New Microbiologica, 44, 4, 191-204, 2021, ISSN 1121-7138

REVIEW

Evolution of viruses and the emergence of SARS-CoV-2 variants

Giulia Freer[1*], Michele Lai[1*], Paola Quaranta[1], Pietro Giorgio Spezia[1], Mauro Pistello[1,2]

[1]*Retrovirus Center, Virology Section, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa;*
[2]*Virology Unit, Pisa University Hospital, Pisa, Italy*

**These authors contributed equally*

SUMMARY

Life implies adaptation. This is one of the fundamental principles that has permitted most living species to survive through ages in an ever-changing environment. Spontaneously occurring events have shaped also virus populations and their fitness. Thanks to their plasticity, viruses have thrived in extremely dissimilar conditions. Unsurprisingly, SARS-CoV-2, the etiological agent of COVID-19, is no exception. Thanks to an unprecedented rate of molecular tracing and sequence scrutiny, the virus was followed in all its changes and shown to evolve in such a way as to possibly determine subsequent waves of infection after the first global and massive outbreak. This review illustrates the major modifications occurred to the virus since its discovery. We describe the potential advantages that these changes conveyed as regards SARS-CoV-2 transmissibility, resistance to host innate and adaptive barriers and molecular diagnosis.

Received December 12, 2021 *Accepted* December 15, 2021

INTRODUCTION

As of November 2021, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has accounted for nearly 260 million infections and over five million deaths [WHO Coronavirus (COVID) Dashboard, https://covid19.who.int]. These dramatic figures show that despite a worldwide mass vaccination campaign that injected over 7 billion vaccine doses, SARS-CoV-2 is still able to circulate. Capillary tracing of infected people clearly demonstrates that COVID vaccines work well but booster doses are necessary to reduce waning of immunity and prolong protection from infection. Countries with higher responses to vaccination campaigns resisted better to subsequent pandemic waves. As further confirmation that vaccination is effective, categories who did not want or could not be vaccinated are the most vulnerable to infection and disease and, often, a source of local outbreaks (Khan *et al.,* 2021). Unfortunately, vaccination took place prevalently in high-income countries, where preventive and safety measures were enforced with uneven scales and policies with the result that, on the verge of 2021 winter season, the incidence of infection is surging again, albeit at a different scale all over the world.

Since the emergence of the first massive outbreak, the world has witnessed subsequent waves of infection that periodically swept Western countries. Real-time monitoring and molecular characterization of circulating virus have allowed to dissect the major forces shaping subsequent bursts of cases.

Kinetics and dynamics of the emergence of viral variants have been studied at an unprecedented rate of detail and, at the time of writing, this effort has generated over two million SARS-CoV-2 sequences available *via* the Global Initiative on Sharing All Influenza Data (GISAID). Pinpoint characterization of circulating virus demonstrate that the original virus first identified in the village market of Wuhan in December 2019 (Haage *et al.,* 2021) progressively disappeared in subsequent waves to leave room to viral variants that first surfaced in restricted geographical areas. These were, in turn, diluted out, depending on a variety of factors, and disappeared or became predominant with time. As expected, based on similar studies previously carried out on other viruses, e.g., human immunodeficiency virus (HIV), influenza viruses (IV) and others, mutations were unevenly dis-

Key words:
SARS-CoV-2, COVID-19, fitness, variant of concern, spike protein, COVID vaccine.

Corresponding author:
Mauro Pistello
E-mail: mauro.pistello@unipi.it

tributed across the viral genome. The success of mutations most likely relies on the selective advantages they convey, which, in turn, depend on various circumstances. This review provides a description of mechanisms of virus variability and its payoffs in general. It also describes how these mechanisms shape circulating SARS-CoV-2 variants by pinpointing how their signature mutations influenced transmissibility, resistance to neutralization by specific immune response, and other factors.

How and why do viruses vary?

Adaptation takes place in all living organisms and throughout the evolution scale. Mechanisms and kinetics differ; largely, the appearance of mutations underlie most changes. These consist of wrong bases introduced randomly in the genetic material during nucleic acid synthesis. Therefore, the more an organism replicates, the greater are its mutation rate and plasticity. Mutations may be non-synonymous or synonymous: the former are mutations that lead to amino acid changes, the latter have no effects at the translational level but may play role in genome processing. Once an amino acid has changed because of a non-synonymous mutation, the organism's fitness - e.g. ability to replicate and transmit, resistance to environment (Wargo and Kurath 2012; DeLong *et al.*, 2021) - may also change. If fitness improves, the mutation is transmitted to progeny, if it declines, the mutation is deemed deleterious and diluted out during replication (*Figure 1*). Within these boundaries, there is a wide number of synonymous and non-synonymous mutations that have no obvious effects and are considered neutral in a given environment. If the latter changes, such mutations may not be neutral anymore and result advantageous or disadvantageous. Indeed, as schematically illustrated in *Figure 1*, such mutations may overcome restraints in viral spread and substitute the founder viruses with a new viral strain.

Interestingly, host factors aimed at fighting viral replication may increase mutation rate, mainly by depletion of CpG dinucleotides (a cytosine followed by a guanine in the 5' to 3' direction) in virus genomes by the Zinc finger antiviral protein or by APOBEC3 cytidine deaminases (MacLean *et al.*, 2021). The rate with which the mutations are selected is slow in organisms that have long life cycles and faster in bacteria and viruses that replicate more rapidly (Peck and Lauring 2018). This is not only a question of turnaround time between generations: higher organisms possess a proof-reading system active that corrects

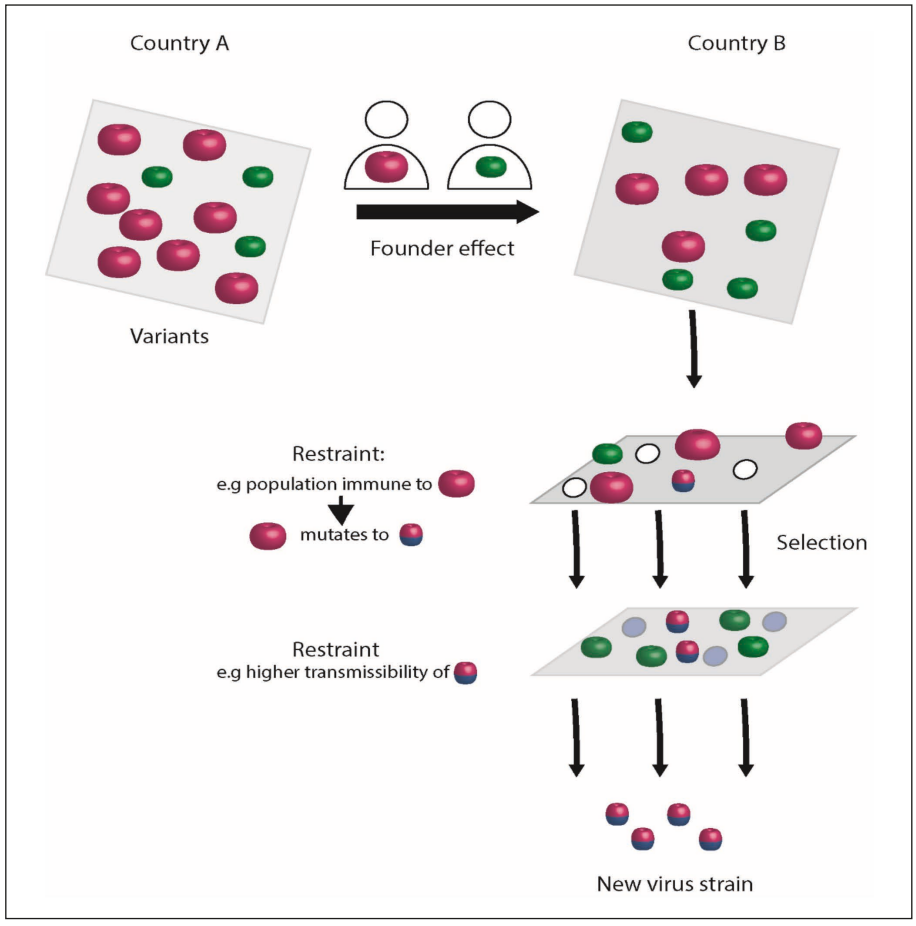

Figure 1 - *Effects of a non-synonymous mutation on the spread of viruses. Schematic illustration on how viruses mutate in response to the environment. Briefly, the founder effect might allow prevalence of variants that have no advantage from a country to another. Then, several restraints, such as the adaptive immunity, might limit the diffusion of some variants and contribute to the emergence of new viral strains with increased environmental adaptation.*

nucleotides erroneously introduced into the newly synthesized strand during the nucleic acid synthesis phase. This system limits misincorporation to a few units per hundred thousand bases. It works so efficiently that higher organisms, like mammals and plants, end up being genetically stable, as opposed to bacteria and viruses. If bacteria tolerate (and take advantage) of moderate numbers of changes, viruses favor mutations and genome plasticity (Weaver *et al.,* 2021). Viruses with a DNA genome tend to exploit cellular DNA polymerases. However, many use viral DNA polymerases which operate faster and are much less accurate compared to cellular DNA polymerases. The net result is a higher rate of misincorporation, in the range of one nucleotide per thousand bases. Genome plasticity of RNA viruses is much higher. Their RNA polymerases are intrinsically inaccurate and have little or no proof-reading system, thus raising the bar of misincorporation to one mutation per few hundred bases. A swarm of viral variants is therefore generated every round of replication, whose genetic complexity and diversity are proportional to viral flexibility (Edwards *et al.,* 2021).

More dramatic changes in positive-sense RNA virus genomes can be brought about by recombination. This mechanism takes place when a cell is co-infected by two compatible viral strains, or even species. It depends on the fact that RNA-dependent RNA polymerases (RdRp) may jump from an RNA template to another one while elongating genomic RNA. The result is a chimeric genome deriving partly from one parental virus and partly from another (Bentley and Evans 2018). Recombination has played a fundamental role in the emergence of SARS-CoV-2 from the original bat host, where the closest bat virus to SARS-CoV-2, RmYN02, is a virus resulting from recombination (MacLean *et al.,* 2021).

The net effect of viral plasticity is fast adaptation to continuously changing conditions in the environment where viruses aim to survive and thrive. For instance, a virus that infects *via* the respiratory tract must, first and foremost, overcome the anatomic-functional barriers of the apparatus (mucus and ciliary movement), survive various substances with antimicrobial activity then resist to a formidable network of resident phagocytic cells. Once it enters cells, it must resist intrinsic cellular restriction factors and, eventually, circumvent an adaptive immune response. Genetic changes helping the virus to bypass the above hurdles are bound to be positively selected. Protein changes that improve viral capacity of replication, persistence in the host, and transmissibility also improve viral fitness (Edwards *et al.,* 2021; Weaver *et al.,* 2021).

Genetic variation is filtered out by inherent properties of the virus, such as constraints of protein conformation, route of transmission, and host cell cycle. Viruses like HIV and hepatitis C virus (HCV), that

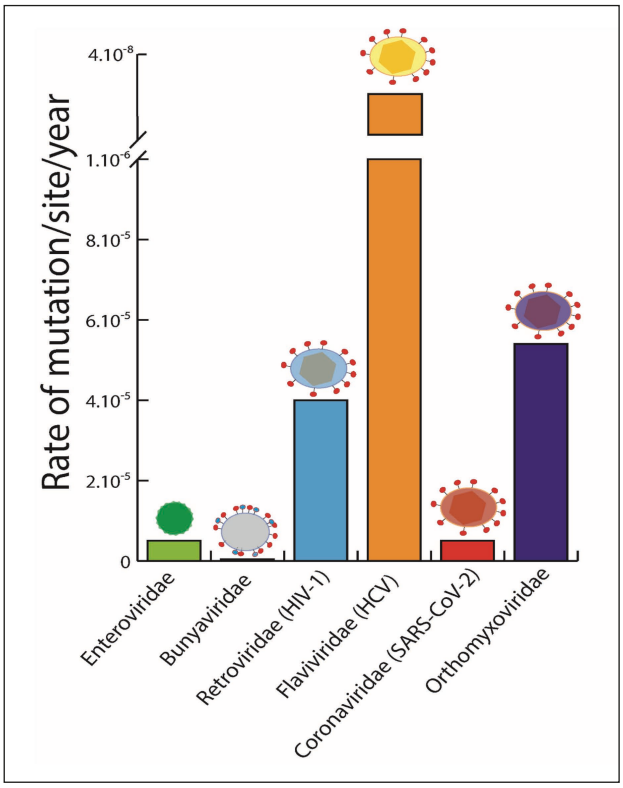

Figure 2 - *Scheme of SARS-CoV-2 variation as compared to other human viruses. The higher genetic stability of SARS-CoV-2, which depends on the proof-reading activity of the RdRp enzyme, is limited to 1.10-3 mutations/site/year. As illustrated, this rate is lower than for other viruses such as Retroviridae, Flaviviridae, and Orthomyxoviridae.*

replicate for long periods in the host, produce millions of viral variants per day in an infected individual (*Figure 2*). These swarms of variants, also called quasispecies, are selected primarily for their ability to evade the immune response and resist to therapeutic treatment. Variants of types A and D IVs and coronaviruses (CoVs), that normally cause rapidly cleared infections, are also purged based on their capacity to resist the external environment and infect individuals of different species (Graham and Baric 2010; Liu *et al.,* 2020; Ciminski *et al.,* 2021). Yet other viruses, like rhinoviruses, seem to be selected for efficient replication in the upper respiratory tract and for being more easily transmitted (Levin *et al.,* 1999; Moya *et al.,* 2000; Domingo-Calap *et al.,* 2019).

Attempts to link single changes to specific properties is an overly simplification for viruses, that encode proteins with multiple functions. Each non-synonymous mutation can, therefore, influence several apparently unrelated features (Snedden *et al.,* 2021). Because of their dynamism and plasticity, each virus has its own mutational frequency that shapes and maximizes its fitness in response to external condi-

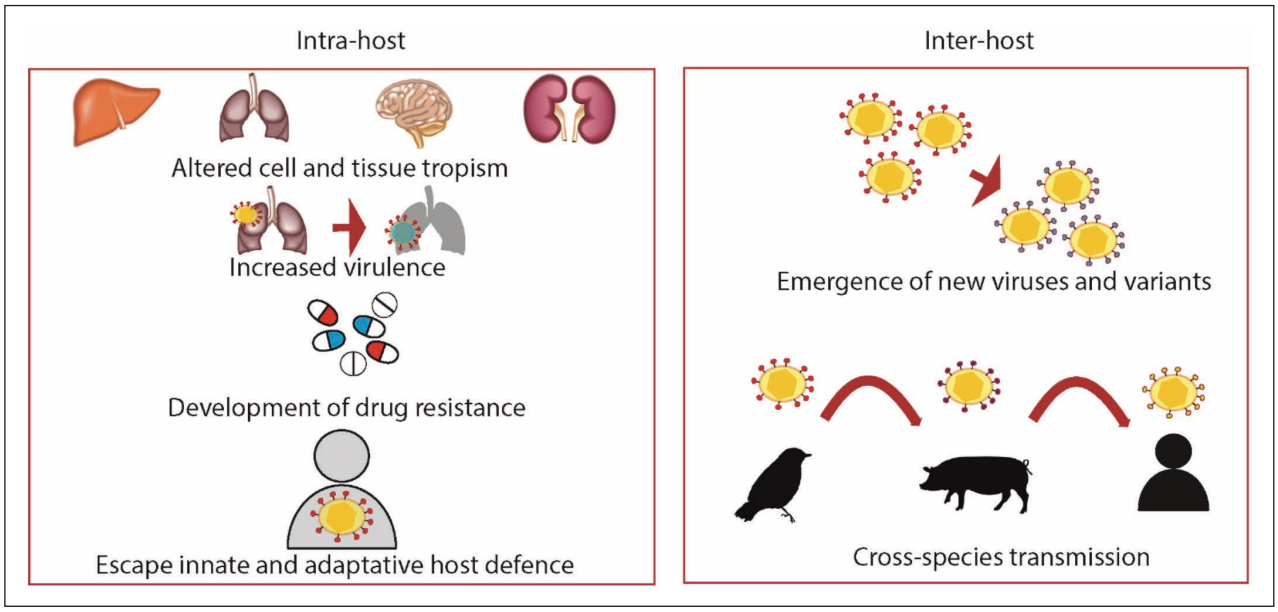

Figure 3 - *Benefits of viral mutations. Payoff of sequence and structural changes in the viral genome might occur intra-host or inter-host. In the first case, mutations might enhance viral tropism to different organs, increase virulence, escape adaptive immunity, or develop drug resistance. In the second case, a new variant might emerge and evolve to increase the number of different animal species infected, including humans.*

tions (*Figure 3*). The net effect is a fine interaction of the virus with its host permitting a level of subtle and prolonged parasitism that has no equal among mammalian microorganisms.

Benefits of viral plasticity

The advantages conferred to a virus by its ability to vary are quite considerable and, in most cases, have pleiotropic effects. Figure 3 shows some of the main benefits of mutations. Regarding sequence changes altering the spectrum of susceptible cells and tissues, a well-known example is HIV-1 and the amino acid changes in the hypervariable V3 domain of gp120, the viral receptor. Depending on a few amino acids in the V3 domain, the virus acquires the ability to interact with either the CXCR4 co-receptor and infect CD4 T-lymphocytes or with CCR5, belonging to monocyte/macrophage lineage. When transmitted *via* mucosa, the virus initiates infection of resident tissue macrophages and then, by subsequent changes in gp120, will infect T-lymphocytes. Of note, macrophage-tropic variants are disseminated in the body by blood macrophages that trespass the brain-blood barrier and reach the central nervous system. Here, they release infectious viruses that, in turn, infect glial and other nervous cells. These cells are resistant to infection by lymphotropic variants that, on the other hand, replicate efficiently in immature and activated T-cells (Burns and Desrosiers 1994; Sundaravaradan *et al.,* 2007). Other examples of changes altering tissue tropism and, therefore, virulence has been reported for Zika virus, an arthropod-borne flavivirus that circulates in various lineages endemic in different geographical areas. Some of these lineages exhibit high neurovirulence that has been correlated to specific mutations conferring the virus high capacity to invade the central nervous system and infect neuroprogenitor cells (Metsky *et al.,* 2017). Another interesting example of genome plasticity linked to different diseases is the feline leukemia virus (FeLV), a retrovirus that exists in three prototypes: one naturally circulating (prototype A), and two generated by either recombination between exogenous and endogenous FeLV strains (B) or mutations within the *env* gene (C). Interestingly, genome rearrangement confers high pathogenicity to type B and C, which become oncogenic causing erythroid aplasia. On the contrary, it is detrimental for their circulation, since only type A, and very few isolates of type B, are further transmitted (Jarrett 1992).

The relationship between viral variability and escape from innate and adaptive immunity has been studied in detail and reported by thousands of publications. Many of them concern HIV-1 and HCV, whose plasticity and ability to circumvent host immune response is well-known (Petrovic *et al.,* 2012; Dustin *et al.,* 2016). There are many other examples. Among them: the porcine epidemic diarrhea virus (PEDV), a member of *Coronaviridae* causes acute diarrhea/vomiting, dehydration, and high mortality in seronegative neonatal piglets. Similarly to anti-SARS-CoV-2 vaccine, the PEDV vaccine is used on a massive scale but does not prevent periodic epidemic waves and violent outbreaks in intensive farming. In

most cases, resurgence of infection is sustained by the emergence of virus variants (Jung and Saif 2015). In addition, the respiratory syncytial virus and IV continuously undergo a process of mutation/selection to escape neutralization by antibodies, called antigenic drift (Ascough *et al.,* 2018).

Mutations in HBV surface protein are responsible for the failure of immune prophylaxis in infants and liver transplant recipients who received HBV vaccine and hepatitis B immune globulin, respectively (Chotiyaputta and Lok 2009).

Mutations associated to the development of full or increased drug resistance have direct impact on patient management and therapeutic strategies. Here again, HIV-1 has been the virus studied most extensively. A wide array of genome changes clearly associated to resistance to single or multiple drugs have been pinpointed (Scourfield *et al.,* 2011; Bandera *et al.,* 2019). In this regard, HIV-1 monotherapy, performed at the dawn of HIV-1 therapy when the number of antiretroviral drugs was very limited, clearly demonstrated that viral evolution *in vivo* can blunt effective therapy in just a matter of days (Mansky 2002). Other important examples in the clinical setting are hepatitis B virus (HBV) and Cytomegalovirus (CMV). HBV replicates through reverse transcription of an RNA intermediate and the inherent lack of proofreading activity causes high mutation frequency. Drug treatment may select for mutations in HBV RNA-dependent DNA polymerase, conferring resistance to nucleoside and nucleotide analog treatment. Indeed, this is one of the main barriers to the success of anti-HBV therapy. The development of drug resistance negates antiviral treatment response and can lead to hepatitis flares and hepatic decompensation unless another drug to which the virus is not cross-resistant is promptly added. As for CMV, this widespread virus establishes latent infection that can become chronically active consequently to deterioration of host immunity. Chronically active CMV infection is one of the most common complications in solid organ transplant recipients, who are at high risks of graft loss, morbidity, and mortality. Preemptive therapy, usually administered for prolonged periods of time in this category of patients, is the only effective measure to abate CMV replication and its consequences. In the last years, the number of clinically approved drugs against CMV increased significantly. These target different viral proteins and combinatorial drug approaches are being developed. Although newly synthesized CMV genomes are highly proof-read, CMV can swiftly develop drug resistance even during combinatorial therapy. This is often achieved through various combinations of mutations, some of which occurring also outside the drug-targeted genome, and difficult to interpret in relation to loss of drug sensitivity (Lurain and Chou 2010; Chou 2020).

As illustrated in the right panel of *Figure 3*, e.g., emergence and re-emergence of viruses and cross-species transmission, are currently the two better known ones because of their implications in the current pandemic. The boundaries between emergence and reemergence are very subtle: there are human viruses that survive in the environment and persist by lurking and slowly evolving in populations where they periodically emerge. Yet others infect multiple hosts and may have parallel evolutions, until mutations occurring in the animal host increase viral fitness to replicate in the human species where they re-emerge starting new outbreaks. Most of these viruses have RNA genomes that display high mutation rates (as high as 1/1,000 bases). As additional sources of variation, they engage in frequent recombination and reassortment, creating novel genotypes from co-circulating strains. Human influenza is a typically emerging infectious diseases, since IV may undergo single point mutations in genes encoding the surface proteins. As mentioned above, these mutations can reduce sensitivity of IV to neutralizing antibodies (NAbs) elicited by prior infections or vaccination and ignite the endemic waves occurring periodically during the winter seasons. Segmented viruses, like IV, can entirely change their pathogenic properties by shuffling their genome segments with those of viral strains infecting different species. This phenomenon, known as reassortment, creates a new virus that may start a pandemic, as has happened in the past on various occasions.

A virus recently re-emerged is Zika virus, that caused a pandemic emergence in very few months. Although Zika virus has been known for decades, it has never caused such a dramatic human outbreak as the one of 2015, when it suddenly spread around the tropical belt, causing millions of infections, birth defects and fetal losses. The cause of the sudden emergence of Zika virus as a pandemic strain is likely to be a mutation resulting in the change of a single amino acid in the external viral glycoprotein occurring in the Brazilian strain (Morens and Fauci 2020; Shan *et al.,* 2020).

Cross-species transmission, known as spillover, has produced significant animal and human morbidity over the past 50 years, i.e., the time period during which this aspect has been monitored closely. It has always taken place over the centuries and has shaped the array of viruses infecting each animal species. In humans, HIV-1 is among the best characterized example. It causes the acquired immunodeficiency syndrome (AIDS) and is the result of multiple cross-species transmission events of simian immunodeficiency viruses (SIVs) naturally infecting African primates. One transmission event, involving SIV from chimpanzees in Cameroon, gave rise to the major HIV-1 M group, that has started the global pandemic causing nearly forty million deaths so far (Sharp and

Hahn 2011). Close investigation of HIV-1 spillover has led to demonstrate that jumping from another species and adaptation into the new host is the result of multiple driving forces. The parental virus needs to have some inherent properties that predisposes it to switch host. Using cellular receptors highly conserved across phylogeny, stability outside the host, suitable routes of transmission, a replicative cycle hijacking common cellular enzymes and biochemical pathways, are only some of the predisposing conditions. Inter-species transmission takes place following genetic mutations or genome rearrangements that may occur before or after the switch. Genetic changes emerging in the new host are positively selected by the environment in the new niche because they increase the fitness of the zoonotic virus resulting in progressive adaptation to the new host. Adaptation implies optimization of replication in a new type of cell(s) and adjustment of cell and organ tropism. Moreover, after evading innate and adaptive immunity, the virus must find a way to start a chain of inter-individual transmission in the new host. In the initial phase, host restriction factors are particularly efficient in restraining the novel virus, yet there are countless numbers of inter-species transmission events among animal species and to humans. The most infamous gave rise to the Spanish flu pandemic that took place after the Great War and killed five times more people than the war itself.

Except for IV and HIV that have passed to humans from avian and farming species and non-human primates, respectively, the majority of zoonotic viruses have been transmitted by bats either directly to humans or through an intermediate animal (Calisher *et al.,* 2006; Wong *et al.,* 2007). Interestingly, bats appear to serve as a reservoir for all known human CoVs and SARS-CoV. The first CoV causing severe acute respiratory syndrome resembling COVID. SARS-CoV was first reported in late 2002 in Guangdong Province, China, not too far from the spillover place of SARS-CoV, the disease, which quickly spread to many countries over a period of 4 months spanning late 2002 and early 2003, infected over 8,000 individuals and killed nearly 800 before it was successfully contained by aggressive public health intervention strategies. Initial assessments determined that the virus crossed to human hosts from zoonotic reservoirs, including bats, Himalayan palm civets (*Paguma larvata*), and raccoon dogs (*Nyctereutes procyonoides*). Like currently circulating SARS-CoV-2 virus, SARS-CoV jumped to humans from animals sold in wet markets in Guangdong Province, and mutations/recombination in the Spike (S) attachment protein, both within and outside of the receptor binding domain (RBD), have likely mediated the emergence of CoVs in the new host population (Graham and Baric 2010; Morens and Fauci 2020).

Spillover and evolution of SARS-CoV-2

SARS-CoV-2 is a positive-sense single stranded RNA virus with a large genome (26.4–31.7 kb) (Woo *et al.,* 2010) belonging to sarbecovirus subgenus of betacoronaviruses. It shares close genomic similarity to SARS-CoV (79% identity) and Middle East respiratory syndrome CoVs (MERS-CoV, 50% identity) (Lu *et al.,* 2020). As mentioned, these viruses have originated from bats and were transmitted to other mammals, including humans. Both SARS-CoV and SARS-CoV-2 enter the human host through interaction of their receptor, the S protein, with angiotensin-converting enzyme 2 (ACE2) present on the membrane of host epithelial cells (Snedden *et al.,* 2021). Specifically, the S protein interacts with ACE2 through the RBD, a protein region eliciting NAbs and targeted by current vaccines. Thus, RBD is a major determinant for viral infectivity and evolution (Boehm *et al.,* 2021; Khan *et al.,* 2021).

It is intriguing that most of the genetic changes found in SARS-CoV-2 can be found in its ancestor virus in bats, demonstrating they took place before spillover to humans and did not occur due to human-to-human transmission (MacLean *et al.,* 2021). However, molecular scrutiny carried out with unprecedented detail since the dawn of COVID pandemic has examined circulating SARS-CoV-2 strains and evidenced that the virus evolves, as expected, but the frequency of mutations in the SARS-CoV-2 genome is lower than for other known RNA viruses, like HIV-1 and IV (*Figure 2*) (Rausch *et al.,* 2020). The higher genetic stability of SARS-CoV-2, and CoVs in general, is due to their RdRp that has a proof-reading activity minimizing base misincorporation during viral replication, limiting SARS-CoV-2 evolutionary rate to approximately 1×10^{-3} substitutions/site/year, (Boehm *et al.,* 2021). Although such relatively high stability of SARS-CoV-2 is good news for vaccine development, the emergence of SARS-CoV-2 variants poses a major challenge for devising measures to counter the virus threat, as new variants continue to emerge. Indeed, some of these are believed to be more infectious than the progenitor virus (Korber *et al.,* 2020; Giovanetti *et al.,* 2021; Groves *et al.,* 2021).

Thus, as for other viruses, zoonotic CoVs co-evolve with their hosts. Viral genetic diversity is selected through host pressure, to which CoVs respond by mutating their genome and with a high frequency of recombination that, at least for some CoVs, can be up to 25% for the entire genome (Snedden *et al.*, 2021). Recombination has not yet been reported for SARS-CoV-2 viral species itself, but *in silico* modeling has predicted high capacity of recombination for this virus too (Banerjee *et al.,* 2020).

A conventional way to infer the type of genetic evolution to which an organism is subjected is to measure the ratio of non-synonymous/synonymous mutations

(dN/dS). dN/dS ratios greater than one, less than one and equal to one indicate positive selection, negative (purifying) selection and neutral evolution, respectively (Edwards *et al.,* 2021, Levin *et al.,* 1999; Kryazhimskiy and Plotkin 2008). Molecular sequencing at the beginning of the pandemic indicated that the virus was rather stable and frequently undergoing purifying selection, i.e., a higher number of synonymous mutations compared to the non-synonymous mutations (Tang *et al.,* 2020; van Dorp *et al.,* 2020). Conversely, the latest reports suggest that SARS-CoV-2 is now under positive selection for changes conferring presumed advantages, such as increased transmission rates (Meng *et al.,* 2021; Volz *et al.,* 2021). Whether acceleration of genetic change is the results of higher rates of individuals with pre-existing immunity and vaccine pressure has not been established (Khan *et al.,* 2021). However, it is generally accepted that antigenic drift is most frequently observed on viral surface proteins that are highly exposed to selection pressure by Nabs (Burns and Desrosiers 1994; Levin *et al.,* 1999; Moya *et al.,* 2000; Domingo-Calap *et al.,* 2019; Shan *et al.,* 2020). As to mutations found in SARS-CoV-2, the virus is no exception, since its S gene, particularly the S1 and RBD coding regions, exhibits the highest non-synonymous mutation rate detected (Boehm *et al.,* 2021; DeLong *et al.,* 2021; Giovanetti *et al.,* 2021). The tendency of the S protein to mutate has been observed across the majority of CoVs (Graham and Baric 2010; Jung and Saif 2015). It is likely, therefore, that continuous circulation of the virus in the face of increasing numbers of vaccinated/immunized people, along with improvement of therapeutic treatment with monoclonal antibodies, will further increase the emergence of SARS-CoV-2 variants. Retrospective research on the evolution of endemic human SARS-CoV-2 may help predict the likely evolutionary trajectory of current pandemic (Singh *et al.,* 2021; Tao *et al.,* 2021).

The emergence of SARS-CoV-2 clades and variants important for public health

The continuous evolution of SARS-C-V-2 leads to the emergence of variants, some of which largely replaced the parental strains and, in turn, gave rise to variant derivatives. Such swarms of virus variants and closely related strains have been clustered into ten clades, at the time of writing: L, V, S, G, GH, GK, GR, GV, GRY and O, named after their most representative mutations (*Figure 4)* (Singh *et al.,* 2021; Snedden *et al.,* 2021). These clades have uneven distribution across geographic areas and the pandemic timeline. To a large extent, clade L, which dominated the beginning of the pandemic, faded out with the emergence of clades O and S in early January 2020. These were, in turn, replaced by clades GH and GR just a few weeks later. Clades GV and GRY appeared in subsequent months and still others emerged in continuous fluctuating waves (Singh *et al.,* 2021). In late 2020, a new clade split from base clade G, forming clade GK (aka Delta variant and Pango lineage B.1.617.2). Variants differed for their fitness and, more relevantly on a public health scale, infectiousness level, virion stability, ability to circumvent mounting population-level immunity and other yet unidentified factors (Lauring and Hodcroft 2021).
Clades contain further distinctive variants identified

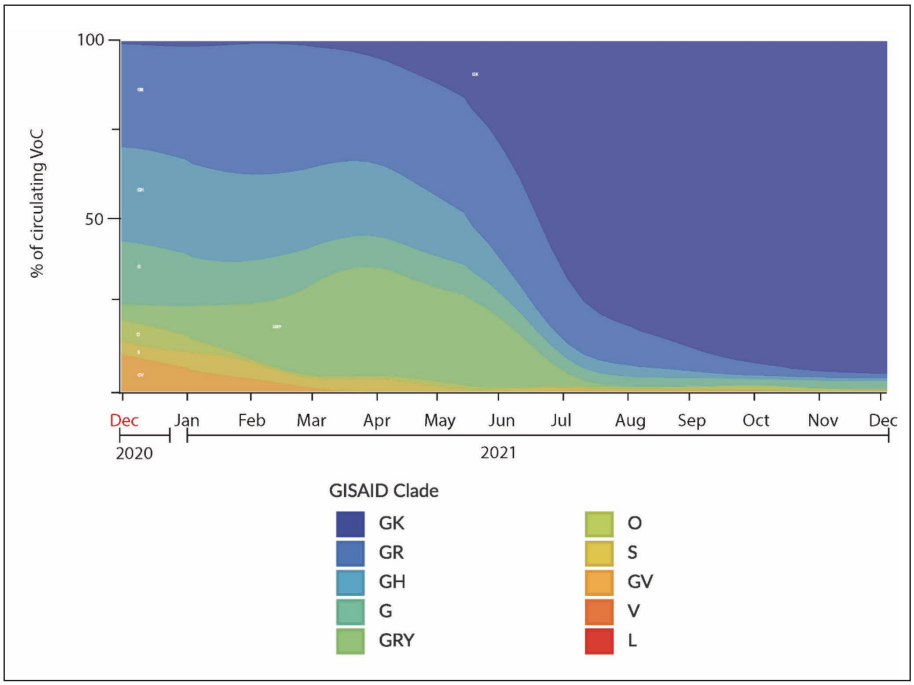

Figure 4 - *Fluctuations of SARS-CoV-2 main clades over time. Changes in prevalence of SARS-CoV-2 clades from December 2020 to December 2021.*

as Variants of Concern (VOC), Variants of Interest (VOI) or Variants Under Monitoring (VUM). This classification was produced by WHO and other relevant international medicine agencies and is presently adopted internationally. VOC implies that a variant deserves full attention, as there is documented evidence indicating significant impact on transmissibility, severity and/or immunity that may influence the epidemiological situation in a geographical area. VOI include variants with genomic properties, epidemiological or *in vitro* evidence that could impact on transmissibility, severity and/or immunity, and on the epidemiological situation in a specific geographical area. However, the evidence is still preliminary or is associated with major uncertainty. Finally, VUM are defined through epidemic intelligence, rule-based genomic variant screening, or preliminary scientific evidence. VUM could have properties of a VOC, but the evidence is weak or has not yet been fully assessed. Segregation in given categories, however, is largely dependent on local factors, i.e. a VUM in areas under development with low vaccination rates is as worrisome as a *bona fide* VOC. Continuous monitoring of variants is pivotal because not all variants necessarily arise thanks to increased viral fitness: when a small number of individuals infected by a lineage account for most of the virus's early spread in specific areas, this may be the only reason underlying the lineage's takeover, giving rise to the so-called 'founder effect' (Callaway 2020; Lauring and Hodcroft 2021; Volz *et al.,* 2021), as schematically illustrated in *Figure 1*.

The most widespread variants are enlisted in *Table 1*. The majority belongs to the most recently defined clades and may harbor either shared or distinctive signature mutations. The emergence of "official" variants was preceded by progressive evolution of SARS-CoV-2, most likely due to adaptation of the virus to the human host. Certain mutations, believed to facilitate transmission and replication in humans, appeared since the beginning of the COVID pandemic and underwent strong positive selection because of the gain of fitness conveyed; these replaced the original sequences in a matter of weeks. Among these, the first to appear was the non-synonymous A to G mutation localized at the nucleotide position 23403 of SARS-CoV-2 genome. It results in the amino acid mutation D614G within subdomain 2 of the S1 gene (Korber *et al.,* 2020; Zhang *et al.,* 2021). Mutants bearing this amino acid change have been detected since the early months of the pandemic and were most likely pivotal for the rapid and successful growth of clade G, now detected globally. It is believed to increase SARS-CoV-2 infectivity possibly due to increased S openness that renders the protein more easily activated by cellular proteases (Eaaswarkhanth *et al.,* 2020; Storti *et al.,* 2021). It has also been postulated that the high replicative capacity *in vivo* and the resulting cytopathic damage of the olfactory epithelium is responsible for anosmia (loss of smell), commonly reported as a symptom since the emergence of D614G mutations (Butowt *et al.,* 2020). Other reports also suggested that the structural modification in the viral receptor also increased susceptibility to NAbs, but this observation was not confirmed by subsequent reports (Eaaswarkhanth *et al.,* 2020; Korber *et al.,* 2020). Others also question its impact on disease severity, because D614G is often accompanied by other mutations inside and outside the S protein (Dao *et al.,* 2021). Among these, the P323L mutation is particularly interesting because it changes RdRp conformation. Position 323 is placed outside the RdRp catalytic site and is believed to interact with cellular and viral proteins aiding viral replication, possibly modulating its proof-reading activity (Kirchdoerfer and Ward 2019; Pachetti *et al.,*

Table 1 - *List of the most representative VOCs. Summary of the reported SARS-CoV-2 variants divided by clade, first detection and Spike mutations.*

	Alpha	*Beta*	*Gamma*	*Delta*	*Omicron*
Scientific name	Pango: B.1.1.7 PHE: VOC-202012/01 Nextstrain: 20I/B 501Y.V1 GISAID: GRY, GR/501Y.V1	Pango: B.1.351 PHE: VOC-202012/02 Nextstrain: 20H/501Y.V2 GISAID: GH/501Y.V2	Pango: P.1 PHE: VOC-202101/02 Nextstrain: 20J/501Y.V3 GISAID: GR/501Y.V3	Pango: B.1.617.2 PHE: VOC-21APR-02 Nextstrain: 21A/S:478K GISAID: G/452R.V3	Pango: B.1.1.529 Nextstrain: 21K GISAID: GR/484A
First detection	September 2020, UK	May 2020, South Africa	November 2020, Brazil	October 2020, India	November 2021, South Africa
Spike mutations	Del 69/70, Del 144, N501Y, A570D, D614G, T716I, P681H, S982A D1118H	L18F, D80A, D215G, Del 241/243, 242-244 del, R246I, K417N, E484K, N501Y, D614G, and A701V	L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, V1176F	T19R, Del 157/158, L452R, T478K, D614G, P681R, D950N	A67V, Del 69/70, T951, G142D, Del 143/145, Del 211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F

2020). Both D614G and P323L are undergoing purifying selection, thereby increasing their frequency and co-presence across viral strains of different clades. It is thus conceivable that both mutations enhance viral fitness and, together with other mutations, are key factors in contributing to the rapid spread of clade G and its derivatives GH, GR, GV, etc. Other important mutations, L37F and G251L in Nsp6 and ORF3a proteins, at first, were thought to reduce viral fitness but, instead, they turned out to contribute to SARS-CoV-2 dissemination by increasing asymptomatic cases. Nsp6 is relatively conserved in other CoVs (DeLong *et al.,* 2021; Giovanetti *et al.,* 2021) and has been shown to inhibit interferon (IFN) type 1 responses and reduce autophagic flux (Xia *et al.,* 2020). L37F mutation has been also observed in MERS-CoV-2 (van Dorp *et al.,* 2020) and is believed to reduce stability of Nsp6 (Bakhshandeh *et al.,* 2021); similarly, G251L, a mutation frequently observed in clade V, has been shown to reduce activity of ORF3a viroporin. The latter has been reported to activate the inflammasome, thus contributing to cytokine storm and necrotic cell death and promote viral release (Siu *et al.,* 2019). As opposed to G251L, mutation Q57H is also present in ORF3a and in strains of clade GH, where it has been shown to promote viral replication (Bakhshandeh *et al.,* 2021). L84S is another mutation of interest that lies within ORF8, a protein implicated in evasion of host immune responses which is likely to facilitate zoonotic transmission and adaptation to novel hosts (Flower *et al.,* 2021). L84S may be important in modulating SARS-CoV-2 virulence and pathogenesis but its importance is questioned, as its prevalence is declining. Mutations in the nucleocapsid (N) protein are considered important for its role in viral replication, virion stability and capacity to antagonize type-I and -II IFN responses (Peng *et al.,* 2020; Ramasamy and Subbian 2021). Further, mutations in this protein are a matter of concern for possible repercussions in the diagnosis of viral infection (Zhou *et al.,* 2021). Indeed, nearly all antigenic tests are designed to detect the N protein that, being present in multiple copies within each viral particle, reduces the gap of sensitivity between antigenic and molecular assays (Safiabadi Tali *et al.,* 2021). Compared to other proteins, N is relatively stable, particularly in its RNA-binding and dimerization domains. Most mutations were found between region type 2 and the linker region (Peng *et al.,* 2020). The most prevalent ones are R203K/G204R, localized in these two domains, in particular G204R is a signature of clade GR (Giovanetti *et al.,* 2021; Singh *et al.,* 2021). These mutations are thought to increase intraviral protein binding affinity and virion assembly (Peng *et al.,* 2020). In addition, because of the rapid expansion of clades bearing these mutations, they are thought to confer immune evasion properties (Xia *et al.,* 2020; Ramasamy and Subbian 2021). At the time of writing, none of the reported mutations has significantly affected detection by available antigen diagnostic tests (Gandolfo *et al.,* 2021; Lai *et al.*, manuscript in preparation).

As shown in *Table 1*, classified variants are mostly defined by mutations in the S gene and, particularly, the RBD region for its chief role in eliciting protective immunity. Accordingly, the RBD is actively targeted by the host antibody response and rapidly evolving (Piccoli *et al.,* 2020; Singh *et al.,* 2021). Most VOCs, identifying SARS-CoV-2 variants associated with greater transmissibility, altered virulence, or the ability to escape natural infection- and vaccine-mediated immunity, contain one or more mutations in RBD (*Table 1*). Some of these mutations are shared among the variants suggesting their early emergence and fitness gain. N501Y, for instance, affects one of the six critical S residues binding to ACE2 and its mutation is believed to increase viral infectivity through stronger interactions with two ACE2 amino acids (Zhang *et al.,* 2021). Although detected occasionally at present, mutations involving other critical S residues within the RBD (namely L455, F486, Q493, S494, and Y505) should be closely monitored as they may also facilitate S-human ACE2 binding (van Dorp *et al.,* 2020). N501Y is often accompanied by other mutations.

The highly transmissible B.1.1.7 variant, aka alpha variant, rapidly emerged over one year ago in Southern England, and was characterized by higher rates of non-synonymous mutations and deletions than expected from the mutation rate of SARS-CoV-2 (*Figure 2*). In addition to N501Y and the above mentioned D614G, VOC alpha contained three amino acid deletions, the two consecutive amino acid positions 69 and 70 and another downstream (144), and P681H mutation, as listed in *Table 1*. The swift escalation of transmission in the general population clearly demonstrates that this rearrangement provided a replicative advantage and increased transmissibility. It is not clear whether these changes confer any resistance to antibody-mediated neutralization and monoclonal antibody treatment (Falcone *et al.,* 2021; Singh *et al.,* 2021; Tao *et al.,* 2021), but it has been recently shown that this variant enters faster into cells *in vitro* compared to its progenitor strain (Storti *et al.,* 2021), whereas 681 mutations may facilitate S1/S2 cleavage by furin which, in turn, enhances infectivity (Li *et al.,* 2020; Peacock *et al.,* 2021). P681H is also within an antigenic epitope recognized by B and T lymphocytes, implicating host immune response alterations (Piccoli *et al.,* 2020). Recent molecular surveys show that B.1.1.7 is likely evolving in a 2.0 version by stably acquiring the RDB mutation E484K. This mutation is of interest as it has been associated with antibody resistance (Wise 2021). Other B.1.1.7 mutations located outside S could also have contributed to its remarkably fast

spread, but there are no firm data. Recently, the alpha variant has been de-escalated and is no longer considered VOC because its circulation drastically dropped in recent months and there is little evidence on its impact on vaccine immunity (ECDC bulletin Nov 04, 2021).

Another important VOC is B.1.351, detected in South Africa at the end of 2020 and also known as beta variant (*Table 1*). This variant contains nine non-synonymous mutations in S gene and shares a few mutations with the alpha variant. Three mutations are within the RBD (K417N, E484K, N501Y) in critical residues interacting with hACE2. In addition, E484K is an important recognition site for Nabs (Barnes *et al.,* 2020; Piccoli *et al.,* 2020); consequently, E484K confers some degree of resistance to antibody-mediated neutralization of SARS-CoV-2 *in vitro* and *in vivo* (Falcone *et al.,* 2021; Tao *et al.,* 2021). The increasing presence of E484K in multiple independent SARS-CoV-2 lineages suggests that this mutation does contribute to immune escape (Tao *et al.,* 2021). Mutation K417N is recognized by NAbs and *in silico* studies hypothesize that it may impact on RBD-hACE2 binding affinity and stabilize E484K (Fratev 2020). Supporting these results, the beta variant has documented reduced sensitivity to NAbs elicited by natural infection by other strains, vaccination, and monoclonal antibodies (Falcone *et al.,* 2021; Fiolet *et al.,* 2021).

The P.1 variant, aka gamma variant, contains the same RBD mutations as the beta variant (K417T, E484K, N501Y). It is likely to have originated in Brazil at the end of 2020 and has since spread to other countries (*Table 1*). The origin of P.1 is not clear; however, its apparently independent evolution from B.1.351 (beta) variant (Tao *et al.,* 2021) and yet the striking similarity observed in their RBD domains indicate that the above three mutations confer an important advantage as regards circumventing antibody-mediated neutralization and enhanced transmissibility. Their emergence, therefore, suggests that the respective progenitor strains underwent convergent evolution. Further supporting this hypothesis is a small deletion in SARS-CoV-2 Nsp6 gene, which is present in both variants, with a thus far unknown functional significance. The same geographical area was swept by another variant, called P.2, whose origin is uncertain and, compared to P.1, retained only E484K (Tao *et al.,* 2021).

Variant B.1.617.2 or delta is, at the moment, the most successful VOC, as it has spread with remarkable speed all over the world, accounting for over 95% new cases in Italy. This variant was first identified in India in late 2020 and contains L452R, T478K, and P681R, along with the D614G mutation within the S protein (*Table 1*). A mutation at position 452 of the S protein from an uncharged, hydrophobic leucine (L) residue into a positively charged, hydrophilic arginine (R) residue could increase electrostatic interaction with the negatively charged amino acids of ACE-2 binding site and promote viral infectivity and replication (Li *et al.,* 2020). Since the region around position 452 is also immunogenic, it has been postulated that this amino acid change may also hamper antibody-mediated neutralization and cellular immune recognition (Li *et al.,* 2020). Similarly to the S477N mutation, T478K is thought to increase S-ACE-2 binding affinity and, being within a neutralizing epitope that also contains E484K/Q, it may contribute to immune escape as well (Tao *et al.,* 2021). The role of L452R is uncertain but it does enhance resistance to NAbs, when found in combination with T478K (Wall *et al.,* 2021). Why the delta variant surged so rapidly and practically wiped out all circulating strains is unclear. It has been shown that this VOC has increased replication efficiency in the human airway system relative to its ancestral strain; moreover, contemporary variants show enhanced S1/S2 cleavage, similarly to the alpha variant sharing the P681R mutation (Peacock *et al.,* 2021). All things considered, the delta variant is a VOC showing NAb resistance comparable to the beta variant but transmissibility higher than the alpha variant (Tao *et al.,* 2021). Possibly due to its continuous spread, even among vaccinated individuals (Fiolet *et al.,* 2021), the delta variant is currently evolving. It seems to be undergoing purifying selection toward acquisition of a further mutation K417N, also localized in the S gene where it is known to alter susceptibility to antibody-mediated neutralization and increase transmissibility (Fiolet *et al.,* 2021; Tao *et al.,* 2021).

Among VOCs, the most recent addition is B.1.1529, better known as omicron variant. This variant bears over 30 amino acid changes, three small deletions and one small insertion in the S protein probably acquired by recombination (Callaway 2021). Nine mutations are localized in RBD. This variant was first detected in mid-November in Botswana and, by the end of the month, was found in several European countries and Israel. Because it is the most divergent variant detected so far in a significant number of countries, the omicron variant has been immediately classified as VOC for its potentially increased transmissibility and resistance to vaccine immunity, and increased risk for reinfections (ECDC Bulletin Nov 26, 2021).

In addition to the above mentioned VOCs, there is a growing number of VOI and VUM that has raised attention because of their scattered increasing prevalence and reduced sensitivity to vaccine-mediated immunity. For SARS-CoV-2, evolution is clearly funneled by two main factors: social distance and public health containment measures. The first privileges mutations enhancing host transmission - both from human to humans and to other species, for instance pets(Parkhe and Verma 2021); vaccination, on the

other hand, selects for mutations conferring ability to reduce sensitivity to Nabs and escape elicited immunity. Given that all vaccines immunize against the S protein, it is not surprising that several mutations localized within S or RBD are slowly emerging and positively selected among many other possible sequence changes. Also, it is likely that mutations associated to resistance to immunity promote the spread of infection, either *per se* or with the aid of other co-mutations. Indeed, in addition to the above mentioned mutations, it is worth mentioning N439K E484K, Q677H, and F888L and other mutations believed to confer some degree of NAb evasion (Fiolet *et al.,* 2021; Tao *et al.,* 2021).

Other variants are under close monitoring. B.1.525, also known as Nigerian variant, surfaced in this country in December 2020, then it was found in the United Kingdom and has since spread internationally (Tao *et al.,* 2021). Initially, its dissemination was fast but then it progressively slowed down concomitantly with steep escalation of the alpha variant. The coincidence of the two events suggests that the alpha outcompeted the Nigerian variant for unknown reasons. The Nigerian variant is nevertheless lurking around and produces small outbreaks in restricted geographical areas. It shares the 69-70del observed in the alpha variant together with E484K, Q677H, and F888L. Based on observations on MERS-CoV-2 and related CoVs, it is believed that Q677P/H and F888L affect S protein cleavability and host cellular entry, respectively (Forni *et al.,* 2015). Variant B.1.526 from New York contains S mutations D253G, D614G, and A701V, along with either E484K or S477N. Mutations A701V and D253G may impact SARS-CoV-2 cleavability and antibody-mediated neutralization, respectively (Tao *et al.,* 2021). Other mutations, such as S477N, D614G and E484K are shared with multiple other variants and are likely to play a role in B.1.526 expansion. A.23.1 variant that emerged in Uganda and spread through the country *via* land, by means of trucks crossing the country and exporting the variant to countries nearby. In fact, other variants were mainly observed in clusters close to international airports or near the main ports of entry of the country (Bugembe *et al.,* 2021). A.23.1 has had limited diffusion in other countries, including Europe, where it was mainly detected in immigrants (Massimo *et al.,* 2021) and generated small clusters. Interestingly, despite its very low prevalence, this variant was found in healthcare workers who had been vaccinated or had recovered from infection within the previous three months (Pistello *et al.*, manuscript in preparation). This variant was able to diversify into a new variant, the UK A.23.1 sub-lineage VUI-202102/01 containing the additional immune escape mutation E484K (Tao *et al.,* 2021). Ongoing molecular sequencing surveys evidence that new or VOC-derived sub-variants are emerging and nearly all contain a combination of the S mutations mentioned above. This suggests that, at least for the S gene, these mutations target crucial positions as regards sensitivity to immune response and efficiency of viral entry into respiratory epithelial cells. Possible exceptions are R.1 variant from Japan, which contains potential immune escape mutations W152L and E484K (Hirotsu and Omata 2021), and certain emerging lineages from North and South America. These lineages bear S13I, W152C, and L452R/Q mutations, also contributing to resistance to antibody-mediated neutralization (McCallum *et al.,* 2021). Interestingly, variant C.37 originating from Peru also shares the Nsp6 deletion found in P.1 and other variants. The presence of mutations conserved across variants could be due to dissemination of a limited number of strains that underwent independent evolution in different geographical areas. Alternatively, they may indicate convergent evolution of diverse strains. Monitoring these variants for transmission, expansion and evolution should shed light on this matter and may assist in predicting long-term efficacy of the current vaccines or assist in the design of novel vaccination strategies.

CONCLUSIONS

SARS-CoV-2 infection is so widespread globally that it will most likely become endemic in the human population, as observed for the other zoonotic HCoVs, NL63, OC43, HKU1, and 229E (Morens and Fauci 2020; Bakhshandeh *et al.,* 2021; Boehm *et al.,* 2021; Khan *et al.,* 2021). Further, as inferred from the epidemic waves that followed since the COVID pandemic start, SARS-CoV-2 circulation recalls endemic HCoVs that cause seasonal outbreaks during the winter in temperate regions (Audi *et al.,* 2020). Cold temperatures stabilize enveloped virions, increasing their stability outside the host. In addition, low temperatures allow viruses to remain suspended in the air longer, enhancing aerosol transmission of respiratory viruses (Harper 1961; Polozov *et al.,* 2008). Chances of infection are further increased by the immunosuppressive effects of cold temperatures and dry environments on a potential host. Current ongoing vaccination, together with the spread of infection, will progressively create an immune population where SARS-CoV-2 circulation is likely to decrease, especially in warmer climates.

How SARS-CoV-2 will evolve from here is uncertain. The current SARS-CoV-2 pandemic is fostered by asymptomatic, pre-symptomatic, or otherwise unrecognized cases (Gandhi *et al.,* 2020; Morens and Fauci 2020). This variety of COVID cases, combined with mounting population-level immunity, will probably create the conditions for a reduction in pathogenicity. The virus is likely to continue to evolve and adapt to the human population, while the emergence

of novel variants may fix those mutations conferring increased transmissibility and resistance to host immunity. On the other hand, mutations that will attenuate pathogenicity of the present SARS-CoV-2 will also be selected: hints in this direction are being observed by the presence of scattered and recurring mutations, e.g., P323L, L37F, G251V, and Q27stop, that are speculated to reduce disease severity. Therefore, like for other zoonotic viruses, it is expected that SARS-CoV-2 will progressively adapt and become less pathogenic in humans (Burns and Desrosiers 1994; Domingo-Calap *et al.,* 2019; DeLong *et al.,* 2021; Edwards *et al.,* 2021).

As seen for influenza, natural and acquired immunity is likely to become the main driver shaping the progressive adaptation of SARS-CoV-2. Growing evidence in the field points to the fact that a third vaccine dose is required to reinforce protection, implying that elicited SARS-CoV-2 immunity, induced either by natural infection or vaccination, wanes with time (Röltgen and Boyd 2021). How fast immunity declines below sufficient protection levels is not yet clear but, using again endemic hCoVs as a reference, protection from SARS-CoV-2 infection or reinfection is not likely to extend beyond six-twelve months, again depending on circulating variants and their inherent resistance to specific immunity (Fiolet *et al.,* 2021; Milne *et al.,* 2021; Townsend *et al.,* 2021). Several studies have demonstrated that reinfection with endemic hCoVs is common within 80 days to one year from prior infection (Edridge *et al.,* 2020; Rees and Waterlow 2021). Reinfections are usually mild or go unnoticed, and multiple exposures may be necessary to maintain good levels of protection (Hamady *et al.,* 2021; Milne *et al.,* 2021).

Endemic circulation in a population, either naïve or with suboptimal protection from infection, increases the possibility for a virus to generate variants progressively more resistant to specific immunity and adapted to humans. It is unclear whether co-circulation of HCoVs shapes the evolving trajectory of SARS-CoV-2. There is evidence of some degree of cross-protection across the same genus of HCoVs, but not between genera. Although theoretically possible for their phylogenetic vicinity, it is unclear whether SARS-CoV-2 infection may boost immunity against other beta-CoVs, such as HCoV-OC43, SARS-CoV and MERS-CoV (Crowley *et al.,* 2021; Hsieh *et al.,* 2021). Conversely, there is no indication that antibody-dependent enhancement from prior SARS-CoV-2 or other HCoVs infection or vaccination plays a role in SARS-CoV-2 dissemination and disease severity (Lee *et al.,* 2020; Wen *et al.,* 2020). Therefore, it is unlikely that cross-reactive antibodies generated against endemic HCoVs will steer SARS-CoV-2 evolution. However, sudden and unpredictable changes of direction may derive from recombination of SARS-CoV-2 with other HCoVs. This possibility is not so remote, as inferred by the high frequency of recombination observed during co-infections by CoVs including SARS-CoV-2 (Goldstein *et al.,* 2021; Parkhe and Verma 2021).

Continuous monitoring of future emerging SARS-CoV-2 variants and our rapidly expanding knowledge regarding the effect of SARS- CoV-2 S mutations on antigenicity and other aspects of virus biology will help tracking the changes flagged as potentially significant. This will guide the implementation of targeted control measures and drive the development of tests for further laboratory characterization. The integration of these data will facilitate early detection of potential VOCs, i.e., before they have spread widely, and assist in the design of updated vaccines. Future generation vaccines should be tailored to emerging antigenic variants and maximally cross-reactive against all circulating variants. Monitoring the evolution of SARS-CoV-2 will also contribute to our understanding of the molecular mechanism leading to cross-species viral transmission and subsequent adaptation to novel hosts. These goals will be achieved through close and continuous international collaboration and rapid and open sharing of data.

Acknowledgements

This work was supported by bando PRIN: Progetti di ricerca di rilevante interesse nazionale, Bando 2017, Prot. 2017KM79NN: "Addressing viral neuropathogenesis: Unraveling the molecular and cellular pathways of viral replication and host cell response and paving the way for the development novel host-targeted, broad spectrum, antiviral agents"; bando Ricerca COVID-19 Toscana 2020 "Suppression of Airborne Viral Epidemic Spread by Ultraviolet light barriers (SAVES-US)"; POR-FESR 2014-2020 Regione Toscana " from TUScany and its Cultivations new Anti-VIral and Antibacterial substances (TUSCAVIA)"; POR FESR Regione Lombardia 2014-2020: "PAN-ANTI COVID 19: Produzione librerie di anticorpi neutralizzanti ad uso terapeutico da pazienti COVID-19"

References

Ascough S., Paterson S., Chiu C. (2018). Induction and Subversion of Human Protective Immunity: Contrasting Influenza and Respiratory Syncytial Virus. *Front Immunol.* **9**, 323.

Audi A., AlIbrahim M., Kaddoura M., Hijazi G., Yassine H.M., et al. (2020). Seasonality of respiratory viral infections: will COVID-19 follow suit? *Frontiers in Public Health.* **8**, 576.

Bakhshandeh B., Jahanafrooz Z., Abbasi A., Goli M.B., Sadeghi M., et al. (2021). Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus. *Microbial Pathogenesis*. 104831.

Bandera A., Gori A., Clerici M., Sironi M. (2019). Phylogenies in ART: HIV reservoirs, HIV latency and drug resistance. *Current opinion in pharmacology.* **48**, 24-32.

Banerjee A., Doxey A.C., Tremblay B.J.-M., Mansfield M.J., Subudhi S., et al. (2020). Predicting the recombination potential of severe acute respiratory syndrome coronavirus 2 and Middle East respiratory syndrome coronavirus. *The Journal of General Virology*- **101**, 1251.

Barnes C.O., Jette C.A., Abernathy M.E., Dam K.A., Esswein S.R., et al. (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*. **588**, 682-687.

Bentley K., Evans D.J. (2018). Mechanisms and consequences of positive-strand RNA virus recombination. *Journal of General Virology*. **99**, 1345-1356.

Boehm E., Kronig I., Neher R., Eckerle I., Vetter P., et al. (2021). Variants: the Pandemics within the Pandemic. *Clinical Microbiology and Infection, Elsevier BV*.

Bugembe D.L., Phan M.V., Ssewanyana I., Semanda P., Nansumba H., et al. (2021). Emergence and spread of a SARS-CoV-2 lineage A variant (A. 23.1) with altered spike protein in Uganda. *Nature Microbiology*. 1-8.

Burns D., Desrosiers R. (1994). Envelope sequence variation, neutralizing antibodies, and primate lentivirus persistence. *Simian Immunodeficiency Virus*. 185-219.

Butowt R, Bilinska K., Von Bartheld C.S. (2020). Chemosensory dysfunction in COVID-19: integration of genetic and epidemiological data points to D614G spike protein variant as a contributing factor. *ACS chemical neuroscience*. **11**, 3180-3184.

Calisher C.H., Childs J.E., Field H.E., Holmes K.V., Schountz T. (2006). Bats: important reservoir hosts of emerging viruses. *Clinical microbiology reviews*. **19**, 531-545.

Callaway E. (2020). The coronavirus is mutating--does it matter? *Nature*. **585**, 174-178.

Callaway E. (2021). Beyond Omicron: what's next for COVID's viral evolution. *Nature*. **600**, 204-207.

Chotiyaputta W., Lok A.S. (2009). Hepatitis B virus variants." *Nature reviews Gastroenterology & hepatology* **6**: 453-462.

Chou, S (2020). "Advances in the genotypic diagnosis of cytomegalovirus antiviral drug resistance. *Antiviral research*. **176**, 104711.

Ciminski K., Chase G.P., Beer M., Schwemmle M. (2021). Influenza a viruses: understanding human host determinants. *Trends in Molecular Medicine*. **27**, 104-112.

Crowley A.R., Natarajan H., Hederman A.P., Bobak C.A., Weiner J.A., et al. (2021). Boosting of Cross-Reactive Antibodies to Endemic Coronaviruses by SARS-CoV-2 Infection but not Vaccination with Stabilized Spike. *Medrxiv*.

Dao T.L., Hoang V.T., Colson P., Lagier J.C., Million M., et al. (2021). SARS-CoV-2 Infectivity and Severity of COVID-19 According to SARS-CoV-2 Variants: Current Evidence. *Journal of Clinical Medicine*. **10**, 2635.

DeLong J.P., Al-Sammak M.A., Al-Ameeli Z.T., Dunigan D.D., Edwards K.F., et al. (2021). Towards an integrative view of virus phenotypes. *Nature Reviews Microbiology*. 1-12.

Domingo-Calap P., Segredo-Otero E., Durán-Moreno M., Sanjuán R. (2019). Social evolution of innate immunity evasion in a virus. *Nature microbiology*. **4**, 1006-1013.

Dustin L.B., Bartolini B., Capobianchi M.R., Pistello M. (2016). Hepatitis C virus: life cycle in cells, infection and host response, and analysis of molecular markers influencing the outcome of infection and response to therapy. *Clinical microbiology and infection*. **22**, 826-832.

Eaaswarkhanth M., Al Madhoun A., Al-Mulla F. (2020). Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? *International Journal of Infectious Diseases*. **96**, 459-460.

Edridge A.W., Kaczorowska J., Hoste A.C., Bakker M., Klein M., et al. (2020). Seasonal coronavirus protective immunity is short-lasting. *Nature medicine*. **26**, 1691-1693.

Edwards K.F., Steward G.F., Schvarcz C.R. (2021). Making sense of virus size and the tradeoffs shaping viral fitness. *Ecology Letters*. **24**, 363-373.

Falcone M., Tiseo G., Valoriani B., Barbieri C., Occhineri S., et al. (2021). Efficacy of bamlanivimab/etesevimab and casirivimab/imdevimab in preventing progression to severe covid-19 and role of variants of concern. *Infectious diseases and therapy*. **10**, 2479-2488.

Fiolet T., Kherabi Y., MacDonald C.-J., Ghosn J., Peiffer-Smadja N. (2021). Comparing COVID-19 vaccines for their characteristics, efficacy and effectiveness against SARS-CoV-2 and variants of concern: A narrative review. *Clinical Microbiology and Infection*.

Flower T.G., Buffalo C.Z., Hooy R.M., Allaire M., Ren X., et al. (2021). Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proceedings of the National Academy of Sciences*. **118**.

Forni D., Filippi G., Cagliani R., De Gioia L., Pozzoli U., et al. (2015). The heptad repeat region is a major selection target in MERS-CoV and related coronaviruses. *Scientific reports*. **5**, 1-10.

Fratev F. (2020). The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: A Free energy of perturbation study. *BioRxiv*.

Gandhi M., Yokoe D.S., Havlir D.V. (2020). Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19, Mass Medical Soc.

Gandolfo C., Morecchiato F., Pistello M., Rossolini G.M., Cusi M.G. (2021). Detection of SARS-CoV-2 N protein allelic variants by rapid high-throughput CLEIA antigen assay. *Journal of Clinical Virology*.

Giovanetti M., Benedetti F., Campisi G., Ciccozzi A., Fabris S., et al. (2021). Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. *Biochemical and Biophysical Research Communications*. **538**, 88-91.

Goldstein S.A., Brown J., Pedersen B.S., Quinlan A.R., Elde N.C. (2021). Extensive recombination-driven coronavirus diversification expands the pool of potential pandemic pathogens. *bioRxiv*.

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

Groves D.C., Rowland-Jones S.L., Angyal A. (2021). The D614G mutations in the SARS-CoV-2 spike protein: Implications for viral infectivity, disease severity and vaccine design. *Biochemical and biophysical research communications*. **538**, 104-107.

Haage V., de Oliveira-Filho E.F., Moreira-Soto A., Kühne A., Fischer C., et al. (2021). Impaired performance of SARS-CoV-2 antigen-detecting rapid tests at elevated and low temperatures. *Journal of Clinical Virology*.

Hamady A., Lee J., Loboda Z.A. (2021). Waning antibody responses in COVID-19: what can we learn from the analysis of other coronaviruses? *Infection*. 1-15.

Harper G. (1961). Airborne micro-organisms: survival tests with four viruses. *Epidemiology & Infection*. **59**, 479-486.

Hirotsu Y., Omata M. (2021). Detection of R. 1 lineage severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with spike protein W152L/E484K/G769V mutations in Japan. *PLoS pathogens*. **17**, e1009619.

Hsieh C.-L., Werner A.P., Leist S.R., Stevens L.J., Falconer E., et al. (2021). "tabilized coronavirus spike stem elicits a broadly protective antibody. *Cell reports* **37**: 109929.

Jarrett O. (1992). Pathogenicity of feline leukemia virus is commonly associated with variant viruses. *Leukemia*. **6**, 153S-154S.

Jung K., Saif L.J. (2015). Porcine epidemic diarrhea virus infection: etiology, epidemiology, pathogenesis and immunoprophylaxis. *The Veterinary Journal*. **204**, 134-143.

Khan W.H., Hashmi Z., Goel A., Ahmad R., Gupta K., et al. (2021). COVID-19 Pandemic and Vaccines Update on Challenges and Resolutions. *Frontiers in Cellular and Infection Microbiology*. **11**.

Kirchdoerfer R.N., Ward A.B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications*. **10**, 1-9.

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. e819.

Kryazhimskiy S., Plotkin J.B. (2008). The population genetics of dN/dS. *PLoS genetics*. **4**, e1000304.

Lauring A.S., Hodcroft E.B. (2021). Genetic variants of SARS-CoV-2 - what do they mean? *Jama*. **325**, 529-531.

Lee W.S., Wheatley A.K., Kent S.J., DeKosky B.J. (2020). Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nature microbiology*. **5**, 1185-1191.

Levin B.R., Lipsitch M., Bonhoeffer S. (1999). Population biology, evolution, and infectious disease: convergence and synthesis. *Science*. **283**, 806-809.

Li Q., Wu J,. Nie J., Zhang L., Hao H., et al. (2020). The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*. **182**, 1284-1294. e1289.

Liu R., Sheng Z., Huang C., Wang D., Li F. (2020). Influenza D virus. *Current Opinion in Virology*. **44**, 154-161.

Lu R., Zhao X., Li J., Niu P., Yang B., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The lancet* **395**: 565-574.

Lurain N.S., Chou S. (2010). Antiviral drug resistance of human cytomegalovirus. *Clinical microbiology reviews*. **23**, 689-712.

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.

Mansky L.M. (2002). HIV mutagenesis and the evolution of antiretroviral drug resistance. *Drug resistance updates*. **5**, 219-223.

Massimo M.C., Fabio T., Di Naro Daniela R.G., Paola S., Claudia M., et al. (2021). First detection of SARS-CoV-2 A. 23.1 sub-lineage in migrants arriving to Italy via the Mediterranean Sea and public health implications. *Travel medicine and infectious disease*.

McCallum M., Bassi J., De Marco A., Chen A., Walls A.C., et al. (2021). SARS-CoV-2 immune evasion by variant B. 1.427/B. 1.429. *BioRxiv*.

Meng B., Kemp S.A., Papa G., Datir R., Ferreira I.A., et al. (2021). Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B. 1.1. 7. *Cell reports*. **35**, 109292.

Metsky H.C., Matranga C.B., Wohl S., Schaffner S.F., Freije C.A., et al. (2017). Zika virus evolution and spread in the Americas. *Nature*. **546**, 411-415.

Milne G., Hames T., Scotton C., Gent N., Johnsen A., et al. (2021). Does infection with or vaccination against SARS-CoV-2 lead to lasting immunity? *The Lancet Respiratory Medicine*.

Morens D.M., Fauci A.S. (2020). Emerging pandemic diseases: how we got to COVID-19. *Cell*.

Moya A., Elena S.F., Bracho A., Miralles R., Barrio E. (2000). The evolution of RNA viruses: a population genetics view. *Proceedings of the National Academy of Sciences*. **97**, 6967-6973.

Pachetti M., Marini B., Benedetti F., Giudici F., Mauro E., et al. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of translational medicine*. **18**, 1-9.

Parkhe P., Verma S. (2021). Evolution, Interspecies Transmission, and Zoonotic Significance of Animal Coronaviruses. *Frontiers in Veterinary Science*. 1149.

Peacock T.P., Goldhill D.H., Zhou J., Baillon L., Frise R., et al. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nature Microbiology*. 1-11.

Peck K.M., Lauring A.S. (2018). Complexities of viral mutation rates. *Journal of virology*. **92**, e01031-01017.

Peng Y., Du N., Lei Y., Dorje S., Qi J., et al. (2020). "tructures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *The EMBO journal*. **39**, e105938.

Petrovic D., Dempsey E., Doherty D.G., Kelleher D., Long A. (2012). Hepatitis C virus–T-cell responses and viral escape mutations. *European journal of immunology*. **42**, 17-26.

Piccoli L., Park Y.-J., Tortorici M.A., Czudnochowski N., Walls A.C., et al. (2020). Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell*. **183**, 1024-1042. e1021.

Polozov I.V., Bezrukov L., Gawrisch K., Zimmerberg J. (2008). Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nature chemical biology*. **4**, 248-255.

Ramasamy S., Subbian S. (2021). Critical Determinants of Cytokine Storm and Type I Interferon Response in COVID-19 Pathogenesis. *Clinical microbiology reviews*. **34**, e00299-00220.

Rausch J.W., Capoferri A.A., Katusiime M.G., Patro S.C., Kearney M.F. (2020). Low genetic diversity may be an Achilles heel of SARS-CoV-2. *Proceedings of the National Academy of Sciences*. **117**, 24614-24616.

Rees E.M., Waterlow N.R. (2021). Estimating the duration of seropositivity of human seasonal coronaviruses using seroprevalence studies. *Wellcome Open Research*. **6**.

Röltgen K., Boyd S.D. (2021). Antibody and B cell responses to SARS-CoV-2 infection and vaccination. *Cell Host & Microbe*.

Safiabadi Tali S.H., LeBlanc J.J., Sadiq Z., Oyewunmi O.D., Camargo C., et al. (2021). Tools and Techniques for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)/COVID-19 Detection. *Clinical Microbiology Reviews*. **34**, e00228-00220.

Scourfield A., Waters L., Nelson M. (2011). Drug combinations for HIV: what's new? *Expert review of anti-infective therapy*. **9**, 1001-1011.

Shan C., Xia H., Haller S.L., Azar S.R., Liu Y., et al. (2020). A Zika virus envelope mutation preceding the 2015 epidemic enhances virulence and fitness for transmission. *Proceedings of the National Academy of Sciences*. **117**, 20190-20197.

Sharp P.M., Hahn B.H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harbor perspectives in medicine*. **1**, a006841.

Singh J., Pandit P., McArthur A.G., Banerjee A., Mossman K. (2021). Evolutionary trajectory of SARS-CoV-2 and emerging variants. *Virology journal*. **18**, 1-21.

Siu K.L., Yuen K.S., Castano-Rodriguez C., Ye Z.W., Yeung M.L., et al. (2019). Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *The FASEB Journal*. **33**, 8865-8877.

Snedden C.E., Makanani S.K., Schwartz S.T., Gamble A., Blakey R.V., et al. (2021). SARS-CoV-2: Cross-scale insights from ecology and evolution. *Trends in microbiology*.

Storti B., Quaranta P., Di Primio C., Clementi N., Spezia P.G., et al. (2021). A spatial multi-scale fluorescence microscopy toolbox discloses entry checkpoints of SARS-CoV-2 variants in VeroE6 cells. *bioRxiv*.

Sundaravaradan V., Das S.R., Ramakrishnan R., Sehgal S., Gopalan S., et al. (2007). Role of HIV-1 subtype C envelope V3 to V5 regions in viral entry, coreceptor utilization and replication efficiency in primary T-lymphocytes and monocyte-derived macrophages. *Virology Journal*. **4**, 1-12.

Tang X., Wu C., Li X., Song Y., Yao X., et al. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*. **7**, 1012-1023.

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*. 1-17.

Townsend J.P., Hassler H.B., Wang Z., Miura S., Singh J., et al. (2021). The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study. *The Lancet Microbe*.

van Dorp L., Acman M., Richard D., Shaw L.P., Ford C.E., et al. (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*. **83**, 104351.

Volz E., Hill V., McCrone J.T., Price A., Jorgensen D., et al. (2021). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*. **184**, 64-75. e11.

Wall E.C., Wu M., Harvey R., Kelly G., Warchal S., et al. (2021). Neutralising antibody activity against SARS-CoV-2 VOCs B. 1.617. 2 and B. 1.351 by BNT162b2 vaccination. *The Lancet*. **397**, 2331-2333.

Wargo A.R., Kurath G. (2012). Viral fitness: definitions, measurement, and current insights. *Current opinion in virology*. **2**, 538-545.

Weaver S.C., Forrester N.L., Liu J., Vasilakis N. (2021). Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. *Nature Reviews Microbiology*. **19**, 184-195.

Wen J., Cheng Y., Ling R., Dai Y., Huang B., et al. (2020). Antibody-dependent enhancement of coronavirus. *International Journal of Infectious Diseases*.

Wise J. (2021). Covid-19: The E484K mutation and the risks it poses, British Medical Journal Publishing Group.

Wong S., Lau S., Woo P., Yuen K.Y. (2007). Bats as a continuing source of emerging infections in humans. *Reviews in medical virology*. **17**, 67-91.

Woo P.C., Huang Y., Lau S.K., Yuen K.-Y. (2010). Coronavirus genomics and bioinformatics analysis. *Viruses* **2**, 1804-1820.

Xia H., Cao Z., Xie X., Zhang X., Chen J.Y.-C., et al. (2020). Evasion of type I interferon by SARS-CoV-2. *Cell reports*. **33**, 108234.

Zhang Y., Sun H., Pei R., Mao B., Zhao Z., et al. (2021). The SARS-CoV-2 protein ORF3a inhibits fusion of autophagosomes with lysosomes. *Cell Discov*. **7**, 31.

Zhou Y., Wu Y., Ding L., Huang X., Xiong Y. (2021). Point-of-care COVID-19 diagnostics powered by lateral flow assay. *TrAC Trends in Analytical Chemistry*. 116452.