Determinants of Genomic Diversity and Impact of Mutations in SARS-CoV-2

Franklin Rómulo Aguilar-Gamboa, MS*; Danny Omar Suclupe-Campos, BIOL†

SARS-CoV-2 has spread widely throughout the world, and multiple mutations and variants of interest were detected in the late 2020 and continue to emerge. Although genetic diversity is common in RNA viruses, these variations at the genetic level have given it certain characteristics associated with transmission, resistance to neutralizing antibodies, and even suspected increased lethality. Better understanding of the genomic diversity in SARS-CoV-2 will help to take appropriate containment measures against the virus. It should be borne in mind that this diversity can originate anywhere in the world, especially in areas where there is a high number of infections. This highlights the need for continuous molecular surveillance to guide development, therapy, vaccine use and health policy. [*P R Health Sci J 2023;42(2):102-110*]

Key words: COVID-19, Evolution, Genome, Viral, Molecular epidemiology, SARS-CoV-2, SARS-CoV-2 variants, Mutation

SARS-CoV-2 is the seventh species of the *Coronaviridae* family after 229E, NL63, OC43, HKU1, MERS-CoV and SARS-CoV, viruses that cause infections in humans (1). SARS-CoV-2 is the cause of COVID-19 disease, which in March 2020 initiated the second pandemic of the 21st century. Although not directly related to SARS-CoV, the clinical picture it produces, categorized as a severe acute respiratory syndrome, merited its current name.

Since the beginning of the pandemic, there has been much discussion about its probable zoonotic origin as a possible source of spread to humans, starting with strains MP789, RmYN02 and RaTG13 (2,3), the latter being the one with the highest whole genome sequence homology (96.2%) with SARS-CoV-2 (4). At present, there is controversy about the reliability of these sequences, since possible deficiencies have been found in these studies that would cast doubt on the true origin of the virus (5).

SARS-CoV-2 has a large positive-stranded RNA genome of 29.903 nucleotides that encodes about 30 mature proteins. The genome includes the six ORFs common to all coronaviruses, and at the 5′ end there are two large ORFs termed ORF1a and ORF1b, which cover more than two-thirds of the genome. Translation of ORF1a produces the pp1a polyprotein. A programmed frameshift -1 located four codons before the end of ORF1a directs a proportion of ribosomes to translate an alternative reading frame to the end of ORF1b, producing the pp1ab polyprotein. Both polyproteins are proteolytically cleaved into 11 or 15 mature nonstructural proteins (nsps), with the nsps within ORF1a and ORF1b responsible for controlling gene expression and viral replication, respectively (6). The last third of the genome encodes four structural proteins that are present in all coronaviruses (E, M, N and S), of these, the S or spike protein is responsible for ACE2 receptor binding, membrane fusion and viral entry through its S1 and S2 subunits respectively. The most notable mutations in SARS-CoV-2 are found in the gene encoding the S protein, of which more than 4000 mutations have been reported (7).

Probable origins of SARS-CoV-2

Since its emergence in Wuhan markets in China in December 2019, SARS-CoV-2 infection has spread worldwide. However, the first cases could be earlier than officially reported in many countries, because the presence of SARS-CoV-2 has been indirectly detected in biological samples preserved in biobanks. Researchers such as Basavaraju et al (8) have suggested that it may have appeared in the United States in December 2019, based on the identification of antibodies to SARS-CoV-2 in sera from blood donations collected by the American Red Cross, prior to the first officially reported case on January 19, 2020. Although it could not be confirmed whether these positive tests were from community transmission or travel, as only 11 of the volunteers reported having been to Asia recently (8).

Carrat et al (9) propose early circulation of the virus in Europe even before the event reported in Wuhan. They theorize that the outbreak may have occurred in November 2019, based on the identification of 13 positive anti-SARS-CoV-2 IgG tests collected from sera from a cohort in France, which were subsequently confirmed by antibody neutralization testing.

*Lambayeque Regional Hospital, Immunology and Virology Laboratories, Inmunología y Virología del Norte´s group, Lambayeque, Peru; †Pedro Ruiz Gallo National University, Faculty of Biological Sciences, Department of Microbiology, Inmunología y Virología del Norte´s group, Lambayeque, Peru.

The authors have no conflict of interest to disclose.

Address correspondence to: Danny Omar Suclupe-Campos, Email: dannysuclupcamp@gmail.com, inmunomicrobiologiaexperimenta@gmail.com

In addition, as they refer in their report, 11 of the participants disclosed having experienced symptoms possibly related to an infection or risk situations of possible exposure to SARS-CoV-2.

The studies cited above reveal that SARS-CoV-2 may have circulated prior to the Wuhan event, but they are based on the presence of antibodies, which could be a cross-reaction rather than true evidence of infection. Precise identification of the origin of SARS-CoV-2 requires certainty from direct data, and so far the closest is obtained from the bat coronavirus RaTG13 (4). This reveals the likely origin from bats, however RaTG13 was reported in 2013 and formed a distinct lineage from SARS-CoV-2, which could not be directly transmitted to humans, necessitating investigation of the presence of a likely intermediate host.

A number of studies have suggested mink, pangolin and snakes as intermediate hosts (10), however the strongest evidence has been found in pangolins. A coronavirus named Pangolin-CoV is 91.02 and 90.55 % identical at the whole genome level to SARS-CoV-2 and RaTG13 respectively, with the S1 protein of Pangolin-CoV being the most closely related to SARS-CoV-2, because five key amino acid residues involved in the interaction with human ACE2 are completely consistent with the SARS-CoV-2 sequence. Thus, pangolin species have been considered as a potential natural reservoir of CoV and a potential intermediate of SARS-CoV-2 (11).

In the investigation of the probable origins of SARS-CoV-2, it is important to identify the actual biological distance of genome sequences. Although traditional methods based on multiple sequence alignments (MSA) are widely used, these analyses do not satisfy the triangular inequality properties of the mathematical distance, which according to some authors guarantees better performance and data identification (12). In this regard, one study employed the "k-mer natural vector method" to encode the complete SARS-CoV-2 genome sequences with high quality in GISAID as vectors in Euclidean space, and after defining a new natural distance between the vectors, they concluded that SARS-CoV-2 most likely existed before the Wuhan outbreak in countries such as France, India, the Netherlands, England and the United States (13).

Genomic nomenclature in SARS-CoV-2

Since the emergence of SARS-CoV-2, many terms have been used to denote the products of genomic variability. Thus, terms such as mutations, variants, lineages, clades and strains have been used interchangeably, especially at the beginning of the pandemic. In this sense, it is necessary to define some concepts related to it in order to understand the impact that mutations have had on the genomic diversity of this virus.

A mutation occurs when there is a change in the sequence of nucleotides that make up the genome of the virus. The most frequent mutation in sequenced genomes worldwide is a transition affecting the nucleotide adenosine 23403, transforming it into a guanosine (A23403G). This mutation defines the so-called G clade of SARS-CoV-2 genomes, prevalent in Europe, Oceania, South America and Africa. The effect of this mutation is an amino acid change that produces the aforementioned D614G. Molecular dating analysis estimated the emergence of this clade in mid to late January (January 10 - 25, 2020) (14). Likewise, three mutations named C14408T, C241T and C3037T show a similar frequency with A23403G. So that, these four mutations almost always coexist in the same genomes, defining clade G, the main clade observed in the viral population (15). In SARS-CoV-2 as in other microorganisms, mutations arise independently several times and are called homoplasies, a term referring to the parallel evolutionary change that causes organisms to present the same independently acquired character. Of these homoplasies, 198 recurrent mutations have been identified in the SARS-CoV-2 genome (16).

Genomes that differ in sequence are often called variants, which are the product of many accumulated or major mutations, so that two variants may differ in one or more mutations. Mutations in the viral genome are conserved and accumulate in subsequent replications of the virus, allowing comparisons to be made to assess their similarity to each other and to generate phylogenetic trees by bioinformatics methods that allow their origin to be traced. These variants are frequently located within lineages, a term that is often used as a synonym for variant, although in other taxonomic scenarios a lineage may harbor two or more variants (17).

Clades are monophyletic groupings that harbor individuals related not only in space and time but also in genomic sequences; therefore, phylogenetic tree analysis is used for this category. Currently, the Nextstrain platform defines 27 phylogenetically related clades to classify SARS-CoV-2 variants, named based on the last two numbers of the estimated year in which they arose (19 to 22) and followed by a generic letter (A, B, C) (Figure 1). Thus, the first clade is named 19A and is considered the root clade from which the others arise (18). For its part, the GISAID database, an initiative originally created to deposit influenza virus genomes, for SARS-CoV-2 nomenclature and classification into clades, uses actual letters based on marker mutations, rather than generic letters (A, B) as in Nextstrain. For example, S-D614G is one of several genetic markers characterizing a new clade that increased sharply since February 2020 and the letter G was chosen to name it at that time. GISAID currently classifies SARS-CoV-2 sequences into 10 clades (Figure 1), where G, GH, GR, GV, GK, and GRA harbor the D614G mutation in the spike glycoprotein of the virus that is present in the newly reported variants of interest and concern (19).

A genetic lineage refers to the set of mutations that connect an ancestral genetic type, and similar to what occurs to define clades, lineage categorization is supported by the elaboration of phylogenetic trees. Thus, for SARS-CoV-2, a hierarchical and dynamic system was used to define the occurrence of new local outbreaks with epidemiological significance, using the command line tool and the web application Phylogenetic Assignment Of Named Global Outbreak Lineages (Pangolin) (21). This tool is based on the nomenclature proposed

by Rambaut et al (22) which assigns lineages to query sequences taking into account multiple sources of phylogenetic and epidemiological information, as well as a variety of metadata associated with the sequence. Pangolin indicates evolutionary relationships between SARS-CoV-2 lineages, in which each successive character denotes a subgroup of the previous one stipulated by letters, e.g., lineage B.1.1.7. This system seems to be the most appropriate, as it eliminates names that associate a variant or lineage with the country or region in which it was identified, avoiding designating them as British, South African, Brazilian variants, among others (21).

Lastly the definition of strain is the term that has given rise to most confusion by some media (23,24). In this regard, the World Health Organization (WHO) refers to the UK variant as follows, "more than 50 % of the isolates were identified as variant strain" (25). In another report it states that "studies in human respiratory cells and in animal models showed that, compared to the initial virus strain, the strain with the D614G substitution has increased infectivity and transmission" (26). As can be read in both reports, the WHO clearly identifies the term strain to refer to the product of a viral isolation, ensuring its purity and the absence of subpopulations or "quasispecies" that usually originate once they enter the susceptible host. This definition is complemented by Fauquet et al (27) who state that "strains are viruses belonging to the same species and differing in stable and heritable biological, serological and/or molecular characteristics".

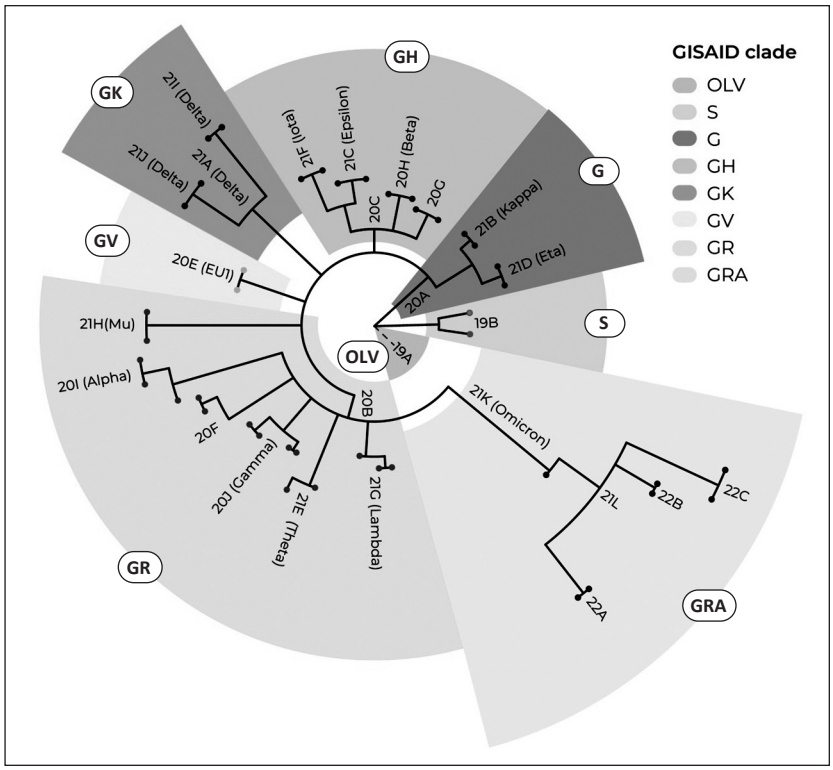

Figure 1. Schematic representation of the phylogenetics of SARS-CoV-2 among the different nomenclature systems. The sunray diagram represents the clades according to GISAID, and inside it contains their counterparts in Nextstrain. Figure modified from Ref (20), taking data from Ref (18).

In the most pragmatic sense of the term, a strain is characterized by being the product of a viral isolation and sequencing of the complete genome. For this, there must be some particularity or attribute that merits the performance of this work. For this reason, during the present pandemic it has been opportune to obtain viral strains when the virus first appeared and originated the outbreak (Wuhan-Hu-1 strain) (28), when it underwent genetic changes that produced phenotypic or biological patterns of interest (D614G strain) (29), when it represented a high number of cases in a given spatiotemporal location (strain Slovenia/SI-4265/20) (29), or when it was isolated to serve as reference patterns in the development of research such as the huge catalog compiled by the European Virus Archive-Global (30).

Determinants of SARS-CoV-2 genomic diversity

Most mutations that occur in the SARS-CoV-2 genome do not have a noticeable effect on the spread, virulence of the virus, and course of the disease (31). But there is concern that a change in the emerging mutations could lead to an increase in the severity or failure in the effectiveness of vaccines already developed. For this to occur, however, a substantial change in the genome sequence of SARS-CoV-2 would be necessary, which is unlikely due to the peculiarities of its genome and in general of most coronaviruses. Thus, mutations in the spike have received special attention, because it is the main target of antibody-mediated immunity and the antigen in current vaccines. These mutations occur at all positions and in different combinations of the spike (Figure 3), with more than 160,000 unique sequences reported to date (32).

The genome of the *Coronaviridae* is characterized by its large size compared to other RNA viruses of clinical importance. In the case of ssRNA (-) viruses it is about 10 to 15 kbp, while for ssRNA (+) viruses the genome size is variable. Those that infect bacteria and fungi have an average of 4 kbp, and those that infect algae, plants, invertebrates and vertebrates around 6 to 12 kbp (33). However, a 30 kbp genome represents a considerable challenge, since the errors introduced during replication are an important source of genetic variation that could be critical for all coronavirus populations, and thus require limiting the accumulated errors.

Limiting errors is an activity that favors the maintenance of genome identity, but from an evolutionary point of view

has allowed its growth in length, since if replication fidelity improves, then the upper limit imposed on length will increase, allowing evolution towards a larger genome. This will favor the evolution of a new function, thus improving viral fitness, which in turn will further improve replicative fidelity, and so on (34).

This relationship between genome size, fidelity and viral fitness is reflected in the so-called Eigen threshold, where viral fitness could potentially increase with genome size and replication fidelity (Figure 2). However, exceeding the limit of genome size imposed by the fidelity that can be reached at a given time in evolution leads to an "error catastrophe" (35). The latter event may have been responsible for the demise of SARS-CoV in 2003, where a 29-nucleotide deletion in the ORF8 gene during the early stages of human-to-human transmission led to the loss of viral fitness. Even years later, an *in vitro* study found a 23-fold decrease in viral replication as a result of this (36), deletion, indicating that SARS-CoV became extinct due to a catastrophe mutation of the error.

To overcome the obstacles imposed by the Eigen threshold, coronaviruses evolved to harbor nonstructural protein 14 (nsp14), which accompanies viral replicons during RNA synthesis. It corrects misincorporated ribonucleotides in nascent strands before they can spread, so that nsp14 exhibits exonuclease activity and prevents errors from becoming permanent. This correction capability was unknown among RNA viruses before its discovery in SARS-CoV, and contributes to a replication error rate more than 10-fold lower than that of other RNA viruses (37), which is likely responsible for the low genetic diversity of SARS-CoV-2 compared to other clinically important viruses.

Although exonuclease activity cannot correct insertions and deletions such as that already discussed in SARS-CoV, it is sufficient to maintain a low genetic diversity compared to other viruses such as influenza, hepatitis C and HIV. In the latter two, the great genetic diversity leads to the generation of quasispecies, which impedes the success of vaccines. However, in influenza virus, combined seasonal vaccines (influenza A: H1N1, H2N3, influenza B) have shown an effectiveness of up to 40%, sufficient to stop outbreaks (38). In other RNA-type viruses, the large number of mutations due to the lack of exonuclease activity could have an additive effect and represent a problem mainly in terms of vaccine efficacy.

What is of real concern in SARS-CoV-2 with the emergence of new mutations is a molecular phenomenon called "genetic epistasis," which is defined as the ability of genes to interact with each other. In other words, the problem of harboring a considerable number of mutations in SARS-CoV-2 lies not in the cumulative effect that may be more detrimental, but in the possibility of interaction that these mutations can produce. An example of this can be found in the N501Y mutation, which confers to the alpha variant the ability to spread, and although the delta variant does not possess it, the latter is much more contagious due to the greater number of mutations that interact with each other to improve the transmissibility of the virus.

On the other hand, genetic recombination is another mechanism by which SARS-CoV-2 may acquire some evolutionary advantage. This mechanism of viral genetic interaction is very common in the *Coronaviridae* due to their ability to produce sub genomic RNAs in the transcription of the last third of their genome that code for the structural proteins of the virus. This phenomenon has recently been described with the appearance of the XD or "deltacron" variant in which part of the delta variant genome and part of omicron are evident. Thus, both epistasis and genetic recombination currently represent a challenge for the control of this virus.

The relevance of the variants changes over time, however many mutations remain

After February 2020, viral genomes were observed to have distinct point mutations clearly discernible in different geographic regions (7). Subsequently, between 2020 and 2022 many variants were generated that were categorized of interest and concern according to attributes already described (39). This designation has been transitory in many cases, due to the decrease in the frequency of the same in different locations, and the emergence of new variants taking their place. Thus, the

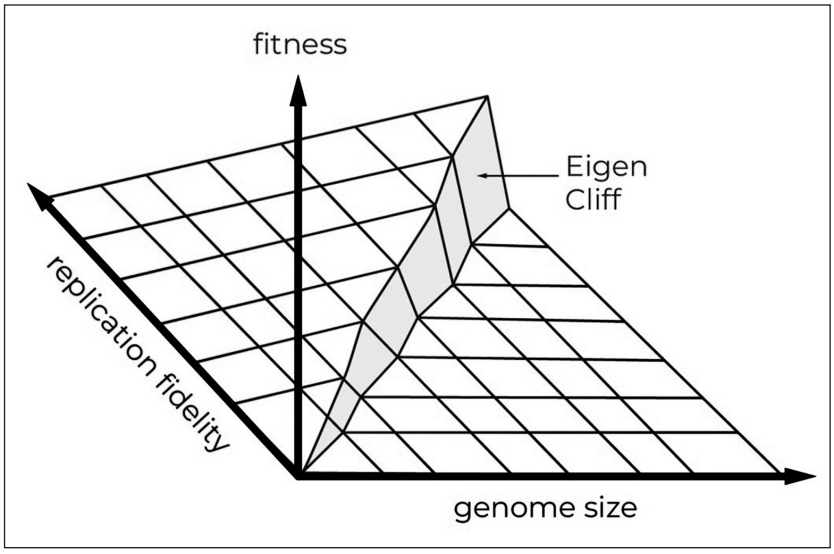

Figure 2. Eigen cliff. Viral fitness can potentially increase with genome size due to the accumulation of new beneficial mutations as long as replication fidelity is maintained, something that has been possible in coronaviruses due to the presence of nsp14 exonuclease activity. However, exceeding the limit of genome size imposed by the fidelity that can be achieved at a given time in evolution leads to "error catastrophe". Figure modified from Ref (35).

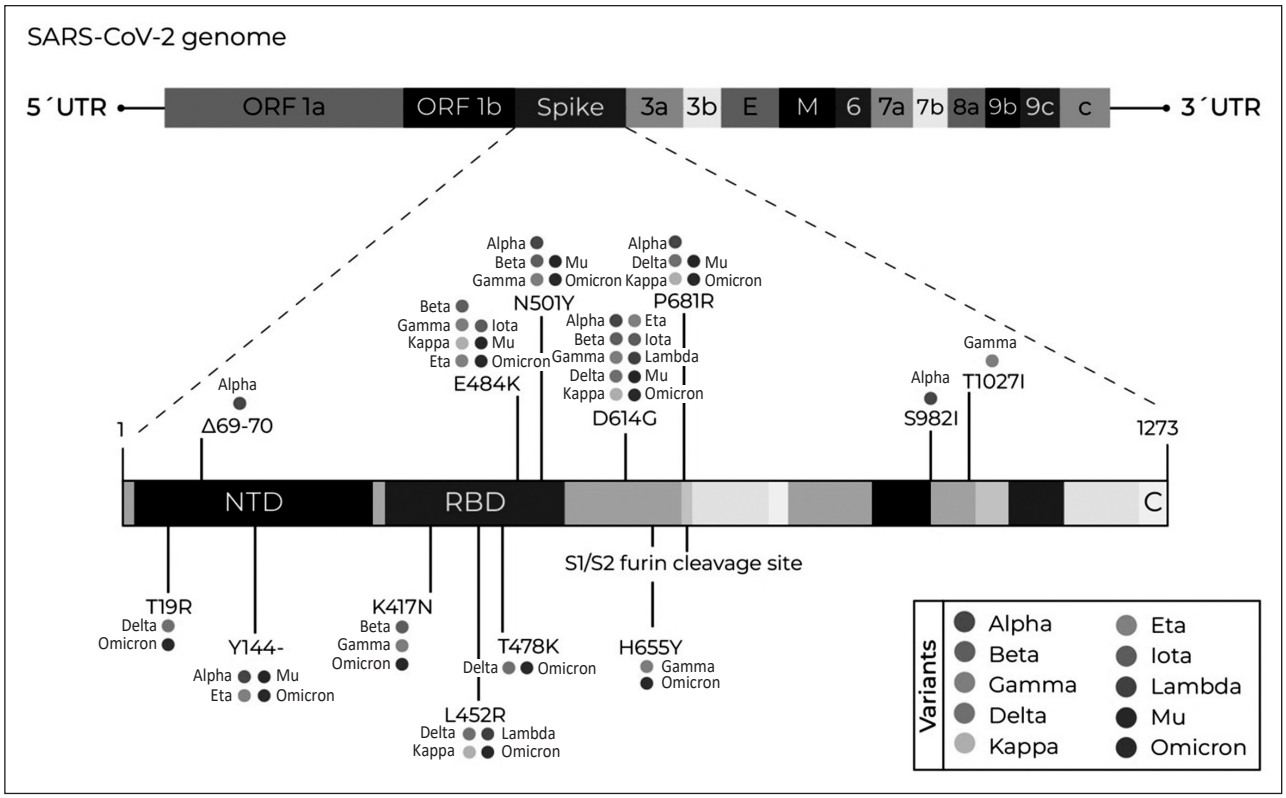

Figure 3. Mutations shared by SARS-CoV-2 variants. Some appear convergently in very distant lineages such as delta and omicron. Figure modified from Ref (41,42).

variants that initially caused concern have now been displaced, and it seems that the trend is towards a constant replacement of these variants (39).

However, what determines the category of variants and the characteristics they display beyond their epidemiological distribution is the set of synonymous and non-synonymous mutations they possess, many of which have remained over time, as in the case of D614G, which has appeared convergently and independently in very distant lineages, as well as E484K, N501Y, T478K, L452R and T19R (Figure 3). In view of this fact, some researchers have defined three types of mutation dynamics: high, medium and low frequency, based on their temporal dynamics that allow us to examine viral adaptation and evaluate the effects of control measures implemented in the evolution of the virus during the pandemic. Thus, it has been possible to determine which medium-frequency mutations are characterized by a high prevalence in specific regions and/or in constant competition with other mutations in various regions (40).

Given the frequency of the same mutations in many variants and their independent occurrence, it is suggested that natural selection has taken the path of their generation and thus the trend towards improved viral fitness. Some harbor attributes that allow the virus to successfully spread and probably allow it to establish seasonal cycles in the future. Among the mutations that are contributing to this and have maintained relevance during the development of the pandemic are the following.

Relevant mutations in the SARS-CoV-2 spike protein

D614G

The mutant form of D614 is known as D614G, or simply G614. It was one of 4 mutations differing from the original Wuhan form that formed the G clade found almost exclusively in Europe (43). It is currently present in 90% of circulating SARS-CoV-2 cases worldwide and is caused by a change from aspartic acid to glycine at position 614 of the spike (44). Its importance lies in different *in vitro* observations, which suggest that viruses carrying G614 have a higher affinity for binding and fusion with ACE2 of host cells, which increases their infectivity and greater propagation, as already demonstrated in case frequency studies at the beginning of the pandemic (45).

D614G was first observed in genomes sampled on January 28, 2020 in a small outbreak in Germany, initiated by a visitor from Shanghai. Therefore, it is likely that the mutation occurred in China before it was introduced multiple times in European countries. This scenario is consistent with the rapid increase in European virus genomes carrying the 614G variant in February and March (45). Another interesting finding is that D614G had a particularly high frequency (87 %) among SARS-CoV-2 specimens sequenced in Italy, which at that time turned out to be the most affected country outside China, with a mortality rate of 7.2 % (46). Thus, according to GISAID data, the frequency of D614G was increasing at an alarming rate throughout March and had an increasing geographic

spread to North America and then to Latin America by mid-April 2020 (47).

The increased frequency of D614G in SARS-CoV-2 was consistent with a selective advantage, although many authors attribute genetic drift and bottlenecks behind this fact, the truth is that it has been associated with attributes such as the ability to decrease the interaction between S1 and S2 units, facilitating the detachment of these and thus improving spike infectivity, activation of the fusion or obtaining ADE antibodies (43).

E484K

It was first identified in the beta variant (B.1.351) in South Africa, however the alpha (B.1.1.7) and gamma (P.1) variants classified at the time as SARS-CoV-2 variants of concern also possess it (39). E484K is a mutation harboring a substitution of glutamate (E) for lysine (K) at position 484 of the receptor binding domain (RBD) of the spike protein (Figure 3).

In vitro and clinical studies have shown that E484K reduces the neutralization of convalescent sera and those vaccinated with Ad26.COV2.S, NVX-CoV2373 and ChAdOx1 nCoV-19. There are also reports of monoclonal antibodies against the E484K mutation, showing a considerable reduction in susceptibility to casirivimab (25-fold), bamlanivimab and etesevimab; however, monoclonal antibodies such as imdevimab, sotrovimab, AZD7442, regdanvimab and REGN-COV2, maintained their efficacy (48). It is important to note the independent and convergent origin of this mutation, since both the delta and lambda variants do not present it, but in a more distant clade, omicron and its sub lineages do, which could reaffirm its importance in immune escape (Figure 3).

P681R

The delta variant rapidly replaced the alpha variant of SARS-CoV-2 worldwide. A mutation that played a key role in this was P681R, which has been shown to facilitate cleavage of the S1 and S2 subunits of the spike protein, enhancing viral fusogenicity and pathogenicity compared to the parental virus (49). In contrast, the spike protein of the alpha variant also has a mutation in the same amino acid (P681H), but the excision of its virions is reduced compared to delta (50). Although the infectivity of delta is not only due to P681R, it has been possible to appreciate that omicron and its sublineages present P681H that characterized alpha (41), thus, it is probable that it fulfills some function and participates in epistasis phenomena that require further study.

N501Y

It was first reported in September 2020 in the alpha variant of SARS-CoV-2. It contains at least 24 mutations in its genome, with the N501Y substitution in the spike being the most worrisome. N501Y is associated with higher transmissibility (70 to 80% more transmissible than the ancestral lineage) (51) and a reduced susceptibility to neutralization by antibodies (52). Enhancement of viral fitness for replication by N501Y in the upper airways has been demonstrated in hamster model and human airway epithelial cells (53), and the impact of N501Y on neutralizing activity, in convalescent sera from patients with COVID-19 and in anti-RBD immunoassays (52). This substitution is also present in the beta, gamma variants, and although delta does not, it was surprisingly reacquired by omicron (Figure 3), suggesting that N501Y is an important adaptive mutation of interest and likely the result of convergent evolution (53).

K417N

It is located in the receptor binding domain (RBD) of the spike protein of the beta, gamma and omicron variants of SARS-CoV-2 (Figure 3). The main characteristic of this mutation is that it gives the virus the ability to escape antibodies (54,55), in addition to reducing binding to ACE2 (56,57). However, when associated with other mutations such as N501Y and E484K, it creates new contacts between proteins that change the internal structural dynamics, thus increasing ACE-2 binding and ultimately infectivity (58). Similarly, the association with other mutations decreases the neutralization of antibodies from sera of patients convalescing from COVID-19 and from those vaccinated with one or two doses of BNT162b2 or mRNA-1273 (59).

T478K

It is found within the RBD of the SARS-CoV-2 spike protein, producing an exchange of uncharged threonine (T) for positively charged lysine (K) at position 478. This mutation can significantly alter the electrostatic surface of the protein, thus affecting the interaction with receptors, antibodies and drugs in a stronger or weaker way. The effect on viral infectivity and evasion of the immune response can be increased if it is combined with other mutations such as D614G, P681H and T732A, however it is unknown how these affect the virulence of SARS-CoV-2. T478K It is very prevalent in sequenced genomes of patients from Mexico and the US, but it has also been observed in variants such as B.1.1.519, delta, omicron distributed in many countries around the world, which makes continuous genetic and clinical monitoring important of this mutation and others in the spike protein, which will allow a better understanding of its effect on COVID-19 (60).

Other mutations of interest whose frequency has decreased

Although many mutations have appeared convergently and some have remained since the beginning of the pandemic, others have not been able to prevail over time. In this group, mention should be made of the S943P mutation, which was the first to occur in the spike protein in isolates from Belgium. The importance of S943P lies in the fact that it was the first evidence of a recombination mutation in an infected host (61). Although this mutation is local in terms of geographic distribution (Belgium), it did not arise as a single lineage from a

local founder effect, but is repeated in many distinctive lineages in the phylogeny circulating simultaneously in this country, a pattern suggestive of recombination.

S943P comes with a double base mutation in its codon, one of which is an AGT (S) -> CCT (P) transversion, therefore, it would require not one but two base mutations to have arisen together and repeatedly (co-selected). This is extremely unlikely and recombination is much more feasible since it is known to play an important role in the evolution of coronaviruses in general (61).

Another mutation that was relevant at the beginning of the pandemic was Y453F, found in the genetic variant known as "Cluster 5" detected in mink in Denmark in November 2020. It was considered to be an adaptive mutation that enhances binding to ACE2, in addition to showing resistance to antibodies from convalescent sera, which led to the mass death of millions of mink in this country, as it could represent a risk for the spread of COVID-19 and the development of vaccines (62,63).

Genetic recombination and the future of SARS-CoV-2

To date, three new variants designated as XD, XF and XE have been detected. The first two are combinations of delta and omicron called "deltacron", while XE is the result of the combination of the original omicron variant (BA.1) and its subvariant BA.2. The transmission capacity in the case of XD and XF are at similar levels, while XE has been considered the most contagious variant of all, since it exceeds the transmissibility to omicron BA.2 by 10% (64,65).

During the first half of 2022, the World Health Organization has warned of the emergence of new variants of omicron. These are BA.4 and BA.5, which are widespread in South Africa (66). Both lineages have mutations that were not present in other omicron variants. In addition, *in vitro* studies show that sera from previously infected or vaccinated people neutralize these lineages to a lesser extent. Although there is no evidence that these lineages cause more severe disease, more hospitalizations or more deaths, they are replacing the other variants worldwide (67).

Conclusions

It is unlikely that the genomic diversity of the virus will increase and produce a considerable immune escape that reduces the effectiveness of vaccines. Compared to other RNA viruses, the genomic diversity of SARS-CoV-2 is low, mainly attributed to the presence of the exonuclease Nsp14, which greatly limits the appearance of mutations of interest. Furthermore, although in a scenario of widespread viral dissemination the number of mutations increases, the effects of one or more mutations are not always additive. In this way, genetic interactions such as epistasis could have a more relevant role in the genomic diversity of the virus, since they can act both positively and negatively, increasing or decreasing the effect of each one. Likewise, the infection of immunocompromised hosts represents a risk due to its ability to motivate mechanisms such as recombination, and that together with epistasis, are actually the true determinants of genetic diversity that currently represent the greatest risk against SARS infection. -CoV-2.

Resumen

El SARS-CoV-2 se ha diseminado ampliamente por todo el mundo, y a finales de 2020 se detectaron múltiples mutaciones y variantes de interés, las cuales continúan apareciendo. Aunque la diversidad genética es común en los virus de tipo ARN, estas variaciones a nivel genético le han otorgado ciertas características asociadas a la transmisión, resistencia a anticuerpos neutralizantes e incluso se sospecha de una mayor letalidad. Una mejor comprensión de la diversidad genómica en SARS-CoV-2 ayudará a tomar medidas de contención adecuadas frente al virus. Hay que tener en cuenta que esta diversidad puede originarse en cualquier parte del mundo, especialmente en aquellas zonas donde existe un elevado número de contagios. Esto pone de manifiesto la necesidad de una vigilancia molecular continua para orientar el desarrollo, la terapia, el uso de vacunas y las políticas sanitarias.

References

1. Centers for Disease Control and Prevention (CDC). Human Coronavirus Types. Center for Disease Control and Prevention. Published 2020. https://www.cdc.gov/coronavirus/types.html
2. Liu P, Chen W, Chen JP. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (Manis javanica). Viruses. 2019;11(11):979. doi:10.3390/v11110979
3. Zhou H, Chen X, Hu T, et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr Biol. 2020;30(11):2196-2203.e3. doi:10.1016/j.cub.2020.05.023
4. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7
5. Deigin Y, Segreto R. SARS-CoV-2's claimed natural origin is undermined by issues with genome sequences of its relative strains. BioEssays. 2021;43(7):2100015. doi:10.1002/bies.202100015
6. Jungreis I, Sealfon R, Kellis M. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat Commun. 2021;12(1):2642. doi:10.1038/s41467-021-22905-7
7. Cosar B, Karagulleoglu ZY, Unal S, et al. SARS-CoV-2 Mutations and their Viral Variants. Cytokine Growth Factor Rev. 2022;63:10-22. doi:10.1016/j.cytogfr.2021.06.001
8. Basavaraju S V, Patton ME, Grimm K, et al. Serologic Testing of US Blood Donations to Identify Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)–Reactive Antibodies: December 2019–January 2020. Clin Infect Dis. 2021;72(12):e1004-e1009. doi:10.1093/cid/ciaa1785
9. Carrat F, Figoni J, Henny J, et al. Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based "CONSTANCES" cohort. Eur J Epidemiol. 2021;36(2):219-222. doi:10.1007/s10654-020-00716-2
10. The Conversation. Origine du virus de la Covid-19 : la piste de l'élevage des visons. Published 2021. https://theconversation.com/origine-du-virus-de-la-covid-19-la-piste-de-lelevage-des-visons-153219
11. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr Biol. 2020;30(7):1346-1351.e2. doi:10.1016/j.cub.2020.03.022

12. Chen J, Ng YK, Lin L, Jiang Y, Li S. On triangular Inequalities of correlation-based distances for gene expression profiles. bioRxiv. Published online January 1, 2019:582106. doi:10.1101/582106
13. Pei S, Yau SST. Analysis of the Genomic Distance Between Bat Coronavirus RaTG13 and SARS-CoV-2 Reveals Multiple Origins of COVID-19. Acta Math Sci. 2021;41(3):1017-1022. doi:10.1007/s10473-021-0323-x
14. Isabel S, Graña-Miraglia L, Gutierrez JM, et al. Evolutionary and structural analyses of SARS-CoV-2 D614G spike protein mutation now documented worldwide. Sci Rep. 2020;10(1):14031. doi:10.1038/s41598-020-70827-z
15. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol. 2020;11. doi:10.3389/fmicb.2020.01800
16. van Dorp L, Acman M, Richard D, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol. 2020;83:104351. doi:10.1016/j.meegid.2020.104351
17. Caponi G. Linajes y sistemas: dos tipos de individuos biológicos. Sci Stud. 2012;10(2):243-268. doi:10.1590/S1678-31662012000200003
18. Nextstrain. Genomic epidemiology of novel coronavirus - Global subsampling. Nextstrain. Published 2022. https://nextstrain.org/ncov/global?c=GISAID_clade&tl=GISAID_clade
19. GISAID. Clade and lineage nomenclature, March 2, 2021. GISAID. Published 2021. https://bit.ly/3bhjvT2
20. Gupta D, Sharma P, Singh M, Kumar M, Ethayathulla AS, Kaur P. Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants. Cell Mol Life Sci. 2021;78(24):7967-7989. doi:10.1007/s00018-021-04008-0
21. Á, O'Toole, E, Scher, A, Underwood, B, Jackson, V, Hill, JT McCrone, C, Ruis, K, Abu-Dahab, B, Taylor, C, Yeats, L, du Plessis, D, Aanensen, E, Holmes, O, Pybus, A R. pangolin: lineage assignment in an emerging pandemic as an epidemiological tool. PANGO lineages. Published 2020. https://doi.org/10.1038/s41564-020-0770-5
22. Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403-1407. doi:10.1038/s41564-020-0770-5
23. Viciosa M. Diferencias entre mutación, cepa, variante y transgénico y en qué te afecta. Newtral. Published 2020. https://www.newtral.es/diferencia-mutacion-cepa-variante-coronavirus/20201223/
24. Hartmann I. Variantes, cepas y linajes: el "capricho mutante" del genoma del coronavirus que explica los casos de reinfección. Clarín. Published 2020. https://bit.ly/2Z0RxBz
25. World Health Organization. SARS-CoV-2 Variant – United Kingdom of Great Britain and Northern Ireland. World Health Organization. Published 2020. https://www.who.int/csr/don/21-december-2020-sars-cov2-variant-united-kingdom/en/
26. World Health Organization. SARS-CoV-2 Variants. World Health Organization. Published 2020. https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/
27. Fauquet CM, Stanley J. Revising the way we conceive and name viruses below the species level: A review of geminivirus taxonomy calls for new standardized isolate descriptors. Arch Virol. 2005;150(10):2151-2179. doi:10.1007/s00705-005-0583-0
28. Pavan M, Bassani D, Sturlese M, Moro S. From the Wuhan-Hu-1 strain to the XD and XE variants: is targeting the SARS-CoV-2 spike protein still a pharmaceutically relevant option against COVID-19? J Enzyme Inhib Med Chem. 2022;37(1):1704-1714. doi:10.1080/14756366.2022.2081847
29. Centers for Disease Control and Prevention (CDC). Science Brief: Emerging SARS-CoV-2 Variants. Center for Disease Control and Prevention. Published 2021. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html
30. EVA-GLOBAL. SARS-CoV-2 collection. EVA-GLOBAL. Published 2021. https://www.european-virus-archive.com/evag-news/sars-cov-2-collection
31. Sonnleitner ST, Prelog M, Sonnleitner S, et al. Cumulative SARS-CoV-2 mutations and corresponding changes in immunity in an immunocompromised patient indicate viral evolution within the host. Nat Commun. 2022;13(1):2560. doi:10.1038/s41467-022-30163-4
32. Maher MC, Bartha I, Weaver S, et al. Predicting the mutational drivers of future SARS-CoV-2 variants of concern. Sci Transl Med. 2022;14(633). doi:10.1126/scitranslmed.abk3445
33. Campillo-Balderas JA, Lazcano A, Becerra A. Viral Genome Size Distribution Does not Correlate with the Antiquity of the Host Lineages. Front Ecol Evol. 2015;3. doi:10.3389/fevo.2015.00143
34. Martinez E, Argibay P. Acerca del origen de la vida: hacia una teoría unificada. Segunda parte. Rev Hosp Ital B Aires Vol. 2006;26(4):164. https://bit.ly/3OusYW1
35. Wolf YI, Koonin E V. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. Biol Direct. 2007;2(1):14. doi:10.1186/1745-6150-2-14
36. Muth D, Corman VM, Roth H, et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep. 2018;8(1):15177. doi:10.1038/s41598-018-33487-8
37. Eckerle LD, Becker MM, Halpin RA, et al. Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus Replication Is Revealed by Complete Genome Sequencing. Emerman M, ed. PLoS Pathog. 2010;6(5):e1000896. doi:10.1371/journal.ppat.1000896
38. Rausch JW, Capoferri AA, Katusiime MG, Patro SC, Kearney MF. Low genetic diversity may be an Achilles heel of SARS-CoV-2. Proc Natl Acad Sci. 2020;117(40):24614-24616. doi:10.1073/pnas.2017726117
39. World Health Organization. Tracking SARS-CoV-2 variants. World Health Organization. Published 2021. https://www.who.int/es/activities/tracking-SARS-CoV-2-variants
40. Justo Arevalo S, Zapata Sifuentes D, J. Huallpa C, et al. Dynamics of SARS-CoV-2 mutations reveals regional-specificity and similar trends of N501 and high-frequency mutation N501Y in different levels of control measures. Sci Rep. 2021;11(1):17755. doi:10.1038/s41598-021-97267-7
41. CoVariants. Shared mutations. CoVariants. Published 2021. https://covariants.org/shared-mutations
42. Ellis P, Somogyvári F, Virok DP, Noseda M, McLean GR. Decoding Covid-19 with the SARS-CoV-2 Genome. Curr Genet Med Rep. 2021;9(1):1-12. doi:10.1007/s40142-020-00197-5
43. Korber B, Fischer WM, Gnanakaran S, et al. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. bioRxiv. Published online January 1, 2020:2020.04.29.069054. doi:10.1101/2020.04.29.069054
44. Ozono S, Zhang Y, Ode H, et al. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. Nat Commun. 2021;12(1):848. doi:10.1038/s41467-021-21118-2
45. Volz EM, Hill V, McCrone JTJT, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 2021;184(1):64-75.e11. doi:10.1016/j.cell.2020.11.020
46. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA. Published online March 23, 2020. doi:10.1001/jama.2020.4683
47. Aguilar-Gamboa FR, Salcedo-Mejía LA, Serquén-López LM, et al. Genomic Sequences and Analysis of Five SARS-CoV-2 Variants Obtained from Patients in Lambayeque, Peru. Roux S, ed. Microbiol Resour Announc. 2021;10(1). doi:10.1128/MRA.01267-20
48. Yang WT, Huang WH, Liao TL, Hsiao TH, Chuang HN, Liu PY. SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations. Infect Drug Resist. 2022;Volume 15:373-385. doi:10.2147/IDR.S344099
49. Saito A, Irie T, Suzuki R, et al. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. Nature. 2022;602(7896):300-306. doi:10.1038/s41586-021-04266-9
50. Liu Y, Liu J, Johnson BA, et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. bioRxiv. Published online January 1, 2021:2021.08.12.456173. doi:10.1101/2021.08.12.456173
51. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Eurosurveillance. 2021;26(1). doi:10.2807/1560-7917.ES.2020.26.1.2002106

52. Lu L, Chu AWH, Zhang RR, et al. The impact of spike N501Y mutation on neutralizing activity and RBD binding of SARS-CoV-2 convalescent serum. EBioMedicine. 2021;71:103544. doi:10.1016/j.ebiom.2021.103544
53. Liu Y, Liu J, Plante KS, et al. The N501Y spike substitution enhances SARS-CoV-2 transmission. bioRxiv. Published online January 1, 2021:2021.03.08.434499. doi:10.1101/2021.03.08.434499
54. Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. bioRxiv. Published online January 1, 2021:2021.02.17.431683. doi:10.1101/2021.02.17.431683
55. Wang Z, Schmidt F, Weisblum Y, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. bioRxiv. Published online 2021:2021.01.15.426911. doi:10.1101/2021.01.15.426911
56. Thomson EC, Rosen LE, Shepherd JG, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell. 2021;184(5):1171-1187.e20. doi:10.1016/j.cell.2021.01.037
57. Starr TN, Greaney AJ, Hilton SK, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020;182(5):1295-1310.e20. doi:10.1016/j.cell.2020.08.012
58. Khan A, Zia T, Suleman M, et al. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. J Cell Physiol. 2021;236(10):7045-7057. doi:10.1002/jcp.30367
59. Fratev F. 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. Published online 2020:2020.12.23.424283. doi:10.1101/2020.12.23.424283
60. Di Giacomo S, Mercatelli D, Rakhimov A, Giorgi FM. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. J Med Virol. 2021;93(9):5638-5643. doi:10.1002/jmv.27062
61. Graham RL, Baric RS. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. J Virol. 2010;84(7):3134-3146. doi:10.1128/JVI.01394-09
62. Ren W, Lan J, Ju X, et al. Mutation Y453F in the spike protein of SARS-CoV-2 enhances interaction with the mink ACE2 receptor for host adaption. Perlman S, ed. PLOS Pathog. 2021;17(11):e1010053. doi:10.1371/journal.ppat.1010053
63. BBC News Mundo. Coronavirus en Dinamarca: cómo el sacrificio de millones de visones para frenar el covid-19 desató una crisis política en el país escandinavo. BBC. Published 2020. https://www.bbc.com/mundo/noticias-internacional-55130946
64. Fernández M. La OMS halla la variante covid XE, la más contagiosa de todas las conocidas. Redacción médica. Published 2022. https://bit.ly/3QHFYsu
65. Redacción médica. La OMS confirma la variante covid Deltacron, que mezcla Delta y Ómicron. Redacción médica. Published 2022. https://bit.ly/3QCNl4x
66. Mateos A. Dos nuevas variantes covid a estudio por su escape inmunitario. Redacción médica. Published 2022. https://bit.ly/3QF4mv2
67. Ministerio de Salud-Perú. Minsa anuncia presencia de linajes BA. 4 y BA. 5 de la variante ómicron en el Perú. MINSA. Published 2022. https://bit.ly/3zWu6xc