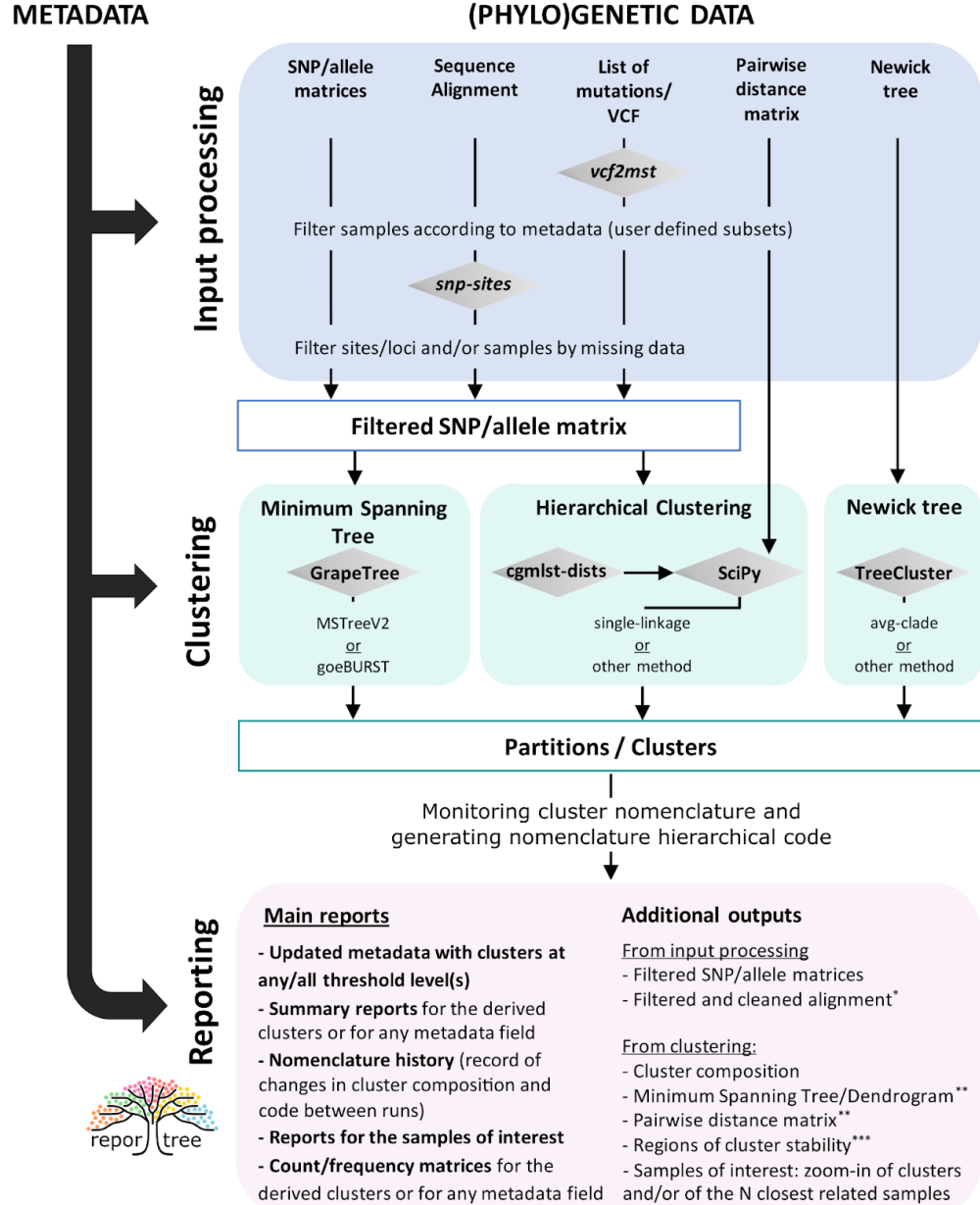


ReporTree can be **smoothly implemented in routine surveillance**, with negligible computational and time costs, contributing to a **sustainable and efficient public health genomics-informed pathogen surveillance**

Designed for surveillance of multiple pathogens



FLEXIBILITY



- Implemented in the **routine surveillance** of *L. monocytogenes* at INSA and has already facilitated the Quality Control Assessments of other [foodborne pathogens](#)
- Being tested for application in the genomics surveillance of other important bacterial pathogens, such as [Neisseria meningitidis](#) and [Mycobacterium spp.](#)



Take-home message



*ReporTree is an **automated and flexible** pipeline, applicable to **multiple pathogens**,
with a concept aligned with **One Health***

- ❖ ReporTree can be easily implemented in [routine surveillance](#) and [outbreak investigation](#) workflows, providing solutions to:
 - ✓ Obtain genetic clusters at all possible thresholds using multiple clustering methods and maintaining cluster nomenclature between different time-points
 - ✓ Apply a dynamic approach to increase the resolution power in clusters of interest
 - ✓ Obtain summary reports with the statistics/trends for the derived genetic clusters

- ❖ Integration in INSaFLU-TELEVIR platform (ongoing)
- ❖ Integration in the COHESIVE platform

Collaboration with:



Acknowledgements

Genomics and Bioinformatics Unit

INSA

Alexandra Nunes

Daniel Sobral

Joana Isidro

João Santos

Luís Coelho

Miguel Pinto

Rita Ferreira

Supervisor: Vítor Borges

Team leader: João Paulo Gomes

GENPAT IZSAM

Adriano Di Pasquale

Nicolas Radomski



Infraestrutura
Nacional de
Computação
Distribuída



Instituto **Nacional de Saúde**
Doutor Ricardo Jorge

veronica.mixao@insa.min-saude.pt

FCT

Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



This work was supported by funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 773830: One Health European Joint Programme.