

Genomics to fight the emerging COVID19 pandemic

G Damaggio¹, F Villani¹, V Colonna¹

¹IGB-CNR, Napoli, Italy

COVID-19 pandemic is caused by the expansion of a new viral strain, SARS-CoV-2, determining an Acute Respiratory Distress Syndrome that in many cases could be fatal. Among the many aspects worth working on, we chose to be active in two ways: contributing to understand the genetic variability of the SARS-CoV-2 genome and to inform the general population within a range of initiatives.

Understanding how the genome of SARS-CoV-2 is composed and changes through time and space is of paramount importance. It is relevant for basic research to reconstruct the pattern of diffusion and trace the evolutionary origin of the virus. Similarly it is important for the applied research, to understand the susceptibility to the disease, improve diagnosis through genetic testing, cure, and the development of effective vaccines. Beside impacting the scientific aspects, mapping of the SARS-CoV-2 genome has social implications in the informed management of the pandemic. It is in fact possible to understand possible changes in transmissibility as well as to track the dynamics of the outbreaks in order to evaluate the effectiveness of the containment strategies.

In practice we acted through collaborative efforts. Together with the IBB-CNR and within the framework of the project RECOVER-COVID funded by Regione Campania, we evaluated the conservation of the SARS-CoV-2 mutations in time, using publicly available genome sequences. We discovered that more recent sequences are enriched for non-conservative mutations compared to early sequences, suggesting a progressive adaptation of the virus to the host.

We also worked on third-generation sequencing data of SARS-CoV-2. In collaboration with the CEINGE we assembled and analyzed long reads of genomic sequences of SARS-CoV-2 produced using the Oxford Nanopore Technology. Samples were collected in the Regione Campania, bridging a gap in the geographical mapping of the virus genome. Although with a still small sample size, this collection represents the most comprehensive from Regione Campania. Finally, we took part in the COVID biohackathon initiative² (covid-19-bh20-assembly) to work on a project on de-novo assembly of polyA-enriched third-generation sequences of SARS-CoV-2 to determine the suitability of this type of data for assembly.

We took part in a number of initiatives to explain and diffuse the scientific results of the research on COVID19. Effective communication was especially required during the first stages of the pandemic when the response of the general population to the emergency could condition the strategic decision of the policy making organs. Among the initiative we were invited to, we want to highlight a report to the Senate of the Republic to inform the Health Committee about the importance of promoting projects to sequence the SARS-CoV-2 genome, and the participation in the Science Festival Futuro Remoto where we explained to the general public two research papers of great relevance^{3,4}.

Future perspectives. We will collect, sequence and openly publish as many samples of SARS-CoV-2 genomes as we can. This will enable basic and applied research on COVID19. We will also continue the scientific communication campaign as we believe that this pandemic requires full and continuous promotion of citizen science.

References

1. Balasco N, **Damaggio G**, Esposito L, **Villani F**, Berisio R, **Colonna V**, Vitagliano L.

The analysis of conservative and non-conservative amino acid mutations in SARS-CoV-2 variants provides insights into the early evolution of the virus in humans (Submitted)

2. <https://github.com/virtual-biohackathons/covid-19-bh20/wiki/Pangenome>

3. Argimón, Silvia, et al. "Microreact: visualizing and sharing data for genomic epidemiology and phylogeography." *Microbial genomics* 2.11 (2016)

4. Plante, Jessica A., et al. "Spike mutation D614G alters SARS-CoV-2 fitness." *Nature* (2020): 1-9.

Keywords: SARS-CoV-2, genomics, third generation sequencing

Contacts: vincenza.colonna@igb.cnr.it, gianluca.damaggio@igb.cnr.it, flavia.villani@igb.cnr.it.

Website(s): https://colonnalab.github.io/laboratory_WebPage/

<https://www.lemonde.fr/blog/binaire/2020/05/06/sars-cov-2-et-covid-19-on-va-jouer-sur-les-mots/>

Other: