New Microbiologica, 46, 1, 1-8, 2023, ISSN 1121-7138

REVIEW

Genomic epidemiology and phylogenetics applied to the study of SARS-CoV-2 pandemic

Alessia Lai[1,2], Annalisa Bergna[1,2], Carla Della Ventura[1,2], and Gianguglielmo Zehender[1,2]

[1]*Department of Biomedical and Clinical Sciences, University of Milan, Italy*
[2]*CRC-Coordinated Research Center "EpiSoMI", University of Milan, Italy*

SUMMARY

The study of characteristics, prevalence and patterns of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections is significant to monitor and define the status of the pandemic, helping to design and evaluate control strategies. In this setting, the continuous emergence of new variants and their dynamic of replacement underline the importance of implementing genomic epidemiology and phylogenetic methods for the molecular monitoring and surveillance of this new virus. The current profile of the pandemic can change rapidly when new variants emerge and spread, impacting epidemiology and public health in terms of prevention and treatment and making it necessary to develop new molecules and formulate vaccines.

In this paper, we reviewed and synthesized the main studies on molecular genomics and phylogeny of SARS-CoV-2 during the pandemic, and highlighted their contributions to our understanding of this new emergent pathogen.

Received January 16, 2023 *Accepted* January 26, 2023

INTRODUCTION

From the end of 2019 to the present, the COVID-19 (coronavirus disease 2019) pandemic has represented a unique phenomenon: the real-time evolutionary process of a zoonotic virus progressively adapting to its new human host after a spillover event.

The adaptation process of SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) started with a series of mutations causing increased efficiency to transmission to human host (Korber *et al.*, 2020; Zehender *et al.*, 2020) and leading to a progressive increase in genetic diversity that required constant monitoring (Srivastava *et al.*, 2021). While this diversity remained low in the first months of the pandemic (Balloux *et al.*, 2022), it increased when variants of concern (VOCs) emerged (Boehm *et al.*, 2021).

According to WHO classification, these variants are characterized by an increased transmissibility, or changes in COVID-19 epidemiology, or increased virulence or changes in clinical disease presentation, or decrease in effectiveness of public health and social measures or available diagnostics, vaccines and therapeutics. Variants classification also includes VOIs (variants of interest) and VUMs (variants under monitoring); the former characterized by genetic changes that are predicted or known to affect virus characteristics and identified to cause significant community transmission or multiple COVID-19 clusters in multiple countries, the latter presenting genetic changes that are suspected of affecting virus characteristics, with indication of future risk (https://www.who.int/activities/tracking-SARS-CoV-2-variants).

Like other RNA viruses, SARS-CoV-2 has high genetic variability, partially mitigated by the presence of an exonuclease coded by the nsp14 gene, associated with the viral RNA-dependent RNA-polymerase (RdRP), able to remove the non-complementary nucleotides wrongly incorporated in the nascent chain during RNA duplication (proofreading activity). Nevertheless, the virus genome, having an average length of 29,000 nucleotides, is subject to an evolutionary rate of about 0.4-2 mutations every 1,000 sites per year, corresponding to an accumulation of about 2 amino acid changes per month (Kistler *et al.*, 2022).

An additional mechanism of SARS-CoV-2 diversity is represented by recombination, a common evolutionary process for RNA viruses, particularly for coronaviruses (Graham, Baric, 2010; Focosi, Maggi, 2022). Recently, recombination has been estimated to occur in less than 10% of sequenced genomes (Turakhia *et al.*, 2022; VanInsberghe *et al.* 2021).

Key words:
Phylogeny, genomic surveillance, viral evolution, SARS-CoV-2.

Corresponding author:
Alessia Lai
E-mail: alessia.lai@unimi.it

Initially, SARS-CoV-2 included only two closely similar lineages: A and B (Rambaut *et al.*, 2021) separated early in the Wuhan outbreak, with lineage B now being more widely distributed. This observation raised the hypothesis of at least 2 distinct spillover events from animals to humans (Pekar *et al.*, 2022). Subsequently, these lineages began to evolve independently, accumulating a series of mutations (particularly lineage B) that gave rise to a series of lineages/sub-lineages and clades/sub-clades that form the basis of today's very complex classification.

During this period, viral evolution was predominantly characterized by a slow mutational pattern dominated by neutral evolution (MacLean *et al.*, 2021) except for two mutations (D614G and P323L, in the Spike protein and in the viral RNA polymerase RNA-dependent, respectively).

In April 2020, the first variant of great epidemiological impact, classified as B.1 lineage and characterized by the D614G mutation, became established in Europe and then all over the world. A few weeks after, it became the predominant lineage worldwide and the progenitor of all subsequent variants (Korber *et al.*, 2020).

In late 2020, the first VOCs were described and assigned to a Greek letter according to the classification of the WHO, starting with the Alpha variant (lineage B.1.1.7, clade 20I), which was firstly identified in South-East Britain and rapidly spread worldwide during 2021, to account for half of the global viral isolates in April. This VOC was characterized by 7 amino acid substitutions and 2 deletions in the Spike protein.

At the same time, two other VOCs emerged. The first, named Beta (B.1.351/20H), first detected in South Africa, encompassed nine mutations, 8 amino acid substitutions and one deletion, in the Spike protein. At the end of June 2021 this variant was associated with half of all cases in Sub-Saharan Africa and caused the second wave of COVID-19 in South Africa.

The second, named Gamma (P.1/20B), was identified during an epidemic resurgence of cases in Manaus, Brazil, and in travelers from Brazil. This variant was characterized by as many as 11 amino acid substitutions in the Spike protein and was widespread mainly in South America, accounting for up to 80% of infections in Brazil? that country in the late spring-summer of 2021.

However, by late summer 2021 the incidence of the major VOCs was declining, and the Alpha variant was quickly replaced globally by the emerging VOC Delta (B.1.617.2). This VOC, identified in India, was characterized by a unique combination of mutations in the Spike protein, which conferred increased transmissibility as well as some ability to escape neutralization. The Delta variant spread rapidly, becoming dominant worldwide (more than 85% of global prevalence in November 2021) (Boehm *et al.*, 2021), and producing a number of sublineages (more than 40) classified as AYs (Dhawan *et al.*, 2022b).

All these early variants were characterized by increased transmissibility compared to pre-existing variants, evidenced by an increase in effective reproductive numbers (by 25-40% for Alpha, Beta and Gamma and up to 97% for Delta) and secondary attack rate. They also showed some resistance to monoclonal antibodies (mABs), and reduced susceptibility (albeit not significantly) to vaccines (Tao *et al.*, 2021).

In late 2021, the Omicron variant (B.1.1.529/21K) emerged in South Africa and spread rapidly, replacing the earlier variant. This new lineage was characterized by a surprisingly large number of substitutions in the Spike protein, able to cause substantial modification of the protein conformation and of the epitopes recognized by the neutralizing antibodies.

This, together with an increased ACE2 (Angiotensin-Converting Enzyme 2) binding affinity reflecting on higher transmissibility than the previous variants, drove the worldwide spread of this variant and of its numerous sub-lineages (BA.1, BA.2, BA.3, BA.5 and BA.4), which at present represent almost the totality of infections on a global level. Today, the continuous evolution of Omicron sublineages has produced a "swarm" of variants, of which the most prevalent are only a fraction of those circulating in the population, and it is not clear what, if any, variant will succeed in prevailing over the others in different epidemiological and geographical settings (Dhawan *et al.*, 2022a). Despite a reduced virulence suggested for this variant (Bálint *et al.*, 2022), it spread to a predominantly immunized global population thanks to past infections and extensive vaccination. However, the spread of Omicron in the mostly non-vaccinated population of Hong Kong in March 2022 caused the highest number of deaths from COVID-19 in older people ever recorded in that geographic area.
(https://www.info.gov.hk/gia/general/202203/14/P2022031400687.htm).

EVOLUTIONARY DYNAMICS AND MUTATION LANDSCAPE OF SARS-CoV-2 VOCS

Variants of concern dynamic

The major issue regarding VOCs is their dynamic of "replacement" (*Figure 1*), i.e., their potential to outcompete and rapidly replace formerly prevalent lineages that become extinct, first in the areas where they likely emerged and subsequently spread to many countries, similarly to that observed for influenza A virus (Chavda *et al.*, 2022; Funk *et al.*, 2021; Galloway *et al.*, 2021). The main difference between SARS-CoV-2 and the influenza virus is that the VOCs did not derive from each other, but have different an-

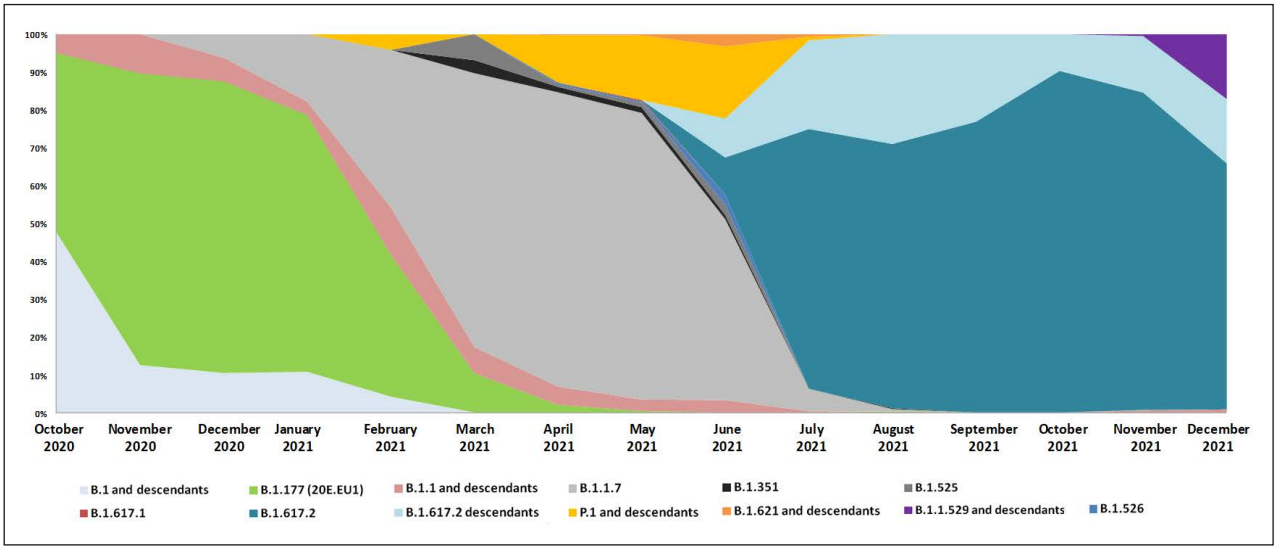

Figure 1 - *Dynamics of the SARS-CoV-2 epidemic in Italy: variant prevalence in terms of lineages (Modified from Lai A, Bergna A, Menzo S, et al. Virol J. 2021 and Lai A, Bergna A, Della Ventura C, et al. Viruses 2022). B.1.1.7: Alpha variant, B.1.351: Beta variant, B.1.525: Eta variant, B.1.617.1: Kappa variant, B.1.617.2: Delta variant, P.1: Gamma variant, B.1.621: Mu variant, B.1.1.529: Omicron variant, B.1.526: Iota variant.*

cestors in early lineages that were replaced by them. This dynamic of replacement of variants underlines the importance of epidemiological genomics and phylogenetic methods for the molecular monitoring and surveillance of SARS-CoV-2. The emergence and spread of variants can rapidly change the current profile of the pandemic in terms of epidemiological and clinical impact, deaths, need for new active drugs and formulation of vaccines.

In the evolution of the SARS-CoV-2 genome, one of the most evident phenomena is the high number of sites characterized by convergent evolution, i.e., mutations that arise independently in different evolutionary lineages (homoplasy). These mutations tend to recur several times, as they are capable of conferring significant advantages in terms of fitness (Daniloski *et al.*, 2021; Gobeil *et al.*, 2021; Wang *et al.* 2021; Hoffmann *et al.*, 2021). The fitness effect is influenced by the genetic background of the viral lineage, co-circulating lineages, host immunity, and some epidemiological factors such as different mitigation measures.

Another phenomenon is the combined action of different mutations (epistasis) mainly due to the compensatory mechanism that reduces the cost in terms of fitness of a particularly advantageous mutation (such as those conferring immune escape). For example, the N501Y mutation, which arose independently in several VOCs (Alpha, Beta and Gamma), increases ACE2-binding affinity and compensates for mutations, decreasing Ab neutralization such as K417N/T (Rochman Nash *et al.*, 2022; Moulana *et al.*, 2022). This may explain the frequent re-emergence of some mutations often present in combination with each other (linkage disequilibrium) (McLeod, Gandon, 2022).

Evolution of mutations

Lineages can grow in frequency due to stochastic effects. It is expected that mutations associated with successful clades will change over time, reflecting both a changing fitness landscape and the stochastic nature of evolution. Mutations that transcend this or are associated with successful lineages at multiple time points have important adaptive functions.

In fact, a single mutation can increase or decrease one function, and simultaneously increase or decrease the second one. In particular, the Spike protein is used by the virus to enter human cells via the ACE2 receptor (in particular the Receptor Binding Domain-RBD), but it also represents the main target of neutralizing antibodies, being an immunodominant antigen. For example, the D614G mutation in the RBD region, is a non-synonymous substitution replacing the amino acid D (aspartic acid) with a less bulky amino acid G (glycine). This contributes to a more flexible hinge region that enables more efficient cutting for receptor binding (Korber *et al.*, 2020; Turoňová *et al.*, 2020), giving the virus a selective advantage in infection and transmission. However, this mutation also seems to increase susceptibility to neutralizing antibodies (Weissman *et al.*, 2021).

On the contrary, the mutation N439K has been reported associated with both increased affinity for ACE2 and reduction of neutralization by some monoclonal antibodies and convalescent plasma (Thomson *et al.*, 2021; Zhou *et al.*, 2021). Other mutations, such as K417N/T, present in VOCs Beta, Gamma and more

recently Omicron, were associated with immune escape from neutralizing antibodies, but also with a decrease in ACE2 binding-affinity (Starr *et al.*, 2020). Globally, mutations increasing ACE2 binding affinity appeared to be prevalent in the first months of the pandemic, while those conferring immune escape were less numerous and appeared associated with variants circulating in populations characterized by a large spread of infection in previous waves (Krause *et al.*, 2021).

Role of specific mutations

Since November 2020, a substantial increase in the number of SARS-CoV-2 sites undergoing accelerated evolution was observed. This change was first driven by the emergence of the 501Y lineages (B.1.1.7/501Y.V1, B.1.351/501Y.V2, P.1/501Y.V3) (Tegally *et al.*, 2021; Faria *et al.*, 2021) (*Figure 2*).

Besides the broad landscape of point mutations that affect the Subunit 1 (S1) and Subunit 2 (S2)-RBD regions, specific deletions critically target the S1 domain, identified as members of the mutational panel predominantly in Alpha, Beta, Delta, Eta, Theta variants (Tao *et al.*, 2021).

For example, two deletions in the Alpha variant (69-70del and 144del), with some amino-acid substitutions in the Spike protein, conferred to Alpha higher transmissibility than the wild type (B.1 lineage), but only a modest advantage in resisting neutralizing antibodies. Some deletions are frequently observed in other regions as ORF7 and 8 correlated to low replication load. Because the virus tries to survive by modifying its strength at a low level of activity, point deletions under the pressure of natural selection and evolution sometimes act negatively to viral spreading (Akkiz, 2021; Korber *et al.*, 2020).

Understanding the role of the balance of mutational modifications (substitutions/deletions) in the viral genome is a critical step for increasing specific molecular knowledge.

Alpha, Beta and Gamma variants

Recent works (Martin *et al.*, 2021; Faria *et al.*, 2021) highlighted positive selection events in 151 individual codon sites, of which 80 in Alpha, 41 in Beta and 37 in Gamma variants, indicating a substantial adaptation of these variants from the time of their appearance to March 2021. Of these mutation sites, 22 are defining the lineage: 8/11 for V1, 4/14 for V2 and 13/17 for V3. Among these 22, most noteworthy

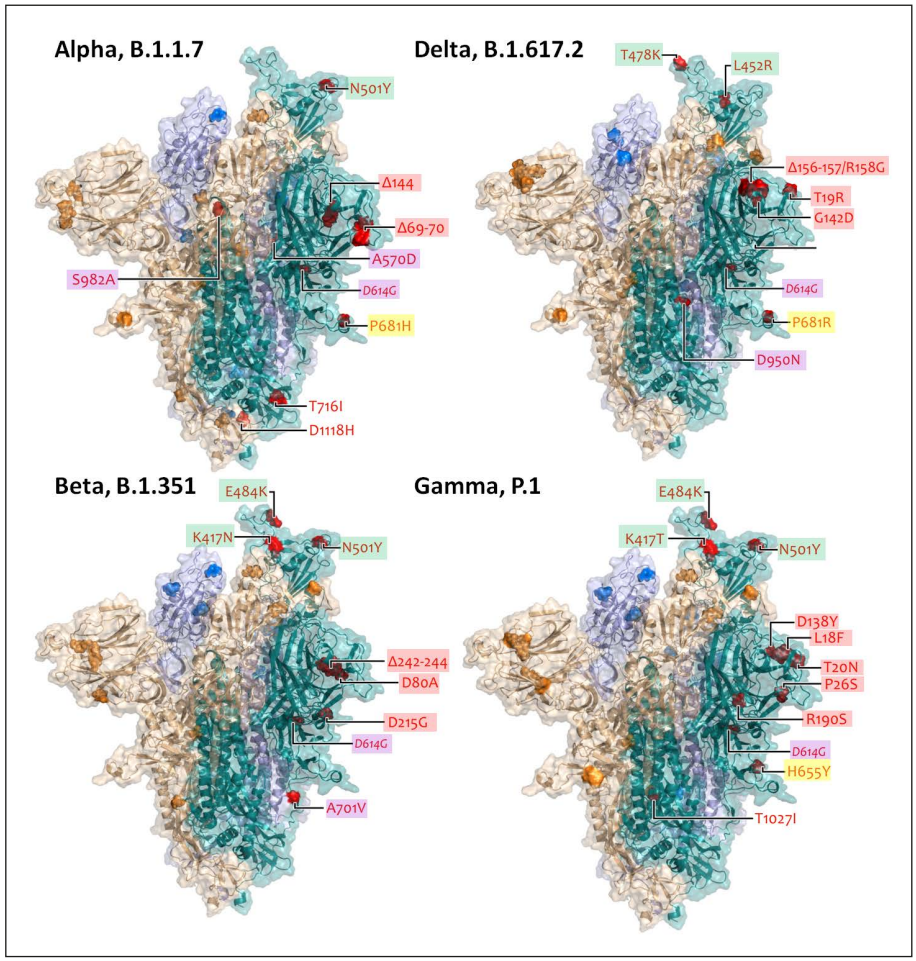

Figure 2 - *Spike protein structures of main variants of concern (VOCs) showing locations of mutated amino acid residues. The three protein chains composing spike are shown in teal, blue and gold. In this images spike is shown in its 'open' conformation. The first relevant substitution, D614G, shared by all lineage B.1 descendants is italicised. Mutation are coloured based on each patter: light red, mutations in the N-terminal domain (NTD), green receptor binding domain (RBD) mutations, yellow fusion cleavage site mutations, pink interchain contact mutations (modified from https://sars2.cvr.gla.ac.uk/cog-uk/).*

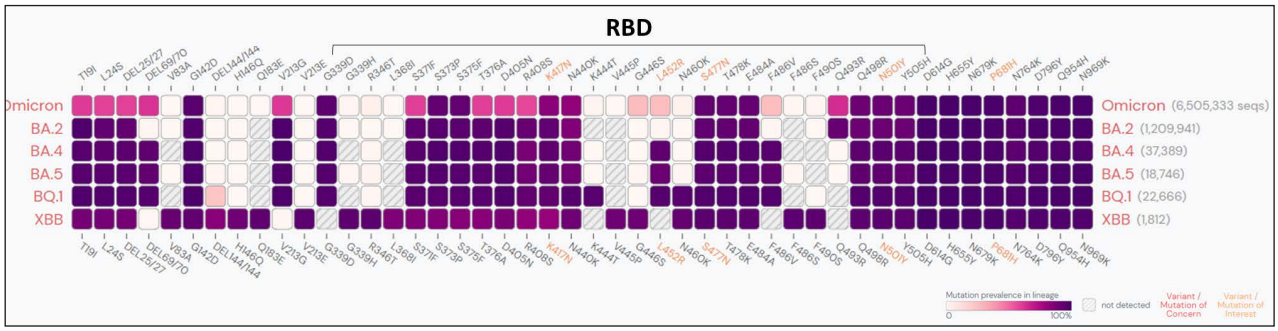

Figure 3 - *Mutations prevalence in spike region across Omicron sublineages (from https://outbreak.info/compare-lineages). Mutations in receptor binding domain (RBD) are indicated.*

for evidence of lineage-specific positive selection are codons S/18, S/80, S/417, S/501, S/655, and S/681, as they are either suspected or known to harbor mutations with potentially significant impacts on fitness (Garry *et al.*, 2021; Greaney *et al.*, 2021a; Greaney *et al.*, 2021b; Lubinski *et al.*, 2022; McCallum *et al.*, 2021; Starr *et al.*, 2020; Wang P. *et al.*, 2021; Wang Z. *et al.*, 2021b; Zahradník *et al.*, 2021).

Convergent or divergent mutation patterns were observed at some mutational sites which, before March 2021, were evolving under positive selection. Moreover, signature mutation sites with lineage-specific signals of positive selection predominantly showed divergent mutations at the amino acid substitution level, according to lineage. Some of these divergent mutations fall within a portion of the Spike N-terminal domain that is an "antigenic supersite", targeted by multiple monoclonal and infection-induced antibodies neutralizers (McCallum *et al.*, 2021).

Delta variant

Since the global dissemination of Delta VOC, continued adaptive evolution has led to a set of Delta sub-lineages, with distinct combinations of mutations, still on the spike protein. The mutation of L452 replaces arginine with leucine, and is found in the RBD's hydrophobic plaques in the S protein, not making any direct contact with the ACE2 receptor. The contemporary presence of this mutation and the T478K mutation is responsible for the increased transmissibility and the more significantly decreased sensitivity to neutralizing antibodies of this variant compared to the Alpha variant (Kudriavtsev *et al.*, 2022) (*Figure 2*).

Recurrent mutations at S:681 enhance S1/S2 subunit cleavage (Lubinski *et al.*, 2022; Liu *et al.*, 2021), which is essential for spike-mediated cell entry (Hoffmann *et al.*, 2020) and contributes to increased viral replication (Liu *et al.*, 2021).

Omicron variant

It is possible that some mutations display a large degree of parallelism due to specific within-host pressures that occur in secondary infections of partially immune individuals, despite having only modest effects on between-host transmission (Chavda *et al.*, 2022).

With the progressive extension of vaccination and the global circulation of the new Omicron variant, the level of seroprevalence in the population grew considerably in the first months of 2022 (https://www.who.int/news/item/03-02-2022-true-extent-of-sars-cov-2-infection-through-seroprevalence-studies). In this context, the high level of population cross-immunity became the prevailing selective force driving the emergency of immune escape mutations.

The Omicron RBD showed 15 mutations, of which only two, L452R and T478K, were shared with the Delta variant and exhibited a similar deletion ORF1a:3674-3676, providing an example of deletion in the Nsp6 region associated with consequential VOC viruses. While these mutations individually reduce receptor binding capacity, the combination with other mutations such as Q498R or N501Y increased this affinity (Moulana *et al.*, 2022) (*Figure 3*).

Among the Omicron sub-lineages, the BA.1 became extinct while the continuous evolution of BA.2 and BA.5 is giving rise to an increasing number of descendants. The fact that BA.4 and BA.5 displayed identical mutational patterns in the 5′ genome region (from ORF1ab to E), but exhibited genetic divergence in the 3′ region (from M to the 3′ genome end), suggested their origin through a recombination event, with breakpoint between the E and M genes (Tegally *et al.*, 2022).

A common mutation of both sub-variants, the F486L, first observed in viruses infecting minks and from human cases linked to mink farms, is in a key site for escape of vaccine-elicited and infection-elicited RBD-targeted antibodies.

RELEVANT STUDIES OF SARS-CoV-2 PHYLOGENY

Thanks to next-generation sequencing tools, the scientific community has collected an unprecedented amount of SARS-CoV-2 genomic data, allowing the

tracking of SARS-CoV-2 evolution. Because of the recent origin of SARS-CoV-2, its initial diversity was relatively limited. This implies that short-term transmission patterns may not leave a detectable footprint in virus genomes, resulting in poorly resolved genomic reconstructions.

Moreover, large spatiotemporal biases exist in the available genome data; around 41% of currently available genomes from Europe have been sampled in the UK, whereas Italy, despite having experienced a similar number of cases and likely an earlier epidemic onset, represents only 1.2% of the genome collection on GISAID (www.gisaid.org, last access on 24 January 2023). Both low sequence diversity and sampling bias confuse the interpretation of transmission patterns so that highly similar SARS-CoV-2 genomes from the same or different locations do not necessarily imply direct linkage. To solve these problem, the integration of additional sources of information is useful. Epidemiological information provides important context to assess genomic sampling biases, and these can be used to subsample genomes by location in situations where large collections of isolates are available (Attwood *et al.*, 2022).

Most of the available phylogenetic studies are related to a limited geographical area or are focused on a new variant emerged recently. In fact, it is difficult to estimate epidemiological parameters (also known as phylodynamic) and to reconstruct evolutionary history when heterogeneous contexts are evaluated.

Europe

A preliminary study evaluating strains collected from January to October 2020 highlighted different situations in the 10 European countries considered. In detail, for Switzerland, Norway, the Netherlands and Belgium more viral import than export events were estimated throughout the considered period and many independent transmission chains since the epidemic started to unfold. The highest proportions of introduction events in the first wave, with a subsequent reduction to predominantly export events, was estimated for Portugal.

France, Italy and Spain resulted characterized by high viral export during the first wave, with the proportions of introductions that remained relatively low for Italy and Spain following the first wave, while in France these proportions were high from mid-June until the end of July. These countries also had comparatively lower entropy values early in the epidemic, with an increase for France by the start of summer and a more gradual increase over time for Italy. However, in Spain the genetic complexity of SARS-CoV-2 transmission chains remained limited (Lemey *et al.*, 2021).

Although in the UK and Germany the viral flow in and out of the country was initially relatively balanced, in the UK a high number of cross-country transmissions was indicated, as more than 2,800 independent introduction events were identified. An increased proportion of introductions and entropy for UK and Germany was observed from mid-June and late summer, respectively. In countries that experienced high summer incidence (e.g., Spain, Portugal, Belgium, and France), the introduction events led to fewer descendants (Lemey *et al.*, 2021).

Phylogenetic analyses conducted on Italian sequences collected from the end of February to mid-June 2020 showed that clusters observed in the first weeks of the epidemic frequently included international isolates and only pure Italian clusters were observed mainly after the lockdown and distancing measures were adopted. Two distinct phylogeographic pattern were observed: the first involved lineage B from China to Veneto, giving rise to a cluster that apparently disappeared in that region; the second was characterized by lineage B.1, entering in Lombardy and spreading further to other Italian regions (mainly in the North) and other European countries. Lineage B.1.1 most probably evolved within Italy and spread from central to other Italian regions, particularly in the South, and to European countries, while lineage B.1.1.1 most probably developed in other European countries, entering Italy only in the second half of March and apparently remaining localized in Piedmont until June 2020. The reconstructed ancestral scenario suggests a central role played by China and Italy in the widespread diffusion of the highly transmissible D614G variant in Europe in the early phase of the pandemic. More dispersed exchanges involved several European countries from the second half of March 2020. Phylogenetic data indicated that outbreaks in both Germany and Italy were independently introduced from China, thereby excluding German responsibility for initiating the Italian outbreak (Lai *et al.*, 2022; Worobey *et al.*, 2020).

Data from Finland (Truong Nguyen *et al.*, 2022) indicated that one third of cases reported in this country derived from a single introduction from Spain during spring 2020. Despite being widespread in Europe during the summer of 2020, the 20E.EU1 variant (Hodcroft *et al.*, 2021; Lemey *et al.*, 2021) was not detected at a high level, suggesting that it did not widely contribute to the second wave in Finland. The phylogeographic reconstructions reported that more than 40 independent introductions from Italy, Austria and Spain occurred in this country during the second week of March, as supported by epidemiological records of travels. A few additional introductions were reported from Germany, Sweden, Switzerland, France, the United Kingdom, Denmark, Turkey, the Netherlands, Latvia, Estonia, Poland, Norway and Hungary. The small number of lineages characterizing the large majority of introductions highlights the extensive heterogeneity in SARS-CoV-2 transmission dynamics underlying the establishment of local transmission chains.

North/South America

Multiple introductions of genetically similar viruses were reported in the United States, starting from the beginning of February 2020, with the largest transmission cluster dated mid-February 2020. Italian and European clusters were indicated as the source of multiple introductions to the US.

Phylogeographic analyses revealed at least 102 international introductions of SARS-CoV-2 in Brazil, considering international strains collected from December 2019 to April 2020 (Candido *et al.*, 2020). In the first phase of the epidemic, authors find an increasing number of international introductions until mid-Mar 2020, predominantly acquired from Italy and the US. After this initial phase, the estimated number of international imports decreased concomitantly with the decline in the number of international passengers travelling to Brazil. In contrast, despite the declines in the number of passengers travelling on national flights, an increase in viral lineages exchanged among Brazilian regions, at least until early April 2020, was observed and characterized by long distance movements. These findings emphasized the roles of within and between migrations as a key driver of both local and inter-regional virus spread, with highly populated and well-connected urban conurbations in the southeast region acting as the main sources of virus exports within the country.

RECENT HYPOTHESIS ON PANDEMIC EVOLUTION

In agreement with a recent revision (Balloux *et al.*, 2022), the main forces driving the dynamics of a respiratory virus epidemic are related to: seasonality (due to environmental, physical climate-related factors and human behaviors); level of cross-immunity in the population; viral evolution (antigenic drift), and application of mitigation measures during the epidemic waves. These four forces affect the Rt value of the epidemic. Indeed, the unstable level of cross-immunization of the population, due to the waning of the response, and the ability of the virus to produce escape mutants resulted in continuous fluctuations of Rt.

During the early phase of the epidemic, the low level of immunization of the population meant that seasonality had little effect, since the other three forces were prevalent. As the level of population cross-immunity increases due to natural infections and/or vaccination, it could lead SARS-CoV-2 to acquire a seasonal epidemic dynamic similar to other respiratory viruses (such as influenza or the other common human coronaviruses) (Balloux *et al.*, 2022).

CONCLUSIONS

Determining the transmissibility, prevalence and spread of SARS-CoV-2 infections is central to understanding the impact of the pandemic and to the design of effective control strategies (Attwood *et al.*, 2022). In this context, phylodynamic approaches that combined evolutionary, demographic, and epidemiological concepts have helped track virus genetic changes, identify emerging variants and inform public health strategy.

Because changes in mobility may impact global and local transmission of SARS-CoV-2, the combination of genomic and mobility data can complement traditional surveillance approaches (Candido *et al.*, 2020). Phylogenetic studies highlight the importance of continued global genomic surveillance as an early warning system, giving countries time to prepare and mitigate the public health effects of new emerging variants (Tegally *et al.*, 2022).

References

Akkiz H. (2021). Implications of the Novel Mutations in the SARS-CoV-2 Genome for Transmission, Disease Severity, and the Vaccine Development. *Front Med (Lausanne)*. **8**, 636532.

Attwood S.W., Hill S.C., Aanensen D.M., Connor T.R., Pybus O.G. (2022). Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nat Rev Genet*. **23**, 547-562.

Bálint G., Vörös-Horváth B., Széchenyi, A. (2022). Omicron: increased transmissibility and decreased pathogenicity. *Signal Transduct Target Ther.* **7**, 151.

Balloux F., Tan C., Swadling L., Richard D., Jenner C., et al. (2022). The past, current and future epidemiological dynamic of SARS-CoV-2. *Oxford Open Immunology*. **3**.

Boehm E., Kronig I., Neher R.A., Eckerle I., Vetter P., et al. (2021). Novel SARS-CoV-2 variants: the pandemics within the pandemic. *Clin Microbiol Infect*. **27**, 1109-1117.

Candido D.S., Claro I.M., de Jesus J.G., Souza W.M., Moreira F.R.R., et al. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. **369**, 1255-1260.

Chavda V.P., Bezbaruah R., Deka K., Nongrang L., Kalita T. (2022). The Delta and Omicron Variants of SARS-CoV-2: What We Know So Far. *Vaccines (Basel)*. **10**.

Daniloski Z., Jordan T.X., Ilmain J.K., Guo X., Bhabha G., et al. (2021). The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife*. **10**.

Dhawan M., Saied A.A., Mitra S., Alhumaydhi F.A., Emran T.B., et al. (2022). Omicron variant (B.1.1.529) and its sublineages: What do we know so far amid the emergence of recombinant variants of SARS-CoV-2? *Biomed Pharmacother.* **154**, 113522.

Dhawan M, Sharma A Priyanka, Thakur N, Rajkhowa TK, et al. (2022). Delta variant (B.1.617.2) of SARS-CoV-2: Mutations, impact, challenges and possible solutions. *Human Vaccines & Immunotherapeutics*. **18**, 2068883.

Faria N.R., Mellan T.A., Whittaker C., Claro I.M., Candido D.D.S., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. **372**, 815-821.

Focosi D., Maggi F. (2022). Recombination in Coronaviruses, with a Focus on SARS-CoV-2. *Viruses*. **14**, 6.

Funk T., Pharris A., Spiteri G., Bundle N., Melidou A., et al. (2021). Characteristics of SARS-CoV-2 variants of concern B.1.1.7, B.1.351 or P.1: data from seven EU/EEA countries, weeks 38/2020 to 10/2021. *Euro Surveill.* **26**.

Galloway S.E., Paul P., MacCannell D.R., Johansson M.A., Brooks J.T., et al. (2021). Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morb Mortal Wkly Rep.* **70**, 95-99.

Gobeil S.M., Janowska K., McDowell S., Mansouri K., Parks R., et al. (2021). Effect of natural mutations of SARS-CoV-2 on spike structure, conformation and antigenicity. *bioRxiv*.

Graham R.L., Baric R.S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol.* **84**, 3134-3146.

Greaney A.J., Loes A.N., Crawford K.H.D., Starr T.N., Malone K.D., et al. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe.* **29**, 463-476.

Greaney A.J., Starr T.N., Gilchuk P., Zost S.J., Binshtein E., et al. (2021). Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe.* **29**, 44-57.

Hodcroft E.B., Zuber M., Nadeau S., Vaughan T.G., Crawford K.H.D., et al. (2021). Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature.* **595**, 707-712.

Hoffmann M., Arora P., Groß R., Seidel A., Hörnich B.F., et al. (2021). SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell.* **184**, 2384-2393.

Hoffmann M., Kleine-Weber H., Pöhlmann S. (2020). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell.* **78**, 779-784.

Kistler K.E., Huddleston J., Bedford T. (2022). Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. *Cell Host Microbe.* **30**, 545-555.

Korber B., Fischer W.M., Gnanakaran S., Yoon H., Theiler J., et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell.* **182**, 812-827.

Krause P.R., Fleming T.R., Longini I.M., Peto R., Briand S., et al. (2021). SARS-CoV-2 Variants and Vaccines. *New England Journal of Medicine.* **385**, 179-186.

Kudriavtsev A.V., Vakhrusheva A.V., Novoseletsky V.N., Bozdaganyan M.E., Shaitan K.V., et al. (2022). Immune Escape Associated with RBD Omicron Mutations and SARS-CoV-2 Evolution Dynamics. *Viruses.* **14**.

Lai A., Bergna A., Toppo S., Morganti M., Menzo S., et al. (2022). Phylogeography and genomic epidemiology of SARS-CoV-2 in Italy and Europe with newly characterized Italian genomes between February-June 2020. *Sci Rep.* **12**, 5736.

Lai A., Bergna A., Menzo S., Zehender G., Caucci S., et al.. (2021). Circulating SARS-CoV-2 variants in Italy, October 2020-March 2021. *Virol J.* **18**, 168.

Lai A., Bergna A., Della Ventura C., Menzo S., Bruzzone B., et al.. (2022). Epidemiological and Clinical Features of SARS-CoV-2 Variants Circulating between April-December 2021 in Italy. *Viruses.* **14**, 2508.

Lemey P., Ruktanonchai N., Hong S.L., Colizza V., Poletto C., et al. (2021). Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature.* **595**, 713-717.

Liu Y., Liu J., Johnson B.A., Xia H., Ku Z., et al. (2021). Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *bioRxiv.*

Luan B., Wang H., Huynh T. (2021). Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. *FEBS Lett.* **595**, 1454-1461.

Lubinski B., Fernandes M.H.V., Frazier L., Tang T., Daniel S., et al. (2022). Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience.* **25**, 103589.

MacLean O.A., Lytras S., Weaver S., Singer J.B., Boni M.F., et al. (2021). Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLOS Biology.* **19**, e3001115.

Martin D.P., Weaver S., Tegally H., San J.E., Shank S.D., et al. (2021). The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell.* **184**, 5189-5200.

McCallum M., De Marco A., Lempp F.A., Tortorici M.A., Pinto D., et al. (2021). N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell.* **184**, 2332-2347.

McLeod D.V., Gandon S. (2022). Effects of epistasis and recombination between vaccine-escape and virulence alleles on the dynamics of pathogen adaptation. *Nat Ecol Evol.* **6**, 786-793.

Moulana A., Dupic T., Phillips A.M., Chang J., Nieves S., et al. (2022). Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat Commun.* **13**, 7011.

Pekar J.E., Magee A., Parker E., Moshiri N., Izhikevich K., et al. (2022). The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science.* **377**, 960-966.

Rambaut A., Holmes E.C., O'Toole Á., Hill V., McCrone J.T., et al. (2021). Addendum: A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* **6**, 415.

Rochman Nash D., Faure G., Wolf Yuri I., Freddolino P.L., Zhang F., et al. (2022). Epistasis at the SARS-CoV-2 Receptor-Binding Domain Interface and the Propitiously Boring Implications for Vaccine Escape. *mBio.* **13**, e00135-00122.

Srivastava S., Banu S., Singh P., Sowpati D.T., Mishra R.K. (2021). SARS-CoV-2 genomics: An Indian perspective on sequencing viral variants. *J Biosci.* **46**.

Starr T.N., Greaney A.J., Hilton S.K., Ellis D., Crawford K.H.D., et al..(2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell.* **182**, 1295-1310.

Tao K., Tzou P.L., Nouhin J., Gupta R.K., de Oliveira T., et al. (2021). The biological and clinical significance of emerging SARS-CoV-2 variants. *Nature Reviews Genetics.* **22**, 757-773.

Tegally H., Moir M., Everatt J., Giovanetti M., Scheepers C., et al. (2022). Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat Med.* **28**, 1785-1790.

Tegally H., Wilkinson E., Giovanetti M., Iranzadeh A., Fonseca V., et al. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature.* **592**, 438-443.

Thomson E.C., Rosen L.E., Shepherd J.G., Spreafico R., da Silva F.A., et al. (2021). Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell.* **184**, 1171-1187.

Truong Nguyen P., Kant R., Van den Broeck F., Suvanto M.T., Alburkat H., et al. (2022). The phylodynamics of SARS-CoV-2 during 2020 in Finland. *Communications Medicine.* **2**, 65.

Turakhia Y., Thornlow B., Hinrichs A., McBroome J., Ayala N., et al. (2022). Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature.* **609**, 994-997.

Turoňová B., Sikora M., Schürmann C., Hagen W.J.H., Welsch S., et al. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science.* **370**, 203-208.

VanInsberghe D., Neish A.S., Lowen A.C., Koelle K. (2021). Recombinant SARS-CoV-2 genomes are currently circulating at low levels. *bioRxiv.*

Wang P., Nair M.S., Liu L., Iketani S., Luo Y., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature.* **593**, 130-135.

Wang Z., Schmidt F., Weisblum Y., Muecksch F., Barnes C.O., et al. (2021). mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature.* **592**, 616-622.

Weissman D., Alameh M.G., de Silva T., Collini P., Hornsby H., et al. (2021). D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. *Cell Host Microbe.* **29**, 23-31.

Worobey M., Pekar J., Larsen B.B., Nelson M.I., Hill V., et al. (2020). The emergence of SARS-CoV-2 in Europe and North America. *Science.* **370**, 564-570.

Zahradník J., Marciano S., Shemesh M., Zoler E., Chiaravalli J., et al. (2021). SARS-CoV-2 RBD *in vitro* evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv.*

Zehender G., Lai A., Bergna A., Meroni L., Riva A., et al. (2020). Genomic characterization and phylogenetic analysis of SARS-COV-2 in Italy. *J Med Virol.* **92**, 1637-1640.

Zhou W., Xu C., Wang P., Luo M., Xu Z., et al. (2021). N439K Variant in Spike Protein Alter the Infection Efficiency and Antigenicity of SARS-CoV-2 Based on Molecular Dynamics Simulation. *Front Cell Dev Biol.* **9**, 697035.