SENSIBLE: implementing data-driven early warning systems for future viral epidemics

Anna Bernasconi[1], Matteo Chiara[2], Tommaso Alfonsi[1], and Stefano Ceri[1]

[1] *Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*

[1] *Department of Biosciences, Università degli Studi di Milano, Milan, Italy*

CAiSE'24

https://sensible-prin.github.io/

PROJECT CONCEPT

Time span: December 2023 - November 2025.

In March 2020, COVID-19 was declared a global pandemic, spurring intense research efforts. Genomic surveillance emerged as a crucial defense against the virus.

The **SENSIBLE project**, supported by the Italian Ministry of University and Research (through NextGeneration EU funding), aims to **develop methods for analyzing viral genomes and implement early warning information systems based on data-driven analysis**, exploiting data from past epidemics for validation.

SENSIBLE will create an **integrated framework for genomic surveillance** of viral pathogens, utilizing data-driven and knowledge-based analyses. By leveraging knowledge from COVID-19, it seeks to enhance understanding of viral pathogens and aid healthcare decision-making.

OBJECTIVES

Data-driven Methods

Derive effective methods for data-driven identification of emerging viral pathogens

Surveillance

Build an objective framework for genomic surveillance in current and future epidemics

Warning

Implement an early warning system, to assist decision-making in healthcare

METHODOLOGIES

SENSIBLE will explore and harness data from different domains of interest, including: analyses of available data, mapping of equivalent/matched information from similar pathogens, computation or prediction of novel features and properties of the virus under study. The framework developed by SENSIBLE will feature four tasks:

Data-driven analysis of pathogens' evolution

The project partners will leverage previously developed methods and tools for the genomic surveillance. The range of applications of these methods will be extended to identify minimal subsets of actionable data, and evaluate their validity and robustness.

Data and knowledge-based analysis

To translate viral evolutionary dynamics into a collection of "epidemiologically-relevant" annotations of the viral genome, different sources of information will be gathered. Annotations available from existing resources will be retrieved and integrated into an internal knowledge base; missing data will be computed (via bioinformatics tools) or predicted (via automatic learning procedures).

Ranking and prioritization

Based on the evolutionary observations and the detailed functional annotations from previous tasks, a prioritization score will be computed to assign a "level of concern" to emerging viral pathogens and or to novel viral lineages. The score will be exploited to develop a ranking system, which will be evaluated according to the heuristics to be developed.

Validation and testing

The initial development and setup of SENSIBLE will be performed on a selection of use cases from the COVID-19 pandemic. Possible candidates to showcase the system and provide an unbiased evaluation are Monkeypox (2022), Zika (2015-2016), and Swine Influenza (2009).

EXPECTED RESULTS

We will tackle two key epidemiological questions and identify key metrics for raising alerts and early warnings in both scenarios:

Minimal actionable data

What is the minimal amount of data production/availability required to set up an effective surveillance system? Can genomic surveillance be applied even in scarce/low resources settings?

Prioritization of emerging pathogens

If a new mutation or pattern of mutations arise in a human pathogen, how does this impact its epidemiological features?

Insights for public health

How can a genomic surveillance tool based on small viral genomes dataset help inform decisions of public health institutions?

TEAM

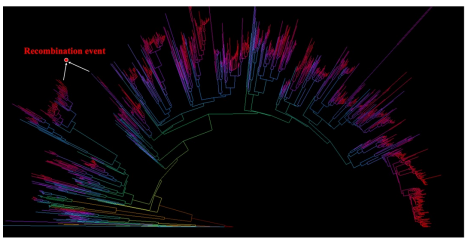
Dr. Anna Bernasconi, PI
Assistant Professor
(Politecnico di Milano)

Prof. Matteo Chiara, co-PI
Associate Professor
(Università degli Studi di Milano)

Prof. Stefano Ceri
Full Professor
(Politecnico di Milano)

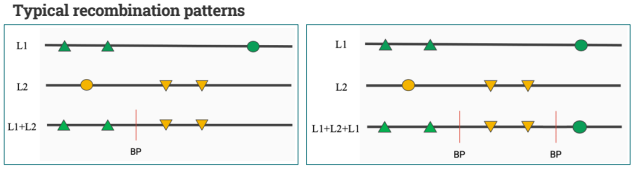
Dr. Tommaso Alfonsi
Postdoc Researcher
(Politecnico di Milano)

PRELIMINARY RESULTS

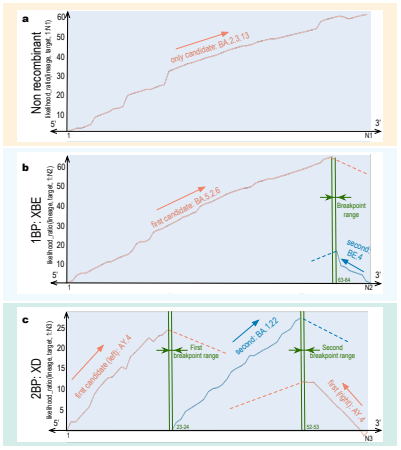


- 100 recombination events recognized in the SARS-CoV-2 virus evolution history
- Lineage XBB descendants are considered Variants Of Interest (WHO), leading to
  - reduced neutralization by antibodies generated against previous infection or vaccination;
  - reduced efficacy of treatments;
  - predicted increase in transmissibility or disease severity.

Typical recombination patterns



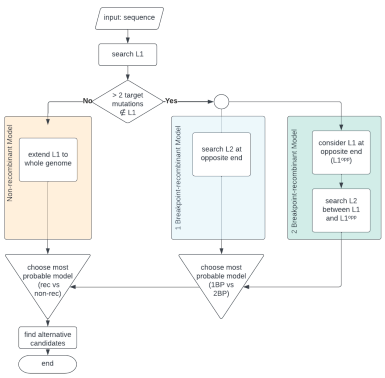
Intuition



- **Data-driven**. The input knowledge is inferred from data: mutation probabilities that can be computed in few minutes
- **Independent from** the **classification** system in use (Nextstrain clades, PANGO lineages, GISAID clades, ...)
- **Independent from phylogenetic** analysis
- **Applicable to any monopartite virus** with a sound classification system.
- Applicable to **pandemic-scale scenarios**
- Higher sentivity w.r.t. sota methods and manual approach
- Potentially impactful for genomic surveillance practices

Method

- Up to 2 breakpoints are detected in the current version of RecombinHunt
- Only the mutations are considered for determining the breakpoint location. This implies a certain degree of uncertainty in the position of the breakpoint.



The work is supported by Ministero dell'Università e della Ricerca (PRIN PNRR 2022 "SENSIBLE" project, n. P2022CNN2J), funded by the European Union, Next Generation EU, within PNRR M4.C2.1.1. Politecnico di Milano CUPD53D23017400001; Università di Milano CUP G53D23006690001. PI Anna Bernasconi, co-PI Matteo Chiara.