INSaFLU-TELE-Vir: an open web-based bioinformatics suite for influenza and SARS-CoV-2 genome-based surveillance

Miguel Pinheiro‡, Ricardo J Pais§, Joana Isidro§, Miguel Pinto§, Carlijn Bogaardt|, Joaquin M Prada|, Daniel L Horton|, João P Gomes§, Vítor Borges§

‡ Institute of Biomedicine-iBiMED, Aveiro, Portugal
§ National Institute of Health Dr. Ricardo Jorge (INSA), Lisbon, Portugal
| University of Surrey, Surrey, United Kingdom

Corresponding author: Vítor Borges (vitor.borges@insa.min-saude.pt)

Abstract

A new era of virus surveillance is emerging based on the real-time monitoring of virus evolution at whole-genome scale (World Health Organization 2021). Although national and international health authorities have strongly recommended this technological transition, especially for influenza and SARS-CoV-2 (World Health Organization 2021, Revez et al. 2017), the implementation of genomic surveillance can be particularly challenging due to the lack of bioinformatics infrastructures and/or expertise to process and interpret next-generation sequencing (NGS) data (Oakeson et al. 2017).

We developed and implemented INSaFLU-TELE-Vir platform (https://insaflu.insa.pt/) (Borges et al. 2018), which is an influenza- and SARS-CoV-2-oriented bioinformatics free web-based suite that handles primary NGS data (reads) towards the automatic generation of the main “genetic requests" for effective and timely laboratory surveillance. By handling NGS data collected from any amplicon-based schema (making it applicable for other pathogens), INSaFLU-TELE-Vir enables any laboratory to perform multi-step and intensive bioinformatics analyses in a user-oriented manner without requiring advanced training.

INSaFLU-TELE-Vir handles NGS data collected from distinct sequencing technologies (Illumina, Ion Torrent and Oxford Nanopore Technologies), with the possibility of constructing comparative analyses using different technologies. It gives access to user-restricted sample databases and project management, being a transparent and flexible tool specifically designed to automatically update project outputs as more samples are uploaded. Data integration is thus cumulative and scalable, fitting the need for both routine surveillance and outbreak investigation activities.

The bioinformatics pipeline consists of six core steps:

1. read quality analysis and improvement,

©
.

2. human betacoronaviruses (including SARS-CoV-2 Pango lineages) and influenza type/subtype classification,
3. mutation detection and consensus generation,
4. coverage analysis,
5. alignment/phylogeny,
6. intra-host minor variant detection (and automatic detection of putative mixed infections).

The multiple outputs are provided in nomenclature-stable and standardized formats that can be visualized and explored *in situ* or through multiple compatible downstream applications for fine-tuned data analysis.

Novel features are being implemented into the INSaFLU-TELE-Vir bioinformatics toolkit as part of the OHEJP TELE-Vir (https://onehealthejp.eu/jrp-tele-vir/) project, including rapid detection of selected genotype-phenotype associations, and enhanced geotemporal data visualization.

All the code is available in github (https://github.com/INSaFLU) with the possibility of a local docker installation (https://github.com/INSaFLU/docker). A detailed documentation and tutorial is also available (https://insaflu.readthedocs.io/en/latest/).

In summary, INSaFLU supplies public health laboratories and researchers with an open and user-friendly framework, potentiating a strengthened and timely multi-country genome-based virus surveillance.

Keywords

Genomic surveillance; SARS-CoV-2; influenza; sequencing; bioinformatics; public health.

Presenting author

Vítor Borges

Presented at

One Health EJP Annual Scientific Meeting Satellite Workshop 2021 Software Fair

Funding program

One Health EJP has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 773830

Grant title

Software development is being co-funded by the European Commission on behalf of OneHealth EJP TELE-Vir project (https://onehealthejp.eu/jrp-tele-vir/).

Hosting institution

National Institute of Health Dr. Ricardo Jorge (INSA), Lisbon, Portugal.

Conflicts of interest

None.

References

- Borges V, Pinheiro M, Pechirra P, et al. (2018) INSaFLU: an automated open web-based bioinformatics suite from-reads for influenza whole-genome-sequencing-based surveillance. Genome Med 10 (46). https://doi.org/10.1186/s13073-018-0555-0
- Oakeson KF, Wagner J, Mendenhall M, Rohrwasser A, Atkinson-Dunn R (2017) Bioinformatic Analyses of Whole-Genome Sequence Data in a Public Health Laboratory. Emerg Infect Dis 23 (9): 1441-1445. https://doi.org/10.3201/eid2309.170416
- Revez J, Espinosa L, Albiger B, Leitmeyer KC, Struelens MJ (2017) Survey on the Use of Whole-Genome Sequencing for Infectious Diseases Surveillance: Rapid Expansion of European National Capacities, 2015-2016. Frontiers in public health 5: 347. https://doi.org/10.3389/fpubh.2017.00347
- World Health Organization (2021) Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. https://www.who.int/publications/i/item/9789240018440. Accessed on: 2021-4-16.