COMMENTARY
SARS-CoV-2 variants: what have we learnt so far?

Luigina Ambrosio[1*], **Matteo Chiara**[2,3*], **Alessandra Lo Presti**[1*], **Piero Poletti**[4*], **Claudia Alteri**[5,6], **Davide Cacchiarelli**[7,8,9], **Valentina Bouchè**[7], **Michele Morgante**[10,11], **Andrea Ballabio**[7,9,12,13], **Carlo Federico Perno**[6], **Stefano Merler**[4], **Graziano Pesole**[3,14] and **Paola Stefanelli**[1*]

[1]*Dipartimento Malattie Infettive, Istituto Superiore di Sanità, Rome, Italy*
[2]*Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy*
[3]*Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari, Consiglio Nazionale delle Ricerche, Bari, Italy*
[4]*Centro Health Emergencies, Fondazione Bruno Kessler, Trento, Italy*
[5]*Dipartimento di Oncologia ed Emato-Oncologia, Università degli Studi di Milano, Milan, Italy*
[6]*Area di Ricerca di Medicina Multimodale, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy*
[7]*Istituto Telethon di Genetica e Medicina (TIGEM), Pozzuoli, Naples, Italy*
[8]*Scuola Superiore Meridionale (SSM), Università degli Studi di Napoli "Federico II", Naples, Italy*
[9]*Dipartimento di Scienze Mediche Traslazionali, Università degli Studi di Napoli "Federico II", Naples, Italy*
[10]*Dipartimento di Scienze Agroalimentari, Ambientali e Animali, Università degli Studi di Udine, Udine, Italy*
[11]*Istituto di Genomica Applicata, Udine, Italy*
[12]*Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA*
[13]*Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, USA*
[14]*Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, Università degli Studi di Bari "Aldo Moro", Bari, Italy*
These Authors contributed equally

Abstract
Besides the timely detection of different SARS-CoV-2 variants through surveillance systems, functional and modelling studies are essential to better inform public health response and preparedness. Here, the knowledge available so far on SARS-CoV-2 variants is discussed from different perspectives, in order to highlight the relevance of a multidisciplinary approach in countering the threat posed by this insidious virus.

The remarkable capacity of viruses to adapt rapidly to new hosts and environments is highly dependent on their ability to generate genomic diversity in a short period of time. RNA viruses introduce and select mutations in their genome faster than DNA ones, thus evolving rapidly [1]. This high evolutionary rate causes accumulation of mutations over time. While most of these mutations are expected to be either deleterious and get rapidly eliminated, or relatively neutral with no detectable effects, some of them confer a selective advantage to the viruses [2]. The potential epidemiological consequences of novel mutations are closely related to their impact on viral replication and transmission and on the competition between co-circulating strains.

In the case of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), modern high throughput sequencing technologies have been applied to obtain millions of viral genome sequences in near real time, providing the unprecedented opportunity to identify and track most of the novel mutations as they accumulated in the genome. This allowed the development of effective "genomic surveillance" systems for monitoring viral variants associated with changes in transmissibility, disease severity and immune evasion.

At the time of writing, more than 233 thousand mutations have been independently identified for SARS-CoV-2 genome [3]. However, while surveillance systems succeeded in compiling an extensive catalogue

COMMENTARY

of viral diversity, our understanding of all the possible functional implications is ongoing.

Most efforts to identify and characterize epidemiologically relevant mutations, in particular those potentially associated with immune escape, have been focused on the spike protein (S) gene and specifically its receptor-binding domain (RDB), defining the portion that mediates the recognition of ACE2 receptor on human host cells and representing the main target of neutralizing antibodies [4]. However, considerable efforts have been made also to investigate the effects of different mutations on viral proteins targeted by antiviral therapies, e.g., the catalytic subunit of the RNA-dependent RNA polymerase (RdRp), non-structural protein 12 (nsp12) [5], or by molecular diagnostic assays based on real-time reverse-transcriptase polymerase-chain-reaction (rRT-PCR), e.g. nucleocapsid protein (N) [6].

Although the role of several mutations in conferring a selective advantage to SARS-CoV-2 has been established [7], some important limitations persist. First of all, while genetic/phylogenetic approaches have been successfully applied for the identification of genomic sites and mutations potentially under positive selection, these do not provide direct mechanistic insights and the inferred estimations may be confounded by genetic drift, founder effect and sampling bias [8]. Moreover, decipher the impact of novel mutations, or identify potential epistatic interactions (*i.e.*, the combined effect of two or more mutations), through functional studies typically requires a time frame not always compatible with the rapid spread of successful variants [9].

Significant advances in viral population genetics and *in silico* protein structure prediction will be required in the future to develop novel and more accurate tools for predicting the fate of viral mutations and variants. The application of population genetics theory and models is not straightforward for SARS-CoV-2 dynamics, due to the inherent complexity of the viral population structure, transmission mechanisms and a continuous alteration of viral population size. Methods based on artificial intelligence have been recently developed [10], but the extent to which such methods can be applied for the prediction of the structural and functional effects of single or multiple amino acid substitutions, also in relation to protein-protein interactions, is not completely clear.

With the aim to facilitate the monitoring of the virus evolution, and the tracking of variants of potential epidemiological relevance, different nomenclature and classification systems have been proposed by the scientific community to classify clusters of SARS-CoV-2 strains with common sets of defining mutations. Among these systems, the most common ones are:

1. Nextstrain, which labels clades that persist for several months and have significant geographic spread with an *ad hoc* number-letter combination (*i.e.*, the estimated year of appearance, followed by a letter). As of January 2022, it includes 24 clades (the complete list is available at https://nextstrain.org/ncov/gisaid/global [11]).

2. Global Initiative on Sharing All Influenza Data (GISAID), based on marker mutations within high-level phylogenetic groupings corresponding to clades, from the early split of S and L to the further evolution of L into V and G, and later of G into GH, GR and GV, and more recently GR into GRY [12]. In late 2020, a new clade split from base clade G forming clade GK. The recent emergence of the Omicron variant (November 2021) caused the introduction of the GRA clade in GISAID.

3. Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin, introduced in April 2020) [13], which combines considerations based on evolutionary history, geographic spread and overall prevalence to define coherent groups of genomic sequences or "lineages" (the full list of lineages is available at cov-lineages.org, [13]). As of January 2022, the Pango nomenclature includes more than 1550 distinct lineages.

While the Nextstrain and GISAID nomenclature system provide a large-scale overview of the clade trends, Pangolin captures a more fine-grained representation of local/regional outbreaks/clusters of SARS-CoV-2 [12].

In addition to these classification systems, the World Health Organization (WHO) has introduced a simplified nomenclature based on the Greek Alphabet (*i.e.*, Alpha, Beta, Gamma, Delta, Omicron) that appears to be easier and more practical also for experts from different disciplines [14].

International health authorities have established a series of guidelines for the identification of SARS-CoV-2 variants associated with relevant changes in epidemiological features. According to the European Centre for Disease Prevention and Control (ECDC) [15], SARS-CoV-2 variants are classified as Variants of Concern (VOCs), Variants of Interest (VOIs) or Variants under Monitoring (VUMs), based mainly on their possible association with increased transmissibility, more severe disease and/or reduced serum neutralization. Variants can also be de-escalated from the status of VOC/VOI/VUM when at least one of the following criteria is satisfied: (i.) the variant is no longer circulating; (ii.) the variant has been circulating for a long time without any impact on the overall epidemiological situation; (iii.) scientific evidence demonstrates that the variant is not associated with concerning traits [15].

Different approaches have been applied to quantify the transmissibility of VOCs, and in particular of the Alpha and Delta variants, from household studies analysing secondary cases generated by different variants, to mechanistic transmission models and statistical inference methods applied to population prevalence data on circulating lineages [7, 16]. These studies suggested that Alfa was 45-66% more transmissible than previously dominant variants [16] and that the increased transmissibility of Delta with respect to Alfa may be around 50-60% (*i.e.*, 76-120% more transmissible than lineages circulating in the 2020) [7, 16]. Estimates for the Gamma variant suggest that the increased transmissibility of this variant compared

to historical lineages may range from 3% to 56% [7], depending on different factors. However, the analysis of the emergence of the Delta variant in India, where the observed immunity was, at the time of the Delta appearance, mainly ascribed to natural infection, suggested that this variant may have been associated with a reduced sensitivity to immune responses generated against previous variants.

Regarding Omicron variant, a modelling study highlighted that its emergence in South Africa quickly shifted the SARS-CoV-2 reproduction number from below one to values around 1.7 in about two months since the seeding [17]. In early December 2021, the daily growth rate of Omicron cases in Denmark was associated with a doubling time between 1.8 and 3.2 days [18]. At the same time, a report from United Kingdom described a 20-45% reduction of the risk of hospitalization for Omicron relative to Delta infections [19].

Overall, the replacement of previous lineages by a new emerging variant (*i.e.*, from near 0% to over 50%) may occur in a short period of time and the epidemiological consequences of the progressive spread of a new VOC could markedly differ across distinct geographical areas due to heterogeneous co-circulation of different lineages and/or vaccine uptake [20-22].

It should be also noted that lineages characterized by a shorter generation time (i.e., the interval between infection of primary and secondary cases) or an increased transmission occurring before symptom onset could hasten its spread and drastically reduce the effectiveness of contact tracing operations in interrupting SARS-CoV-2 transmission chains.

The timely detection and interpretation of trends regarding the circulation of SARS-CoV-2 variants are key to trigger the public health response to COVID-19.

Conflict of interest statement

No potential conflict of interest was reported by the authors.

Andrea Ballabio is Co-Founder of Casma Therapeutics and Advisory Board Member of Avilar Therapeutics and Next Generation Diagnostic srl.

Davide Cacchiarelli is Co-Founder, Shareholder and Consultant of Next Generation Diagnostic srl.

REFERENCES

1. Sanjuán R. From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. PLoS Pathog. 2012;8(5):e1002685. doi: 10.1371/journal.ppat.1002685
2. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. Annu Rev Microbiol. 1997;51:151-78. doi: 10.1146/annurev.micro.51.1.151
3. Canakoglu A, Pinoli P, Bernasconi A, et al. ViruSurf: an integrated database to investigate viral sequences. Nucleic Acids Res. 2021;49(D1):D817-D824. doi: 10.1093/nar/gkaa846
4. Huang Y, Yang C, Xu XF, et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. 2020;41(9):1141-9. doi: 10.1038/s41401-020-0485-4
5. Kozlov M. Why scientists are racing to develop more COVID antivirals. Nature. 2022;601(7894):496. doi: 10.1038/d41586-022-00112-8
6. Jaroszewski L, Iyer M, Alisoltani A, et al. The interplay of SARS-CoV-2 evolution and constraints imposed by the structure and functionality of its proteins. PLoS Comput Biol. 2021;17:e1009147. doi: 10.1371/journal.pcbi.1009147
7. Campbell F, Archer B, Laurenson-Schafer H, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill. 2021;26(24):2100509. doi: 10.2807/1560-7917.ES.2021.26.24.2100509
8. Lynch M, Ackerman MS, Gout JF, et al. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 2016;17(11):704-14. doi:10.1038/nrg.2016.104
9. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021;19:141-54. doi: 10.1038/s41579-020-00459-7
10. Kuhlman B, Bradley P. Advances in protein structure prediction and design. Nat Rev Mol Cell Biol. 2019;20(11):681-97. doi: 10.1038/s41580-019-0163-x
11. Nextstrain: Genomic epidemiology of novel coronavirus – Global subsampling. [Accessed: 24 Jan 2022]. Available from: https://nextstrain.org/ncov/gisaid/global.
12. Global Initiative on Sharing All Influenza Data (GISAID). Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses. Available from: www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/.
13. Cov-lineages.org. PANGO lineages. 2020. Available from: https://cov-lineages.org/.
14. World Health Organization (WHO). Tracking SARS-CoV-2 variants. Available from: www.who.int/en/activities/tracking-SARS-CoV-2-variants/.
15. European Centre for Disease Prevention and Control (ECDC). SARS-CoV-2 variants of concern. Available from: www.ecdc.europa.eu/en/covid-19/variants-concern.
16. Lyngse FP, Mølbak K, Skov RL, et al. Increased transmissibility of SARS-CoV-2 lineage B.1.1.7 by age and viral load. Nat Commun. 2021;12(1):7251. doi: 10.1038/s41467-021-27202-x
17. Gozzi N, Chinazzi M, Davis JT, et al. Preliminary modeling estimates of the relative transmissibility and immune escape of the Omicron SARS-CoV-2 variant of concern in South Africa. medRxiv. 2022. doi: https://doi.org/10.1101/2022.01.04.22268721
18. Staten Serum Institut. Scenarios for infections and new admissions caused by the Omicron variant. Report by the Expert Group for Mathematic Modelling, 17 December 2021. Available from: https://en.ssi.dk/-/media/cdn/files/scenarios-for-infections-and-new-admissions-caused-by-the-omicron-variant_201221.pdf?la=en.
19. Ferguson N, Ghani A, Hinsley W, Erik Volz on behalf of the Imperial College COVID-19 response team. Report 50. Hospitalisation risk for Omicron cases in England. WHO Collaborating Centre for Infectious Disease Modelling, MRC Centre for Global Infectious Disease

Analysis, Jameel Institute, Imperial College London. 22-12-2021. Available from: www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-50-severity-omicron/.

20. CoVariants. Overview of Variants in Countries. Available from: https://covariants.org/per-country

21. Stefanelli P, Trentini F, Guzzetta G, et al. Co-circulation of SARS-CoV-2 Alpha and Gamma variants in Italy, February and March 2021. Euro Surveill. 2022;27(5):2100429. doi: 10.2807/1560-7917.ES.2022.27.5.2100429

22. Dhawan M, Priyanka, Sahni A, Choudhary OP. Vaccine inequity and hesitancy: Dual factors in the emergence of novel SARS-CoV-2 variants. Ann Med Surg (Lond). 2022;73:103186. doi: 10.1016/j.amsu.2021.103186

COMMENTARY