POSSIBLE MUTATION PATHWAYS IN SARS-CoV-2

MUHAMMAD TOREQUIL ISLAM 1, CRISTINA QUISPE 2, JESÚS HERRERA-BRAVO 3, ISHAQ N. KHAN 3, SAUD SALEM BAWAZEER 6, MANOJ KUMAR 7, NATÁLIA CRUZ-MARTINS 8,9,10, MIQUEL MARTORELL 11, ANCA OANA DOCEA 12, JAVAD SHARIFI-RAD 13, SEVGI DURNA DAŞTAN 14,15, DOINA DRĂGĂNESCU 16, ANDREEA LETIȚIA ARSENE 17, DANIELA CĂLINA 18

1Department of Pharmacy, Life Science Faculty, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj (Dhaka) 8100, Bangladesh
2Facultad de Ciencias de la Salud, Universidad Arturo Prat, Avenida Arturo Prat 2120, Iquique, 1110939, Chile
3Departamento de Ciencias Básicas, Facultad de Ciencias, Universidad Santo Tomás, Chile
4Center of Molecular Biology and Pharmacogenetics, Scientific and Technological Bioresource Nucleus, Universidad de La Frontera, Temuco, 4811230, Chile
5Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, 25120, Pakistan
6Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Umm Al-Qura University, Makkah, 42, Saudi Arabia
7Chemical and Biochemical Processing Division, ICAR – Central Institute for Research on Cotton Technology, Mumbai-400019, India
8Faculty of Medicine, University of Porto, Porto, Portugal
9Institute for Research and Innovation in Health (i3S), University of Porto, Porto, Portugal
10Institute of Research and Advanced Training in Health Sciences and Technologies (CESPU), Rua Central de Gandra, 1317, 4585-116 Gandra PRD, Portugal
11Department of Nutrition and Dietetics, Faculty of Pharmacy, and Centre for Healthy Living, University of Concepción, 4070386 Concepción, Chile
12Department of Toxicology, University of Medicine and Pharmacy of Craiova, 200349 Craiova, Romania
13Facultad de Medicina, Universidad del Azuay, Cuenca, Ecuador
14Department of Biology, Faculty of Science, Sivas Cumhuriyet University, 58140, Sivas, Turkey
15Beekeeping Development Application and Research Center, Sivas Cumhuriyet University, 58140, Sivas, Turkey
16Department of Pharmaceutical Physics and Informatics, “Carol Davila” University of Medicine and Pharmacy, 020956, Bucharest, Romania
17Department of General and Pharmaceutical Microbiology, “Carol Davila” University of Medicine and Pharmacy, 020956, Bucharest, Romania
18Department of Clinical Pharmacy, University of Medicine and Pharmacy of Craiova, 200349, Craiova, Romania

*corresponding author: ancadocea@gmail.com.

Manuscript received: August 2021

Abstract

Mutations are the best way to generate genetic variation, as they provide the raw material in which evolutionary forces, like natural selection, can act. However, mutations are not always able to change the apparent behaviour of an organism, instead, they are capable of providing normal and abnormal biological functions. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has triggered a new and controversial biological scenario. Despite the hundreds of studies performed, to date, there is no specific treatment available. Several genomic and non-genomic mutations have been also reported, and due to its high genome size and mutation capacity, it can acclimatize to variable environments and has become the leading cause of high infection and mortality rates. This review outlines the possible pathways behind SARS-CoV-2 mutations.

Rezumat

Mutățiile reprezintă cea mai frecventă modalitate de generare a variatiiilor genetice, oferind posibilitatea mecanismelor de selecție naturală să influențeze evoluția speciilor. Totuși, mutațiile nu sunt întotdeauna capabile să schimbe comportamentul aparat al unui organism, dar pot activa funcțiile biologice normale sau anormale. SARS-CoV-2 a declanșat un nou și controversat scenariu biologic. Deși există numeroase studii efectuate și publicate, până în prezent, nu există un tratament specific disponibil. Au fost raportate mai multe mutații genomice și non-genomice ale noului coronavirus. Datorită dimensiunii sale mari și a capacității de a suferi mutații, se poate aclimatiza în medii variabile, devenind cauza principală a ratelor ridicate de infeție și mortalitate. Acest studiu evidențiază mecanismele posibile ce stau la baza mutațiilor SARS-CoV-2.

Keywords: SARS-CoV-2, mutations, origin, transmission, virulence, genetic diversity, vaccines
Introduction

Viruses have either a deoxyribonucleic (DNA) or ribonucleic (RNA) acid with great diversity in their sequences, originated from primitive pre-cell life forms, escaped cell genetic elements, or retrograde evolution [27]. Viruses can infect all kinds of the life-forms present in the ecosystems. Indeed, there is no doubt that viruses have affected the evolution of both single- and multi-cellular organisms from a long time ago [49]. Animal chromosomes are composed of RNA and DNA so that they can mix viral genomes. For example, the genes necessary for the development of mammalian placenta are also found in retroviruses, with many viral gene fragments being also used to fight infections from various origins [82].

The term “mutation” refers to the everlasting changes in the type, number or sequence of nucleotides in nucleic acids. However, cells have powerful repairing systems that can fix the deficiencies occurring during nucleic acids replication, where mutations may occur under normal conditions to a certain extent [16]. Generally, unlike in humans, viruses’ lack this same precise fixing system after replication errors in their viral genomes, with many viral gene fragments being also used to fight infections from various origins [82].

Briefly, the mutation rate is defined as the rate of mutation per site per genome replication or the rate of mutation per site per round of viral replication, so that it should be consistent and exist in an unbiased manner [29]. However, as viruses can use stamping machines and/or geometric genome replication, mutation rates can also vary between them. In the first case, one single virus genome is used as a mould for replication, resulting in a linear gathering of mutations, while in the latter case progeny strands may become moulds for replication themselves, leading to an exponential (or geometric) increase in progeny genomes, that may result in a very different distribution of mutations within the progeny genomes. Luria-Delbruck fluctuation assay, mutation accumulation and composition studies are the most usually performed tests to assess viral mutations [106].

The current positive-strand RNA novel coronavirus thought to have originated from a wet meat market of the Wuhan city of China has rapidly spread to nearly all countries of the world. Since its onset, the scientific community has shifted towards the understanding of various factors involved in infectivity, disease course, prevention and therapeutic measures. For that, several clinical trials have also been undertaken with several drugs and vaccines in an attempt to stop the disease spread and infectivity. Preliminary genome sequencing data from various countries have shown that the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome is prone to several mutations, suggesting a rapid mutation rate [79]. Several potential factors may influence the SARS-CoV-2 mutation rate (Figure 1), for the discovery of a suitable therapy for any viral infection, it is necessary to know about the mechanism of mutation within the virus as these findings have a direct impact on the prospects of an effective vaccine with long-term immune protection.

Figure 1.
Potential factors (cellular metabolism, genomic errors, proteomic errors, environmental factors) influencing SARS-CoV-2 mutation

In this sense, this review aims to discuss the current knowledge on mechanisms of RNA viral mutations, to hypothesize the possible mechanisms that SARS-CoV-2 can use to undergo further mutations, and
also the possible effects of such mutations on virus viability, virulence and infectivity. An extensive search of articles on this topic was conducted in the PubMed and Scopus databases, from December 1, 2019 to October 1, 2020, using Medical Subject Headings (MeSH) terms: “SARS-CoV-2”, “COVID-19”, “viral genome”, “mutations”, “genetic diversity”, “transmission”, “virulence” and “pathogenicity”. Only papers written in English were included, and both abstracts and letters to editors were not included.

SARS-CoV and mutation mechanisms

Mutation and mutation rate of viruses: a brief overview

Due to the presence of a deoxyribose sugar, that contains one less oxygen-containing hydroxyl group, DNA is more stable than RNA [5]. Because of this, RNA viruses or single-stranded viruses mutate faster than DNA viruses or double-strand viruses, making them more adaptable to contrasting environments, and these features are a major cause of its high pathogenicity, attributed to their high genetic diversity [5]. A specific or regular mutation is needed to enable them to survive in particular environments. First of all, genome size negatively correlates with mutation rate [10], where the viral mutation rates depend on polymerase fidelity, sequence context, cell micro-environment, template secondary structure, replication mechanisms, proofreading and access to post-replicative repair [94]. Briefly, viral mutation rates almost range from $10^{-8}$ to $10^{-4}$ substitutions per nucleotide per cell infection (s/n/c), in which RNA viruses occupy the $10^{-8}$ to $10^{-4}$ range (with some exceptions) (Table I).

<table>
<thead>
<tr>
<th>Virus</th>
<th>Genome size (Kb)</th>
<th>Mutation rate (s/n/c)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatitis C virus</td>
<td>9.65</td>
<td>$1.2 \times 10^{-4}$ to $3.5 \times 10^{-5}$</td>
<td>[20, 37]</td>
</tr>
<tr>
<td>Tobacco etch virus</td>
<td>9.49</td>
<td>$3.0 \times 10^{-3}$ to $4.8 \times 10^{-6}$</td>
<td>[92, 109]</td>
</tr>
<tr>
<td>Human norovirus G1</td>
<td>7.65</td>
<td>$1.5 \times 10^{-4}$</td>
<td>[21]</td>
</tr>
<tr>
<td>Poliovirus 1</td>
<td>7.44</td>
<td>$1.1 \times 10^{-3}$ to $3.0 \times 10^{-4}$</td>
<td>[23, 112]</td>
</tr>
<tr>
<td>Human rhinovirus I4</td>
<td>7.33</td>
<td>$4.8 \times 10^{-3}$ to $1.0 \times 10^{-5}$</td>
<td>[43, 116]</td>
</tr>
<tr>
<td>Tobacco mosaic virus</td>
<td>6.40</td>
<td>$8.7 \times 10^{-6}$</td>
<td>[70]</td>
</tr>
<tr>
<td>Bacteriophage Qβ</td>
<td>4.22</td>
<td>$1.4 \times 10^{-4}$</td>
<td>[10]</td>
</tr>
</tbody>
</table>

The previous two coronavirus species, namely middle east respiratory syndrome (MERS) and SARS-CoV presented a similar genome size (30 Kb), with mutation rates of $1.12 \times 10^3$ and $0.80 - 2.38 \times 10^3$ substitutions per site per year, respectively [19, 123], while SARS-CoV-2 has a genome size of 32 Kb and a mutation rate of $\sim 10^{-6}$ mutations per site per cycle [122], which is the largest genome size and highest mutation rate among the RNA viruses, including coronavirus, disclosed till date [17]. SARS-CoV-2 has a high mutation rate (genome sequence varying up to 0.02% or $2 \times 10^{-5}$) in comparison to the human (genome sequence varying only by 0.001% between individuals), so that the viral mutation rate may depend on its replication cycle, required replication materials used by it, and type of host cells and cell support [99]. Thus, differences in exchange rates among viruses could be a result of differences in virus RNA synthesis rates in distinct cell types [44]. As a result of a greater or lesser mutation rate, and with the progress of the pandemic and the spread of the virus at a worldwide level, this continuous mutational process will lead to the appearance of an increasing viral variability [80]. In this process, only the reduction in the number of infections will primarily lead to a drastic decrease in the viral mutation rate, although the various antiviral drugs may also lead to a process of viral adaptation with the appearance of more infectious variants, which at medium-long term may lead to a reduction in the effectiveness of vaccines and even of drugs currently available and under development [80]. Taken together, if these processes are not understood more deeply, resorting to the use of various genomic and molecular techniques, as the virus may still be in a phase of adaptation to the human host, they may be a matter of great concern and extreme impact on public health.

Possible SARS-CoV-2 mutation pathways

Ligand-receptor interactions

It has been decoded that SARS-CoV-2 enters lung cells by interacting with the angiotensin-converting enzyme 2 (ACE2) receptor. However, besides lungs, ACE2 is also present in many other organs, like the small intestine (enterocytes), arteries (smooth muscle and endothelial cells), veins (endothelial cells), cerebral cortex, striatum, hypothalamus, brainstem and eyes [51]. The disclosure also suggests that the gut may be one of the potential sites of virus replication [26], with the transmembrane protease, serine (TMPRSS)-2 and -4 being proclaimed to facilitate the SARS-CoV-2 spike fusogenic activity, thereby promoting the viral entry into the host [119] (Figure 2).

The high mobility group box 1 (HMGB1) protein, also known as high-mobility group protein 1 (HMG-1) and amphotericin encoded by the HMGB1 gene in humans, is a protein predominantly pleiotropic with an important role in innate immunity [50]. This gene delivery agent can bind to the SARS-CoV-2 RNA and bring it to the cytosol via the receptor for advanced glycation end products (RAGE)-lysosomal pathway.

![Figure 2](image-url)
(HMGB1-assisted transfer combined with lysosome leakage). Thus, it may be an additional pathway apart from ACE2 receptors enabling intracellular virus replication. A recent study demonstrated that the virus requires alarming HMGB1 for its cytosolicFc replication [91]. Thus, there is an incidence to find the viral mobile genetic materials in the bloodstream, which can disseminate in many other cells and tissues of COVID-19 patients, making possible the virus replication outside the lung epithelial cells [13].

Figure 2.

Diagram showing the genome organization and representative structure of the SARS-CoV-2
S: spike protein; E: envelope protein; gRNA: genomic RNA; N: nucleocapsid protein; M: membrane protein; TRS-B: transcription-regulatory sequences in the leader; TRS-B: transcription-regulatory sequences in the body; ORF: open reading frame; UTR: untranslated region; AAAA: poly-A tail.

More recently, this virus has been found in the heart, liver and semen of COVID-19 patients [76, 110]. Indeed, one or more amino acid substitutions within specific regions of the viral surface protein may result in viral surface protein mutations [125]. This may affect its extracellular existence, the cell entrance of the virus and overall survival. In contrast, a mutation in viral surface proteins may enhance the binding capacity or even alter the entrance mode selection [14]. Generally, viruses utilize elegant strategies to attach to one or multiple cell receptors, to overcome the plasma membrane barrier, to enter, and access the necessary host cell machinery; for that, they may target particular classes of molecules [69]. In addition, an alteration in host receptor properties by mutation or by other means may also affect viral entrance in host cells. As the virus-receptor interaction is necessary for a virus to enter and replicate, host cells with a mutation inside the receptor gene that prohibits virus infection persist and finally predominate the population. A virus could defeat this issue by changing an amino acid, allowing altered receptor binding [85, 100]. Mutations that modify the interaction to favour the virus or the host are called “positively selected” mutations, while such back-and-forth progression between viruses and their host cells is called “host-virus arms races”. Recently, the Bangladesh Council of Scientific and Industrial Research (Bangladesh) has found several mutations in the spike (S) protein of SARS-CoV-2 in some Bangladeshi samples [72]. Replication and proofreading

As stated above, RNA viruses are very simple structures with limited genomic sizes ranging from about 2 - 32 Kb [63]. Thus, these viruses possess a small coding capacity and act as obligate intracellular parasites, depending on the host cells for replication, energy generating systems, ribonucleotides and deoxyribonucleotides, cellular translation machinery, tRNAs and amino acids to translate their mRNAs, cell enzymes (for posttranslational modification of their proteins, cell structures, like membranes, vesicular compartments and/or cytoskeleton networks) and finally for assembling and transporting them or their components [8]. Most RNA viruses lack a polymerase proofreading system, making them more prone to faults than DNA viruses [108] However, coronaviruses, a family of positive-strand RNA viruses that encode a complex RdRp have a 3’ exonuclease domain. RdRps can delete, insert and mismatch the nucleotides into the genome product [96]. Certain non-segmented RNA viruses may contain an extra protein that could act to decrease RdRp error. In exception, coronaviruses are the larger genome containing RNA viruses (26 - 32 Kb), having about 125 nm in diameter [32] and an evolved proofreading capacity of 3’- to 5’-exoribonuclease within non-structural protein 14 (nsp14-ExoN) [103].
It has been demonstrated that they sustain a relatively large genome due to their genome proofreading capacity [24] since this system can remove faulty insertions at the 3’ end of the genome product during RNA synthesis [98]. It has been observed that the exonuclease activity can be augmented by an extra coronavirus protein [104], on the contrary, any substitution in this system can result in mutation. For example, alanine substitution of ExoN catalytic residues [ExoN(-)] in SARS-CoV disrupted ExoN activity resulted in viable mutant viruses with defective replication, with up to 20-fold-decreased fidelity and increased susceptibility to nucleoside analogues [41]. Generally, defects in this polymerase system result in mutations, as it is the most important factor which dictates how powerfully the viral genomes replicate. The polymerase fidelity varies according to the template features, with misalignments at homo-polymeric runs which may cause frameshift mutations and base substitutions [60]. Thus, a strong selection at the protein level may cause amino acid replacements from the pre-existing RNA secondary structures, which may result in a modified replication fidelity of the viral genome [36].

RNA viruses are unable to target mutations to specific genomic regions to improve their adaptability. Polymerase variants with altered fidelity also affect the viral mutation rate as it can raise the replication fidelity through a reduced helicase activity, augmented replication kinetics, and resistance to low nucleotide concentrations [59]. Such features may be due to the presence of certain substitutions in polymerase mutant, thereby resulting in exonuclease-deficiency and a mutated phenotype [78]. However, this variant may be compensated rapidly through suitable substitution(s) that restores genome replication fidelity in their genetic background. A possible mutation pathway is shown in Figure 3.

**Figure 3.**
Mutation through viral replication and proofreading pathway. Unlike other RNA viruses, SARS-CoV-2 has a proof-reading system; and when the virus replicates, the copy is corrected, which means a lower probability of genetic mutations.

Polymerases are the main cause of the huge mutation rates of coronaviruses, as under selective pressure they lead to fast genomes’ replication and secure efficacious viral infection. Thus, these viruses evolve an equilibrium between fast genome synthesis and mistakes, so that the mutations they afford are acceptable and sometimes profitable, although not essential for their survival. However, in genome synthesis, the rate is a serious factor in virus activity and survival in the host [87].

**Genomic alterations**
There is an inverse correlation between viral genome size and its mutation rate. However, bacteriophage Qβ does not follow this rule [10]. The RNA viruses, including SARS-CoV-2 that replicate within the cytoplasm, without a DNA intermediary, do not incorporate their genomic materials into the host chromosome. However, their largest RNA genome may act as a principal room for the insertion of large foreign genes. Conversely, like many other viruses, it takes advantage of mechano-genomic regulation in host cells for its replication and propagation. But it is to be mentioned that the genomic variables also have a strong correlation with the effective population size of an organism [9]. Reiter et al. [89] reported that viral mobile genetic elements can integrate into the host
chromosome through site-specific recombination within perfect homology between the short region of the host chromosome and viral genetic material. For this, tRNA genes are the typical targets; therefore, this concept may be conserved during the evolution of integrating viruses to host cells [3]. Recently, the BCSIR (Bangladesh) has reported that this virus can kill many beneficial bacteria in humans. Through mutation, an organism increases the probability to transfer changed genetic information to its next-generation(s), despite the mutation rates of RNA viruses being greatly affected by other host-encoded factors. For example, high levels of reactive oxygen species (ROS) in addition to other cell metabolites through viral infections can cause mutations in both host cells and viruses [90]. Certain chemical substances after attaching to nucleic acids can exert mutagenic effects, similarly to high energy radiation exposure, including X-ray and ultraviolet (UV) rays. A possible mutation pathway is shown in Figure 4.

![Diagram](image)

**Figure 4.**
Mutagenic effects on strands merge. RNA polymerase transcription errors and the lack of inefficient correction mechanisms lead to the incorrect incorporation of a base that can change the message of a codon, which allows the genetic material to be re-assorted.

**Lysosomal activity**
After endocytosis, the coalition of the virus envelope with the endosome membranes and the release of viral RNA into the cytosol of infected cells is due to cleavage of the SARS-CoV “S” protein by the lysosomal proteases cathepsin L and cathepsin P in early endosomes [44, 46]. As it is a strongly adjusted process, differences in lysis time can adjust the rate of mutations in viral populations. The retardation in population growth could occur if the lysis takes place before this; as a consequence, the infected cell will release a little amount of viral progeny, leading to a small number of cells infected in the next infection cycle. On the contrary, if it occurs after the optimum, a large amount of progeny will be produced per cell, but the cell-to-cell transmission will be retarded [31]. The best lysis time is related to the time required to start producing progeny virions (lag/eclipse time), the capacity of infected cells to produce virions (yield) and the virus/host population densities (multiplicity of infection). A retarded lysis may augment viral yield per cell, with progeny virions having additional time to accumulate in cells, while increases the virus ability to tolerate mutagenesis.
However, the mutation rate also affects the optimum lysis time [94].

Recombination

Recombination (gain-of-function mutation) promotes the rapid production of genetic diversity in other genomic regions. This process starts when viral polymerase switches between various template genomes among the same cell [101]. It imparts an immediate fitness advantage, elevating the number of potential targets for the occurrence of following selectively important point mutations in a viral genome so that it can be used as an evolutionary process characterized by expansion and contraction of a specific genome region [107].

It has been demonstrated that recombination is a factor that allowed the occurrence of SARS-CoV variants [42]. Briefly, the RNA viruses can recombine by three distinct mechanisms, template-switching recombination, non-replicative recombination, and re-arrangement. Template-switching, also called copy-choice recombination, more frequent to positive-strand RNA viruses, can occur during the RNA synthesis process if the viral polymerase transfers from one template to another, while still connected to the nascent nucleic acid chain [56]. A mosaic genome is produced in this process. It occurs among a series of close similarities to give rise to an asymmetric recombination phenomenon, being also characteristic in coronaviruses [66]. During transcription, the gene expression requires the RdRp to transfer the genomic template from one site to another [97], thereby, the polymerases of such viruses can transmit a diverse template sequence during their RNA synthesis process. Secondly, the non-replicative recombination occurs when two viral RNA fragments are not capable of sharing in replication by themselves. Thus, they are recombined to form functional RNAs [33]. In this process, the RNA strands may be combined either by a transesterification reaction [18] or by ligation, probably by cell ligases [7], creating a novel RNA genome by physical shearing, nuclelease cleavage or cryptic ribozyme activity, which can be moreover refined by similar recombination to expel repeated sequences [45]. On the other hand, re-arrangement occurs during coinfection of a cell with viruses with segmented genomes [113], where a virus can exchange one of its segments for that of another related virus. This can be understood by the multiplicity of infection (MOI) value which refers to the ratio between the number of viruses and the number of infected cells [71]. A high MOI means that cells are co-infected with multiple viruses, while MOI stands for each cell is most likely infected by one virus only. In this case, the mutation rate of a virus may be influenced by the mutation capability and rate of other viruses. High MOI can also produce multiple genomic copies of the same gene in one infected cell, leading to complex and conflicting effects on genome selection. On the other hand, frequent recombination requires more efficient selection, efficient removal of deleterious alleles and the incorporation of adaptive alleles. High rates of recombination/re-arrangement may also lead to the emergence of strains with a more virulent phenotype, which is contrary to high MOI [105]. MOI also indicates the distribution pattern of viral particles at different sites of a host and the probability to establish an infection in a novel host.

Both mutation and recombination processes are firstly governed by the viral polymerase [94, 118]. In general, high mutation rates not only give the virus the ability to produce beneficial mutations quickly, but also amplify the genetic load of the population [30]. On the other hand, frequent recombination promotes the unlinking of beneficial mutations from deleterious genetic origins and their combination into the same genome; therefore, an increase in adaptation is expected in the face to a high number of coexisting alleles in the same population [39]. However, this increase in viral adaptability is observed to some extent, from which high rates of mutation and recombination lead to an accumulation of deleterious alleles in the population [94, 95].

Two mutated virus-genomes infecting the same cell may ensure a possibility to recover wild-type virus provided mutations, which may not affect the same gene and are located at a certain distance from each other [84]. In a single crossing, it may carry both mutated genes in the wild-type virus. The recombination frequency can be used to determine the relative distance between two genes on a virus genome [83]. Simultaneous analysis of genetic recombination and restriction enzymes can identify the exact place where both crossing, and recombination occur. In some animal viruses, reoviruses and orthomyxoviruses, the re-distribution of genetic material or re-assortment has been evident. However, the simultaneous infection with two different viral species with a difference in certain markers, both parents can be identified at a higher percentage yield containing a mixture of the segmented genome. In this case, the progeny virus selects its genome segments from a common pool in the infected cell and contains only a copy of the RNA segment. The mechanism of RNA crossing-over is yet to disclosed; however, recombination presumably happens due to a template transfer during the early stage of synthesis of complementary minus strands. It has been suspected that the virus polymerase may begin on one RNA and then shift the replication template [52]. A possible mutation pathway is shown in Figure 5.
Mutations through recombination. Recombination produces a new virus, with a new combination of genes and new antigenic properties. The emergence of a new viral strain is conditioned by its ability to multiply and spread to susceptible organisms, and the occurrence of epidemics or pandemics depends on the ability of the virus to spread.

**Conditional mutations**

The mutation effects may not only exert a lethal effect but may also help to successfully control the pathogen. For example, certain modifications may result in stable proteins, retaining their function at lower temperatures, while having a configurational change at higher temperatures, being thereby, rapidly hydrolysed by cell proteases. This is called temperature-sensitive (ts) mutation [52]. At first, it has been suspected that SARS-CoV-2 is infectious only at low temperatures. The present scenario indicates that it’s about the temperature adaptability, suggesting possible temperature adaptable mutations of this pandemic [35]. These kinds of mutations generally affect single bases and usually have a substitution character. However, certain mutations can create a code word for an amino acid which can signal a termination, leading to the synthesis of a smaller protein than normal [84]. However, mRNA containing such altered sequences can express to its normal length in cell-free translation systems if a special form of tRNA (e.g., suppressor-tRNA) is added, which can introduce an amino acid at the place of the non-sense codon [84]. Suppression mutations may help us to identify the genes in pathogens as these can compensate for nonsense mutations. Moreover, it is possible to identify the modified gene product by comparing the protein synthesis in suppressor-negative and suppressor-positive cells. Both temperature-sensitive and suppressor mutations are called conditional lethal mutations, as their lethality is expressed under certain conditions [52]. To date, besides humans, SARS-CoV-2 has been found in many animals, including cats, dogs, bats, goats, tigers, snakes, and pangolins. Cell-dependent mutants, also called host-range mutants, are a type of conditional lethal mutant as they can replicate normally in certain types of cells. Although the cell variability among other animals is yet to be investigated, SARS-CoV-2 has been found in many organs in humans, like the lungs, intestine, liver, heart, eye and brain. However, after mutation(s) a pathogen can develop adaptation capacity, which helps it to survive and to replicate in many cell types in diverse animals. Thus, the selection of suitable detection methods may be helpful to identify differences between natural mutations or induced mutations in a particular organism [65, 67].
Immunological selection, one of the most popularly used techniques to detect mutations in viruses, shows that surprisingly a lot of viruses have notifiable high antigen stability [52]. On the other side, interactions between the gene products also depend on the location of replication. Frequently, replication occurs in the same partition (e.g., cytoplasm or nucleus) and has a great possibility of establishing interaction between them. However, viruses budding from the plasma membrane also lead to the occurrence of mixed products, including other viruses that connect the budding-off process. In case of simultaneous infection, both viruses can be neutralized using antibodies that can act only against anyone. But, phenotypic mixing among viruses could not only be seen with closely related ones [52].

**Complementation**

Complementation between two genomes may be required for replication of certain viruses, enabling genetic robustness [77]. Sometimes, intermolecular recombination is not necessary in case of their replication in cells where parts of the virus-genome are established and expressed in the form of gene products [34]. In such a case, the missing gene function can be complemented with the aid of a gene product from integrated virus genomes [52]. But, by this phenomenon, the antagonistic effect is present in defective genomes. Besides this, it is also possible to generate mutant viruses that can change in the polymerase, thus increasing accuracy; these are known as high-fidelity mutants [8, 54]. Different viral genotypes in quasi-species can increase in spreadability. For example, coinfection of mice with wild-type and high-fidelity mutant virus enabled the high-fidelity virus to reach in mouse brain [112]. However, high-fidelity RdRp showed a decreased nucleotide misincorporation rate, while the low-fidelity RdRp has an increased nucleotide misincorporation rate [82]. Thus, the polymerase fidelity affected the accuracy of the RNA synthesis rate while also impacting the virus ability to adapt to the environment [57]. A possible mutation pathway is shown in Figure 6.

![Complementation](image)

**Figure 6.**

Complementation mutation pathway between two genomes may be possible for replication of certain viruses, enabling genetic robustness.

**Genetic robustness**

Genetic robustness refers to the capability of a genome to resist mutations caused due to genetic or environmental imbalances without a phenotypic change [8]. It is a major cause by which a virus can withstand mutations, leading to adaptation to a changing environment. 30 – 40% of virus genomes generated during infection are defective [93], suggesting that most viral genomes are not robust. In general, the genetic background affects robustness, being thus related to the viral swarm rather than individual genotypes [40]. However, neutral mutations may be needed to enable adaption and survival in an extreme environment, requiring an over robustness [28]. Factors affecting mutation frequency, including polymerase fidelity, replication mode, and host cell factors also impact robustness. For instance, it has been seen that robustness is affected by the virus variant containing synonymous mutations [73]. However, the ability to generate large numbers of genomes within individual infected cells also affects robustness as well as the innate immunity, arising from the high adaptive need at the beginning of the infection cycle. In addition, robustness can also be affected by the presence of multiple genes within the same cell, promoting recombination and consequently the occurrence of multiple mutations [73].

Viral pathogenesis and robustness may increase the virus virulence in host organisms [64], however, its converse is also true and under certain conditions [105], suggesting that it is hard to generalize how robustness affects the virus adaptability. A possible mutation pathway is pictured in Figure 7.
Mutations can trigger different phenotypic changes. For instance, mutations affecting the 3rd base in a code word may be ineffective and do not affect the virus replication, with results phenotypically silent.

**Benign relationship**

A benign relationship may be developed in case of long-time co-evolution of a virus with the host. In this period, it will not cause any disease to the host. However, after crossing from its natural host, it may cause diseases [38]. For example, simian immunodeficiency virus (SIV) strains developed a benign relationship in their natural simian hosts, however, when they crossed from its natural host sooty mangabeys into macaques (novel host), it led to acquired immunodeficiency syndrome (AIDS) [106]. SARS-CoV-2 has also found to present genomic similarity with the previous outbreaks of SARS-CoV and MERS, although both were less virulent than it.

**Genetic evolution of SARS-CoV-2: where it comes from, where it is going**

SARS-CoV-2 comes from a viral family that mainly affects animals, such as bats, pigs, mice and even beluga whales. The seven coronavirus species that infect humans (the most recent of which is SARS-CoV-2) have also been identified in various animal species. Particularly, the onset of the COVID-19 epidemic, with the first cases being reported in Wuhan, was closely linked to an agri-food market, before spreading through interpersonal transmission. All this suggests that SARS-CoV-2 is a zoonotic virus able to cross the species barrier from animal to human. Genetic studies have traced the virus pathway from animal to human hosts, tracking genetic variations of the S protein [6]. Thus, through comparing the genetic code responsible for building this protein from different coronaviruses, the SARS-CoV-2 origin can be identified. Both SARS and SARS-CoV-2 virus has a high affinity for human cells receptors, but the structure of the S protein is significantly different between the two viruses. This observation indicates that the S protein in SARS-CoV-2 appears by natural selection, through a long process of accumulation of genetic mutations in successive generations of viruses [48, 102]. An aspect that differentiates SARS-CoV-2 from other known coronaviruses is the presence of a cleavage site between two of the three subunits of S protein. This protein fragment, consisting of a series of basic amino acids (polybasic cleavage site), can be cleaved by furin, an enzyme found in many body tissues. Like the unique structure of S protein among coronaviruses, the cleavage site is different from other protein fragments with similar function in other coronaviruses (SARS-CoV does not have such a site at all). Thus, the genetic structure of SARS-CoV-2 indicates the bat as the zoonotic source of infection, as it is 96.2% identical to a coronavirus in the bat population in China. However, important fragments of S protein differ [47, 88]. The intermediate host between the bat and man is most likely the pangolin, a mammal with a body covered with scales, which is traded illegally in agri-food markets, like in Wuhan, due to its use in traditional Chinese medicine. Indeed, coronaviruses isolated from such animals share similarity with SARS-CoV-2 highly...
specific S-protein fragments responsible for cell binding, so that the accumulation of mutations further raise the infectivity of these strains [53, 120, 121]. Taken together, such observations indicate the history of SARS-CoV-2.

More recently, some studies have confirmed the zoonotic origin of the virus by identifying “missing links”: the coronavirus species that infects pangolin. By analysing tissue fragments, the genome of a coronavirus found in several samples was identified and compared with bat-specific coronavirus and SARS-CoV-2, and nucleotide sequences were similar in proportions ranging from 85.5 - 92.4%. However, the most significant aspect was the high similarity (97.4%) found in the genetic fragment corresponding to the S protein [74, 75]. Also, the fragment responsible for the interaction with the human ACE2 receptor was found identical in a proportion of 84.8% between the pangolin virus and SARS-CoV-2, compared to 80.8 - 81.4% between the bat virus and SARS-CoV-2, indicating a closer phylogenetic link [62].

SARS-CoV-2 is continuously evolving

Comparison of SARS-CoV-2 S protein with the S protein of other coronaviruses reveals another observation of epidemiological significance: the ability of SARS-CoV-2 to bind to human cells is higher than that of other coronaviruses [4]. First, the existence of the polybasic cleavage site (mentioned above), at which it can activate a ubiquitous enzyme in the body (furin) gives the new coronavirus the ability to effectively infect several tissues [1]. By cleavage, S protein fragments responsible for penetration into the cell are activated. The enzyme furin is also richly distributed in the respiratory tract and lung tissue. At the time of infection, SARS-CoV-2 and SARS attach to the same receptor of the human host (the ACE2 receptor). However, the genetic comparison reveals that the surface S protein (which interacts with ACE2) of both viruses differs significantly, with SARS-CoV-2 S protein having a higher affinity for the human receptor [61, 117]. These differences are probably due to the genetic recombination suffered by the virus in its intermediate host, pangolin [124]. An important aspect, which may indicate changes in the virus behaviour in the human body, is the presence of genetic mutations. As the virus infects more people and then multiplies, there is a possibility of mutations, leading to distinct viral strains, with different behaviours, in terms of ability to infect new patients, inter-human transmission, and to trigger the more severe clinical forms of COVID-19. To date, research indicates a low degree of genetic variability in SARS-CoV-2 [11]. However, it is necessary to perform a comparative analysis of the viral genome taken from patients from different regions and times of epidemic, to identify genetic differences. For example, by comparing 103 samples, 2 distinct strains of SARS-CoV-2 coronavirus were identified, called L and S (depending on the amino acid encoded by the different homologous regions of the two strains). Indeed, it is estimated that 70% of COVID-19 patients are infected with strain L, and 30% with the S. Reconstructing the evolution of the two strains, it has been stated that the S strain is in circulation for a long time, while the L strain appears later, evolving from strain S [53].

Given that the L strain is predominant (70%), although it appeared more recently, and, besides, has a greater genetic variability, we can deduce that it is transmitted and replicated faster in the human population [68]. This behaviour allows it to accumulate mutations with each replication, and to spread faster. The first patients diagnosed with COVID-19 in Wuhan were mainly infected with the L strain (96.3%). Subsequently, patients from the rest of China presented the L strain in a proportion of only 61.6%. This can be explained by the prompt measures adopted by the authorities: by early isolation and quarantine, the transmission of the L strain, which was faster and more aggressive, was stopped, while the S strain, which was slower, was able to achieve a higher incidence [55]. However, all such conclusions represent the results of the analysis of a small proportion of samples from the total SARS-CoV-2 infections, so that a similar analysis is needed in other regions affected by COVID-19, such as Europe or the USA. For example, a study with viral samples from Italian patients confirmed that the SARS-CoV-2 virus does not present a pronounced tendency to accumulate genetic mutations: only five viral genetic variants have been described, a low number for such a large number of replicative cycles [22].

SARS-CoV-2 mutations and the impact on potential COVID-19 vaccines development

Viruses evolve and researchers around the world are closely monitoring genetic changes that could affect the transmission of the new SARS-CoV-2 coronavirus and the effectiveness of future treatments or vaccines developed against the virus. Indeed, the continuous sequencing of the viral genome will allow monitoring of SARS-CoV-2 changes (to identify early whether the virus becomes more virulent or transmissible) and may underlie the development of diagnostic tests and a potential COVID-19 vaccine. Based on the genome, it is possible to perform the first tests on human subjects for the formulation of vaccines against SARS-CoV-2 in a few months, which would be a record development time [13]. However, both the identification and study of mutations during a viral epidemic is an essential tool for the development of effective vaccines and antiviral treatments, as well as for monitoring the virus spread in the human population.

The frequency with which natural mutations occurring is higher in the case of RNA viruses due to the specific mechanism of replication. Besides, the high replication rate (~ every 30 min) and repeated replication in host cells increase the chances of mutations. These mutations
could be favoured by natural selection or spread by chance, through genetic drift. This explains why the flu virus, an RNA virus, evolves very quickly and makes it necessary to develop an influenza vaccine every year [58]. It is unclear at this time whether the evolutionary pattern of the new SARS-CoV-2 will allow the development of a single vaccine that will provide long-term protection to all humans or whether new vaccines will be needed as virus mutations accumulate. Immunologists have begun to identify the stable genetic parts of the virus, which stimulate an immune response and appear to be identical between different strains [12]. Taken together, these stable parts will guide the development of new vaccines. Genetic mutations in the virus must be constantly monitored for several reasons [2]: (i) the genetic information of the virus can be used to construct the equivalent of a family tree of the virus (also called a phylogenetic tree) as this family tree can provide data on the geographical spread of the virus, while can also clarify whether the virus has passed from an animal to a human host; (ii) knowledge of the genetic mutations accumulated by the virus is a valuable resource for scientists developing vaccines; (iii) if a link between changes in viral genetic data due to mutation and how the virus causes the disease could also be established, with this monitoring providing information about changes in the virulence of the disease. There is a well-established international network of virologists who study the flu virus and share data on its genetic changes through online databases, such as the Global Initiative on Sharing All Influenza Data (GISAID). GISAID allows not only to get data on the distribution of genetic information but also the comparison of genetic information from various strains of the new coronavirus. These data will assist in identifying the mutations and will also help in the development of vaccines and treatments while allowing the identification of strains from different geographical areas and can be used to track the origin of coronavirus infections [25, 81, 111, 114]. To date, there is no evidence that current SARS-CoV-2 mutations affect the way the coronavirus is transmitted, its ability to cause infections or the disease severity. However, the study of such mutations represents a powerful tool to help in finding solutions for the SARS-CoV-2 epidemic [123]. Researchers have identified various SARS-CoV-2 mutations. The first one to be identified some time ago was D614G, which is a mutation in the SARS-CoV-2, occurring as a result of a change in amino acids in the S protein that the virus uses to enter human cells. The S protein is important because it binds to cell receptors and mediates the fusion of the viral envelope with the host cell membrane, allowing the virus to penetrate and infect the host. The mutation changes the amino acid in position 614, from D (aspartic acid) to G (glycine), the virus strain being transformed from D614 to G614 [86]. Nonetheless, and despite this mutation has become dominant in the initial times of the pandemic, this mutation alone has not been shown to present distinct antigenicity. However, the SARS-CoV-2 virus has undergone several mutations and, by the end of 2020, the B.1.1.7 and B.1.351 variants appeared in South East England and in the Eastern Cape, South Africa, containing, respectively, 8 and 9 peak mutations in addition to D614G along with other mutations, which quickly became dominant and spread to more than 50 countries quickly. More recently, the variant known as P.1 or 501Y.V3 has emerged in Brazil and is spreading rapidly to many other countries [15, 115]. Taken together, all these variants and others that are emerging are raising a growing concern due to their rate of transmissibility, and consequent genetic alterations that can compromise the effectiveness of vaccines and drugs currently in use and under development to fight this infectious disease.

Conclusion and future perspectives

Overall, in terms of organisms’ evolution, mutations occurring at regular time intervals are very important since this process helps to survive and to adapt to the changing environments and maintaining the general well-being. Indeed, the coexistence of organisms also needs to be balanced for the normal functioning of the ecosystem. However, when abnormal mutations occur, the survival of any pathogen and increased susceptibility to infection may be observed, along with adverse effects to other animals or fauna of the ecosystem that are not desirable. According to genome sequence reports around the world, the mutation potential of SARS-CoV-2 is greater than any epidemic or pandemic that has ever appeared. The virus may have been able to reach all parts of the world in a very short time due to its high mutation potential and high infectivity levels, leading consequently to the loss of millions of lives. The current pandemic (SARS-CoV-2) outbreak suggests that it has a higher mutation rate, making challenges in the development of suitable drugs and vaccines, as the virus may rapidly gain resistance to the newly formulated therapeutic strategies. Despite the current efforts and limitations, we believe, clear know-how of effects and results of genetic variability might open new ways for limiting its spread, so that a continuous sequencing of the viral genome will allow proper monitoring of changes in the SARS-CoV-2 genome organization and may serve as the basis for the development of effective diagnostic tests and vaccines.

Conflict of interest

The authors declare no conflict of interest.
References


Genetic and molecular analyses of spontaneous mutants

Graham RL, Baric RS

mutations. without reversion of exoribonuclease


85. Pinkert S, Kopp A, Bruckner V, Fechner H, Beling A, Single-Point Mutations within the Coxackie B Virus


