

# Keynote Speech: Storing and analyzing viral sequences through data-driven Genomic Computing – Abstract

Stefano Ceri<sup>1</sup>

<sup>1</sup>Politecnico di Milano, P. Leonardo Da Vinci 32, 20133 Milano, Italy

## Abstract

The first part of the talk illustrates, in simple and data-inspired terms, what is a viral sequence, what are mutations, how mutated sequences become organized forming a “variant”, what are the effects of individual mutations and of variants, how viral sequences are deposited to public repositories (GenBank, COGUK, GISAID). The second part of the talk presents the systems that were developed within my group, thanks to ERC and EIT funding. Specifically, I will illustrate (i) *VirusSurf*, a search system enabling free meta-data driven search over the integrated and curated databases, now hitting about 3 million SARS-CoV-2 sequences, continuously updated from the above repositories; (ii) *VirusViz*, a data visualization tool for comparatively analyzing query results; (iii) *VirusLab*, a tool for exploring user-provided viral sequences; (iv) *EpiSurf*, a tool for intersecting viral sequences with epitopes - used in vaccine design. I will also hint at ongoing projects for viral surveillance and for exploring a knowledge base of viral resources.

## Keywords

viral sequences, SARS-CoV-2 viral genome, genomic computing

---

SEBD 2021: The 29th Italian Symposium on Advanced Database Systems, September 5-9, 2021, Pizzo Calabro (VV), Italy

✉ stefano.ceri@polimi.it (S. Ceri)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)