Journal of Virology

Computational Biology | Minireview

Evolutionary deletions within the SARS-CoV-2 genome as signature trends for virus fitness and adaptation

Pedro Miguel Carneiro Jeronimo,[1] Cleber Furtado Aksenen,[1] Igor Oliveira Duarte,[1] Roberto D. Lins,[2] Fabio Miyajima[1]

AUTHOR AFFILIATIONS See affiliation list on p. 11.

ABSTRACT Coronaviruses are large RNA viruses that can infect and spread among humans and animals. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for coronavirus disease 2019, has evolved since its first detection in December 2019. Deletions are a common occurrence in SARS-CoV-2 evolution, particularly in specific genomic sites, and may be associated with the emergence of highly competent lineages. While deletions typically have a negative impact on viral fitness, some persist and become fixed in viral populations, indicating that they may confer advantageous benefits for the virus's adaptive evolution. This work presents a literature review and data analysis on structural losses in the SARS-CoV-2 genome and the potential relevance of specific signatures for enhanced viral fitness and spread.

KEYWORDS evolutionary mechanisms, structural mutations, SARS-CoV-2, viral adaptation, deletions

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the viral agent responsible for causing the coronavirus disease 2019 (COVID-19), and it was first identified in Wuhan (Hubei Province, China) (1). COVID-19 was officially declared a pandemic in March 2020 by the World Health Organization, and it became the deadliest pandemic of the 21st Century, accounting for over 765 million cases and 6.9 million deaths worldwide (2).

Coronaviruses are a family of RNA viruses with a potential to infect and spread in various bodily systems, including respiratory and gastrointestinal tracts, as exemplified by certain coronaviruses like porcine epidemic diarrhea virus, which primarily affects the gastrointestinal tract. Coronaviruses are positive-sense single-stranded RNA viruses (+ssRNA) belonging to the *Coronaviridae* family, which includes some of the RNA viruses with the largest genomes, ranging from 27 to 31 Kbp (3). This family is divided into two subfamilies, the *Coronavirinae* and the *Torovirinae*. The *Coronavirinae* subfamily contains four genera, i.e., Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus, with SARS-CoV-2 belonging to the Betacoronavirus genus (4).

Over time, evolutionary mutations accrued by the SARS-CoV-2 genome resulted in enhanced levels of transmission. Successive cycles of virus transmission and disease recurrence led to the emergence of highly competent lineages, namely variants of concern (VOCs) (5). Most of these mutations correspond to base exchanges, in functional terms amino acid substitutions, due to their higher probability to emerge and more predictable impact on the virus function. Conversely, deletions are comparatively less frequent since they are more likely to result in loss of function events, including erroneous translation of proteins, unless no codon frameshifts are introduced in the reading frames. One such event relates to the almost simultaneous emergence of VOCs Alpha, Beta, and Gamma, all of them possessing a characteristic 9-bp deletion at position 11,288–11,296 in the NSP6 gene within the polyprotein ORF1a (6), which then became widespread amongst all variants except VOC Delta. Moreover, since its emergence in

Editor Suchetana Mukhopadhyay, Indiana University Bloomington, Bloomington, Indiana, USA

Address correspondence to Fabio Miyajima, fabio.miyajima@fiocruz.br.

The authors declare no conflict of interest.

Published 13 December 2023

late 2019, SARS-CoV-2 evolution has shown a tendency toward genome size reduction, with deletions being a notable occurrence, as seen for other RNA viruses (7, 8), as opposed to insertion events. These deletions, including in other SARS-related viruses, have been reported to occur more often within particular genomic sites, indicating this phenomenon is a result of the selective pressure in the evolutionary process, possibly influenced by specific gene locations, avoiding key motifs of structural proteins, and by changing host-virus protein interactions. (9). In most cases, deletions negatively affect viral fitness, while a fraction of them persist and eventually fix within viral populations, suggesting either a neutral or a selective advantage. Mahmoudabadi and collaborators proposed a simple strategy to measure the number of ATP and ATP-related molecules (P) needed to "build" a virus. They used the T4 phage and influenza as examples to estimate the number of 50 P per nucleotide in RNA or DNA synthesis and 36 P per amino acid in protein synthesis. Furthermore, they conclude that the burst size, the number of virion particles that emerge from each infected cell, is a function of the needed energy for virus infection (10). Indeed, it has been suggested that the reduction in the genome size might potentially influence adaptative evolution by decreasing the energy required for viral replication and increasing the efficiency of mutants (11). Evolutionary trends may be objectively studied over time and space through comparative analysis of structural changes and base exchanges from representative genomes, thus providing new insights into the evolutionary patterns of SARS-CoV-2 with respect to its sustained transmission and ability to infect/reinfect the host (12).

Evolution, particularly of RNA viruses, is a very dynamic process, often leading to rapid changes over relatively short periods. Recurrent infections coupled with constant virus introduction to previously immunized hosts and the increasing use of antivirals, despite their pharmacological benefits, further drive the co-evolution of the virus. The diminishing effectiveness of vaccines may, at least in part, account for this stochastic yet orderly process (13). Specific genomic signatures that become fixed as part of a continuous process, or a convergence of independent events, are net products of a complex seesaw interaction between pathogens and the host. SARS-CoV-2 mutations, such as S:D614G, are a clear example where variants with specific alterations conferring selective advantage, rapidly increase their frequency and become dominant (13).

This work presents a comprehensive literature review on structural losses accrued by the SARS-CoV-2 genome, coupled with a pattern and cause-effect investigative analysis supported by data available from GISAID EpiCoV (Table S1), a large repository data set maintained by the scientific community. Here, we data mined and analyzed over 5,116,266 sequences to comparatively assess the patterns of deletion events throughout the virus genome, at the same time determining the potential relevance of specific signatures accounting for enhanced viral fitness and the spread of VOCs, while depicting their trajectory from uprise to decline.

SARS-CoV-2 MUTATION RATES AND GENOME SIZE REDUCTION TREND

In any given biological being, mutations within a genome are predominantly substitutions, due to an intrinsic error associated with their replication apparatus. In the case of RNA viruses, mutation rates can be over a million times greater than their corresponding hosts since they have limited ability to correct replication errors as reviewed by Lauring et al. (14). Hence, under selective pressure, their survival and adaptability are reliant on their ability to overcome alterations of detrimental nature and give rise to mutants possessing competitive advantage to enhance their fitness. Unlike small DNA viruses, RNA viruses encode their own replication machinery and are thus able to optimize the mutation rate for their fitness (15). The size of the RNA genome influences directly on the mutation susceptibility as during the replication process of a large genome, its polymerase has a greater probability of misplacing a nucleotide, making larger RNA viruses likely to have evolved to exhibit a proofreading activity (3). On the other hand, genome size would be limited as verified by the error catastrophe hypothesis (16). The error catastrophe hypothesis suggests that the genome size of a virus may be limited

due to accumulated errors during viral replication. In other words, as the genome size of a virus increases, the probability of errors in the replication of genetic material also increases, leading to loss of viability and decreased replication capacity, resulting in an "error catastrophe." This hypothesis proposes that viruses face an evolutionary limitation in terms of genome size due to the detrimental effects of accumulated errors in genetic replication. Indeed, genome size has been shown to have a negative correlation with mutation rate in viruses regardless of the type of genome or replication proofreading capacity (17).

The intense mutation rate of RNA viruses in a single host results in a population of variant genomes, called quasispecies, that are susceptible to competition and selection, as mutations that result in the increase of replicative capacity are more likely to outgrow others and contaminate new hosts (18–20). In the context of specific viral lineage, the acquisition of mutations is a phenomenon that affects the entire population. In simpler terms, all viruses within this lineage undergo mutations, which contribute to the overall genetic diversity. Furthermore, this statement emphasizes a crucial concept: viral genomes that have mutations enabling them to thrive within hosts and, subsequently, infect new hosts have already experienced a selective pressure known as "adaptive pressure." This indicates that these specific viral variants have undergone genetic changes that provide advantages for their survival and replication within their host environments. In essence, the viruses that successfully spread have already been influenced by natural selection to better adapt to their host organisms (20). In this way, although some deletions affect large extensions of genes as a result of frameshift, the main lineage that infects a patient is frequently, to some extent, a strain with enough fitness to overcome others and infect new hosts (20).

SARS-CoV-2 genome encodes a nonstructural protein (nsp) with 3′ → 5′ exoribonuclease activity, the nsp14, which ensures high-fidelity replication, as observed for other coronaviruses like SARS-CoV and MERS-CoV (21). Even though the nsp14 proofreading activity enables the expansion of the viral genome in the order *Nidovirales* (22), it does not appear to be able to protect the genome from loss of material associated with the discontinuous transcription process, common in ssRNA viruses and responsible for increase in deletion occurrence (23). Proofreading activity alone cannot fix deletions, and this may lead to adaptive evolution in SARS-CoV-2 (24). In fact, the large genome of coronaviruses tolerates numerous amount of mutations, including structural changes, evidencing its high plasticity (25). This fact can be observed from the size of the SARS-CoV-2 genomes deposited in the GISAID EpiCoV database. According to Outbreak lineage Covid using GISAID EpiCoV database, deletions are significantly involved in the constitution of new variants and have contributed to the emergence of Alpha, Beta, Gamma, Delta, and Omicron (BA.1) variants, which are, respectively, 18, 18, 5, 12, and 39 nucleotides shorter than the SARS-CoV-2 reference genome from Wuhan (Fig. 1).

To generate a data set for further analysis, high-quality sequences deposited in the GISAID EpiCov database, accessed on 21 August 2022 (n = 5,116,266), were filtered using the parameter "QC status = good" by command line of the Nextclade CLI v.2.4.0 program. A gradual decrease in their size as a function of time is shown in Fig. 2, evidencing the loss of elements as part of the natural evolution of SARS-CoV-2. The Spearman rank correlation coefficient [ρ (rho)] shows a weak and negative correlation −0.31 [95% confidence interval (CI): −0.31; −0.31] at a 95% significance level. The plot shows the correlation of the nucleotide variation (subtraction between inserted and deleted nucleotides in a given sample) over time.

Deletions can cause frameshifts and are thus likely to be deleterious and rapidly disappear from a lineage, generating changes that extend into the genome, and affecting genes after the frameshift. For this reason, larger structural deletions in SARS-CoV-2 genomes have been observed in accessory proteins (i.e., ORF3, ORF6, ORF7a, ORF7b, ORF8a, and ORF8b), in part owing to the lower selection pressures on these respective genes as these are deemed as not essential for viral replication (23, 26).

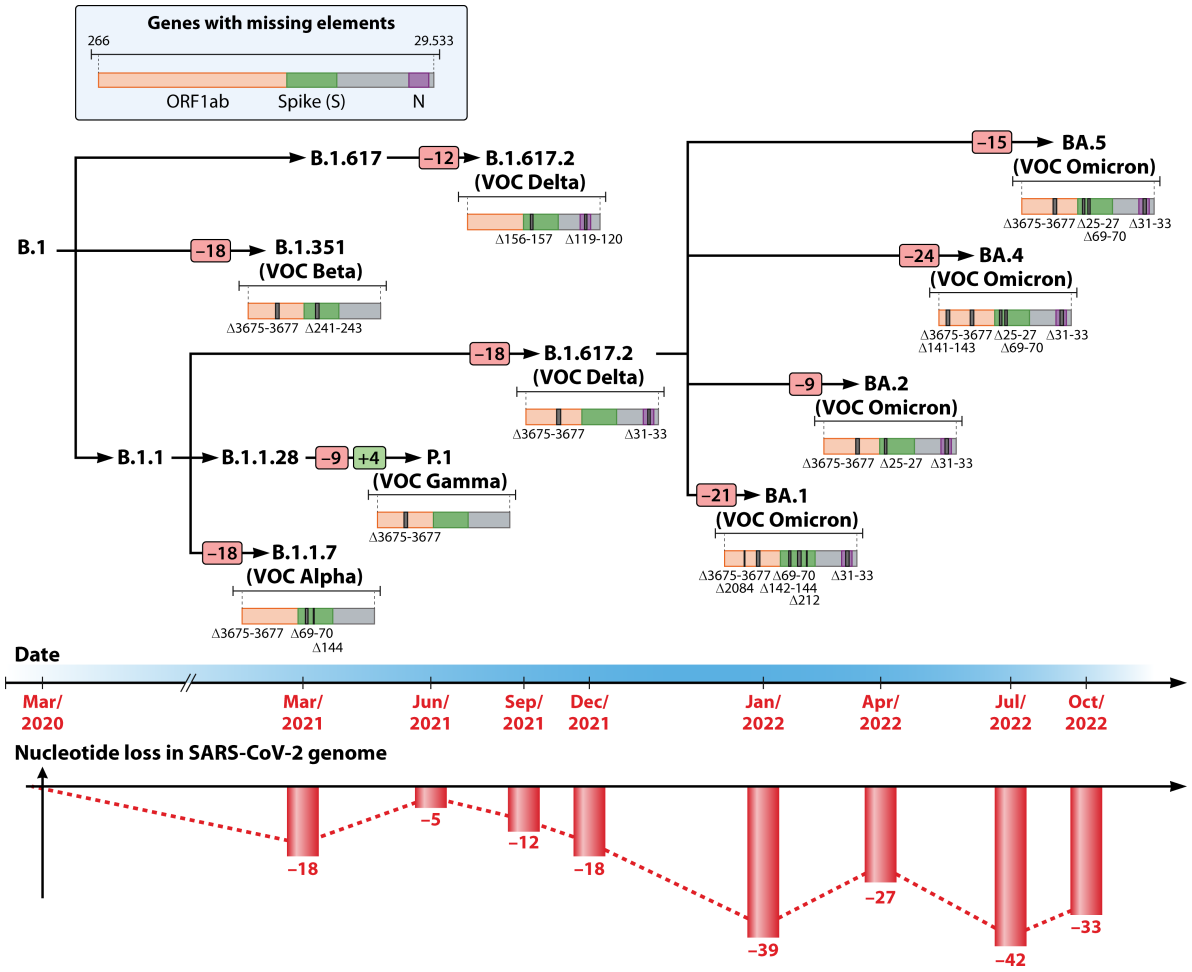


FIG 1 Loss of elements in VOCs of SARS-CoV-2 over time. The schemes represent the different VOCs and their deletion events at the peak of sequenced cases. The left-hand side of each scheme shows the nucleotide acquisition and loss scores, while the bottom bar plot summarizes the scores over time.

A study conducted by Nguyen and colleagues about genomic mutations of SARS-CoV-2 indicated that among more than 5,500 sequences, no structural alterations (either insertions or deletions) were found within the genes encoding structural protein E (envelope) or M (membrane). This leads to the conclusion that these genes are relatively stable and probably more tolerant to deletion; hence, they could be targeted for vaccine and drug development (27).

Even though it is unusual to reconstruct the evolutionary pathway of deletions, it was first reported on an accessory gene by Panzera and collaborators, where they detected a single (Δ12) and a double deletion (Δ12 + Δ68) variant of SARS-CoV-2 within the N.7 lineage derived from the Brazilian B.1.1.33 lineage (12). Both deletions were restricted to a small region of the ORF7a gene and instead of generating a new ORF, it maintained the functionality of both ORF7a and ORF7b proteins, implying that the loss of genetic material may not always impair virus viability. This discovery underscores the significance of deletion events in adding the genetic variability of SARS-CoV-2 as it spreads and transmits within the human population. Moreover, this pattern of loss being more frequent than incorporation of genomic material is not exclusive to SARS-CoV-2 but is also present in other *Betacoronavirus*, including SARS-CoV and SARS-like viruses and other RNA viruses (7–9).

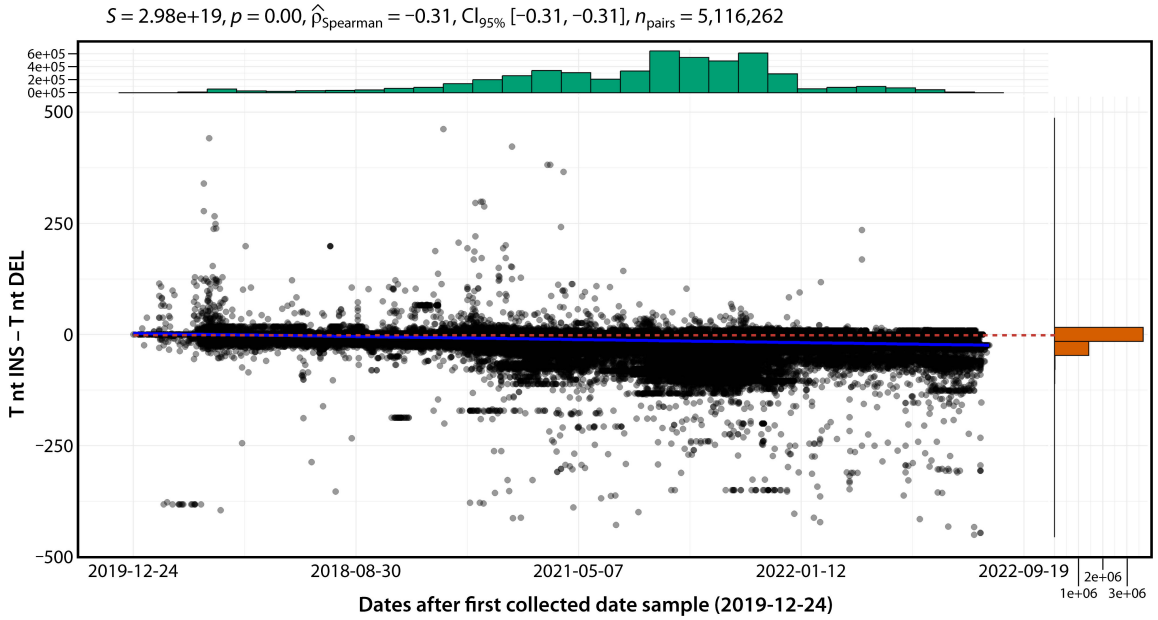


FIG 2 SARS-CoV-2 genome shrinkage over time. This figure illustrates the reduction in nucleotide variation (subtraction of insertion and deletion events) in the SARS-CoV-2 genome over time. The central figure displays the worldwide nucleotide variation over time, with the dotted red line representing no nucleotide variation and the blue line representing the Spearman linear regression. The histograms on the margins show the frequency of dots in the scatter plot.

POTENTIAL FACTORS UNDERPINNING DELETION OCCURRENCE WITHIN SARS-CoV-2 GENOME

Apart from replication errors, viral mutations can occur as a result of editing the genetic material through host-enzymatic modification, spontaneous events, such as chemical reactions, or in favor of specialized viral genomic elements, such as recombinations or genetic shifts (28). Chrisman and collaborators suggested that deletions in the SARS-CoV-2 genome occur as an artifact during the viral replication, in which the RNA-dependent RNA polymerase dissociates from its template and re-associates in a different locus (29). Pereira also suggested that deletions may occur in hairpin regions formed in the RNA secondary structure (30). These hypotheses are supported by other studies, which indicate that the RNA secondary structure, respectively, in retrovirus and Cucumber mosaic virus, can modulate the pace and kinetics of the RNA polymerase, increasing the chance of slippages leading to reading errors (31, 32). Kemp and colleagues demonstrated that the loss of two amino acids (H69 and V70) as well as other reported N-terminal domain deletions within the Spike protein, in addition to increasing its infectivity, occurs at a critical position in overlapping loop structures, in which RNA polymerase activity is often compromised (28). Similar to most viruses from the *Nidovirales* order, SARS-CoV-2 is prone to a common characteristic, which is the discontinuous viral transcription process (23). Hence, the frequency of structural gaps can be incremented by the discontinuous RNA synthesis of the polymerase machinery, which will likely remain uncorrected by the nsp14-exoribonuclease proofreading activity (12).

McCarthy and collaborators identified the incorporation of gaps in the N-terminal domain of the S glycoprotein as an evolutionary pattern defined by recurrent deletions that alter defined antibody epitopes (24). These deletions may converge, and their combined properties result in greater resistance against antibodies elicited by previous exposure. Panzera and collaborators reported cases of immunocompromised patients with long-term infections displaying recurrent deletions in the N-terminal domain of the S glycoprotein (12). This clinical course reinforces the ability of SARS-CoV-2 to evolve through persistent infections, especially in immunocompromised patients.

FROM RANDOM DELETIONS TO EVOLUTIONARY EVENTS

The SARS-CoV-2 genome has experienced numerous mutations during the ongoing pandemic, with many initially considered as random events. However, not all mutations are created equal. Structural variations, particularly deletions within coding regions, often lead to more profound alterations in the amino acid sequences, potentially impacting the resultant protein's function. These deletions can enhance viral fitness and facilitate escape from host immune responses, as illustrated in references (33–35) (Table S2). Some of these seemingly random deletions have become fixed within viral populations, indicating a shift from randomness to functional relevance.

Moreover, the virus has a mechanism to potentially recover information lost to deletions. The recombination process, enhanced by the activity of the NSP-14-ExoN protein (36), serves as a recovery method for SARS-CoV-2 strains that have undergone extensive deletion events.

This section aims to elucidate the journey of deletions from random occurrences to potentially functional and evolutionary significant events, setting the stage for a deeper understanding of the virus's adaptive strategies amidst a pandemic.

Impact of deletions on structural proteins

Deletions occurring within key proteins accounting for cell invasion, such as the Spike protein from SARS-CoV-2, could alter host-pathogen interaction by enabling pathogenesis to occur at additional sites of the body, as seen for feline coronavirus and porcine respiratory coronavirus, thus resulting in increased infection competence and potentially a more severe course of disease (37, 38). This is possible because viral tropism is influenced by the tissue expression of target receptors (23). Dolskiy and collaborators produced transgenic mice with the expression of the human ACE2 receptor in different tissues and reported the presence of significant viral load in the lung, brain, heart, and intestine (39). Interestingly, virus transmission ability/infectivity can be increased by critical deletions of the non-RBD S1 domain of the Spike protein, at the same time promoting immune evasion as demonstrated by decreased performance of neutralizing assays based on antibodies elicited by wild-type variants (40). Deletion such as S:del69/70, although does not account for immune evasion, increases cleaved S2 and spike infectivity (41). The S:del156/157 deletion and the S:del144 map to the same surface of the spike protein, the NTD domain. The deletion of these regions is hypothesized to be involved in disrupting the binding of monoclonal antibodies in neutralization assays (24, 42, 43). For the other structural proteins, E, M, and N, there is minimal to no experimental influence of deletion in these regions. Multiple deletions in the nucleocapsid protein have been reported by Rahman and colleagues in samples from several countries worldwide between March and May 2020(44). The authors predicted, with an *in silico* approach, that the changes on the surface exposed near the C-terminal domain of the encoding gene can have a significant impact on the virus pathogenesis and nucleocapsid-RNA interaction (45). Furthermore, a 12-bp deletion in the E protein accounts for both increased S protein content and upregulation, mainly, of IL-6, CSF2, and CXCL10 cytokines, and a higher level of E-selectin and PTX3 when compared to the wild-type strain (46, 47). For the M protein, however, no experimental or *in silico* data about the influence of deletions were found.

Effects of deletions on accessory and non-structural genes

Studies, on specific deletions of the NSP1 gene, suggest modifications in the protein tridimensional structure (secondary and tertiary structures) and lower viral load and serum IFN-β (48, 49). NSP2, which is suggested to be responsible for binding to RNA and linking viral transcription to viral translation, also presented a deletion of three nucleotides at the genomic position 1,607, without a predicted functional effect in the viral fitness (50). This deletion was frequent in the first months of the pandemic and circulated in SARS-CoV-2 sequences found in population clusters from the Netherlands and China (51).

Silvas et al. (51) generated recombinant SARS-CoV-2 constructs by a reverse-genetic system's approach to provide insights into the contribution of accessory proteins to virus competence (52). Each of the constructs had a deletion in either, ORF3a, ORF6, ORF7a, ORF7b, or ORF8 genes. The study found that ORF proteins contribute to early dissemination and formation of detectable viral plaques, in which viruses lacking particular ORFs (ORFΔ), namely ORFΔ3a, ORFΔ7a, ORFΔ7b, and ORFΔ8 produced significantly smaller plaques than their corresponding wild-type forms. The strong correlation between variations in plaque morphology and size indicates that ORF deletions affect essential virus function and have an impact on its dissemination, transmission, and fitness. For example, the deletion of ORF3a decreases SARS-CoV-2 virulence, indicating that this ORF is a major contributor to viral pathogenesis, with an association with lung physiopathology (52). Similarly, nucleotide deletions within ORF8 genomic regions have been linked with milder infection or yet a lower post-infection inflammation, suggesting its translation is relevant for disease severity (53). Conversely, recent studies have shown that the removal of the ORF8 gene found in some SARS-CoV-2 lineages does not significantly affect mortality rates, with a death rate like the one observed with wild-type variants (52). Furthermore, a previous *in vitro* study showed that the deletion of the ORF8 gene impaired the replication potential of SARS-CoV (54). By contrast, an independent work on SARS-CoV-2 reported a distinctive finding, where they showed a higher replicative fitness *in vitro* associated with a 382-nucleotide deletion comprising both the 3′ end of the ORF7b gene and the transcription regulatory sequence of the ORF8 gene, which resulted in obliteration of its transcription start site (55). Furthermore, a study reported that subjects with this reported 382-nucleotide deletion displayed milder disease manifestation compared to the ones infected by SARS-CoV-2 wild-type variants (56).

During the SARS-CoV epidemics in 2003, a 29-nucleotide deletion in ORF8 resulted in two truncated smaller proteins, namely ORF8a and ORF8b (54). This pattern might result in new proteins that retain their natural function or that can interact with other host cell molecules.

Quéromès and colleagues (56) identified two deletions in ORF6 (a 34-base and a 26-base deletion), resulting in frameshift alterations (57). The sequences with the 34-nucleotide deletion were found to be associated with nosocomial transmission. Interestingly, despite these deletions, no differences in clinical presentation were observed, and the replication kinetics *in vitro* were similar to those of the reference strain. These findings suggest that ORF6 may not be essential for viral replication and transmission, and the SARS-CoV-2 genome may be adapting to new selective pressures without significant implications for current strain transmission.

ORF7a gene, on the other hand, has been subject to some of the largest deletion events within the SARS-CoV-2 genome. One of these deletions refers to a 227-base deletion, resulting in the fusion of the ORF7a with the ORF8 with consequent loss of its function (58). Another one was reported to be as extensive as 892 bases, thus excluding entirely the gene cluster encoding for ORF7a, ORF7b, and ORF8, though the authors observed that the resulting lineages predicted similar viral load compared to the data observed from samples infected with the wild-type variant (59). These findings are supported by other studies indicating that these genes are not essential for viral replication (23, 26), although more recent studies have increasingly shown the importance of this group of genes, particularly ORF7a, for immune evasion and the modulation of host response, particularly inflammation and cell immunity (60).

EXPLORATORY GLOBAL ANALYSIS OF DELETIONS IN SARS-CoV-2 GENOMES

Deletion events across the genome are shown in Fig. 3. These data allow us to identify that ORF1ab, S, ORF7a, ORF8, and N are the genes that have the highest absolute frequency of deleted nucleotides. The high frequency of some regions is expected for different reasons. Among them, independent fixation in multiple VOCs, such as deletion 11,288–11,296 initially present in VOC Alpha and reemerging in VOC Omicron, present

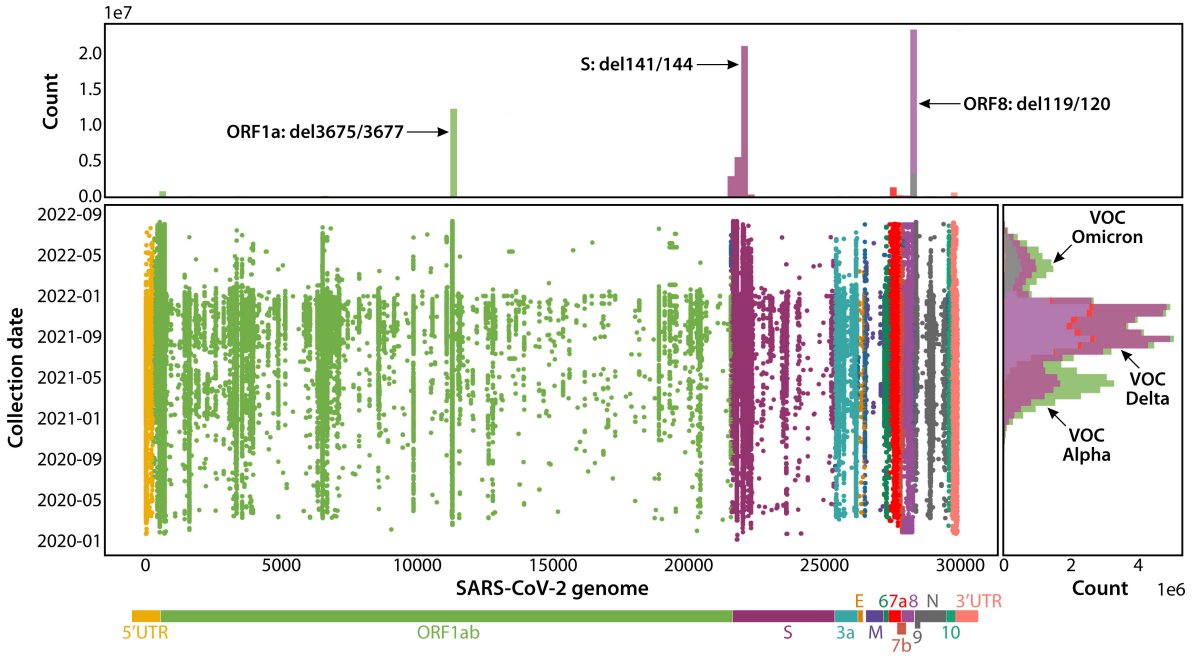


FIG 3 Deleted nucleotides throughout high-quality SARS-CoV-2 genome submitted in GISAID until 21 August 2022. The central scatter plot shows the distribution of deletions across the genome. The marginal histograms display the absolute counting of nucleotide deletions in a given gene and at a specific moment in time, in prevalence at Alpha, Delta, and Omicron VOCs. The plots located in the upper-right corner offer more detailed information about genes across the genome, as well as the deletions that occur most frequently (ORF1a:del3,675/3,677; S:del141/144 and ORF8:del119/120).

in NSP6, within ORF1ab; and deletions 21,765–21,770 and 21,992–21,994 in VOC Alpha and deletion 22,029–22,034 in VOC Delta, all in the Spike protein gene. This bias favors the prevalence of these regions, disappearing and reappearing when one VOC overtakes another, and so they end up being sequenced more frequently.

Figure 3 also shows a reduction in diversity of deletions in ORF1ab and S for VOC Omicron, as of January 2022. Furthermore, the histogram on the right of Fig. 3 shows three peaks with different compositions of regions, which correspond to the signatures of deletions of the most sequenced VOCs in each period: Alpha, Delta, and Omicron. Nevertheless, it is noteworthy that deletions that are recurrent but not fixed result in greater diversity of deletion types, which in turn affects genes where the adaptive cost is relatively lower than in proteins such as Spike. This phenomenon took place in SARS-CoV-2 ORF7 and ORF8 accessory genes, as can be seen in Fig. 4 and 5.

It is readily apparent that in non-accessory proteins such as Spike (Fig. 6), deletions are short and extremely targeted within a specific region (N-terminal domain), whereas in ORF7a, for example, deletions are larger and can be seen over almost the entire length of the gene, although they are also present in a specific domain (ectodomain).

In addition, evaluating the profile of mutations present in ORF1ab that impact the RNA-dependent RNA polymerase (NSP12), NSP7, and NSP8 proteins, it was found that, while for VOC Omicron (n = 370,623) there is no mutation in NSP12 with a frequency greater than 1%, mutations NSP12:G671S (nonpolar to polar amino acid) and NSP12:P323L (nonpolar to nonpolar amino acid) were found in approximately 99% and 100% of Delta (n = 2,938,828) and Alpha (n = 887,403), respectively (Table S3). The existing body of literature presents a notable divergence of viewpoints concerning the ramifications of mutations within NSP12 on its activity. In the pursuit of elucidating this matter, Kannan and colleagues (60) proposed a compelling notion that the NSP12:P323L mutation potentially augments the affinity between NSP12 and NSP8. This proposed enhancement, in turn, holds the potential to reinforce NSP12's processivity. Contrarily, Pachetti and collaborators (61) postulated an alternative perspective that the NSP12:P323L mutation, localized within the interface subdomain, could cause

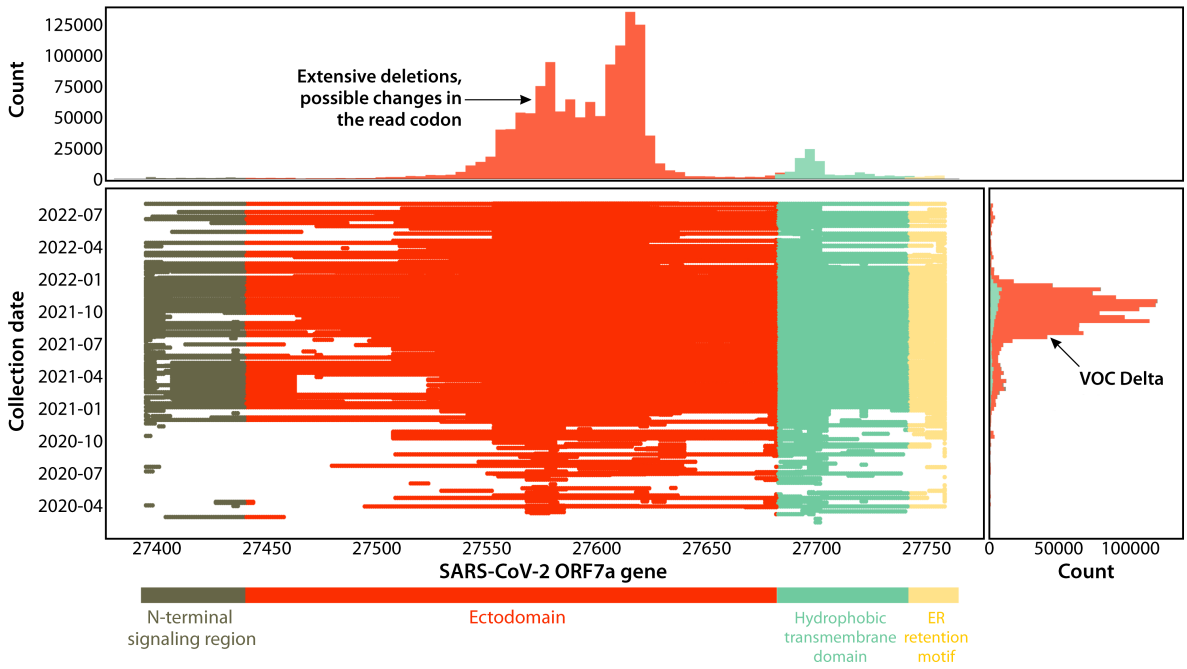


FIG 4 Deleted nucleotides throughout high-quality SARS-CoV-2 gene ORF7a submitted in GISAID until 21 August 2022. The central scatter plot shows the distribution of deletions across the regions of the gene. The marginal histograms display the absolute counting of nucleotide deletions in a given gene and at a specific moment in time, in prevalence at VOC Delta. The plots located in the upper-right corner offer more detailed information about regions across the gene, as well as the large region of deletions occurring most frequently.

disruptions in protein-protein interactions, consequently exacerbating the mutation rate. Similarly intricate is the discourse surrounding the NSP12:G671S mutation. Kim and colleagues (62) propound the intriguing hypothesis that this mutation serves a

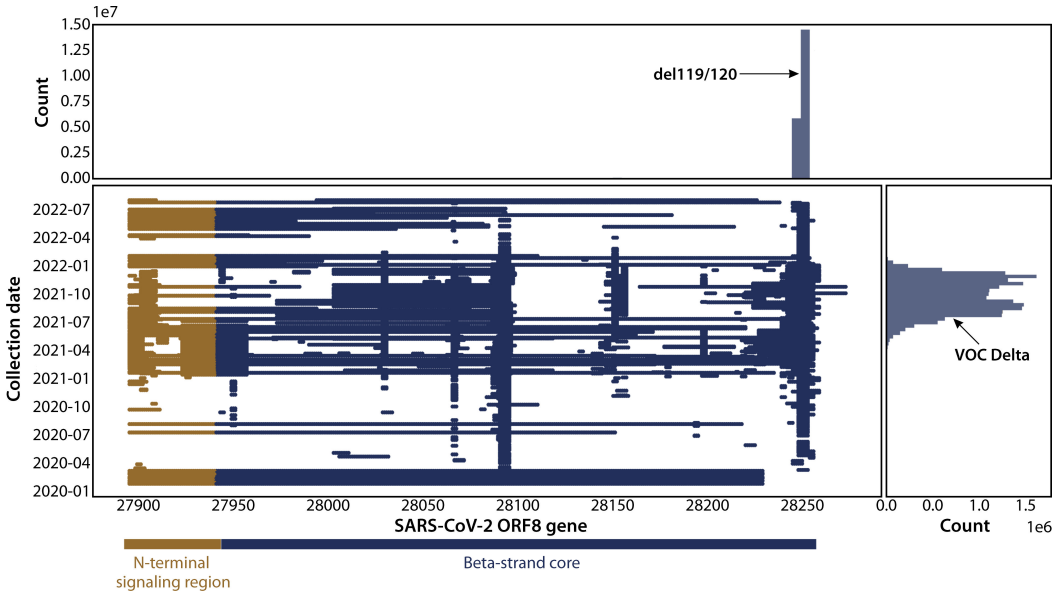


FIG 5 Deleted nucleotides throughout high-quality SARS-CoV-2 gene ORF8 submitted in GISAID until 21 August 2022. The central scatter plot shows the distribution of deletions across the regions of the gene. The marginal histograms display the absolute counting of nucleotide deletions in a given gene and at a specific moment in time, in prevalence at VOC Delta. The plots located in the upper-right corner offer more detailed information about regions across the gene, as well as the region of deletions occurring most frequently (ORF8:del119/120).

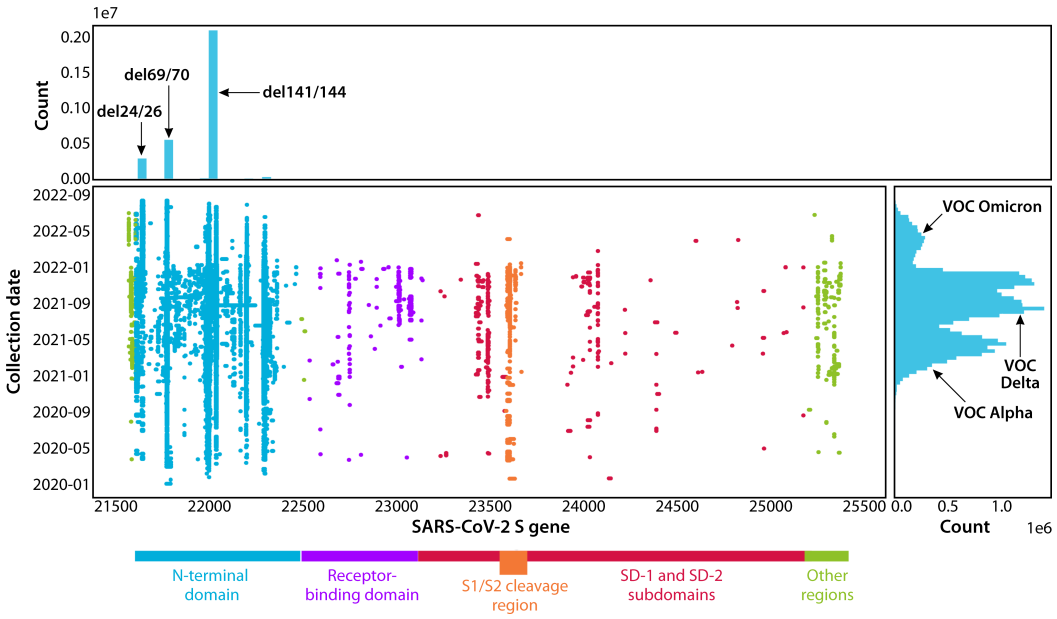


FIG 6 Deleted nucleotides throughout high-quality SARS-CoV-2 gene S submitted in GISAID until 21 August 2022. The central scatter plot shows the distribution of deletions across the regions of the gene. The marginal histograms display the absolute counting of nucleotide deletions in a given gene and at a specific moment in time, in prevalence at Alpha, Delta, and Omicron VOCs. The plots located in the upper-right corner offer more detailed information about regions across the gene, as well as the region of deletions occurring most frequently (S:del24/26, S:del69/70, and S:del141/144).

stabilizing role within the replication complex. This perspective is countered by Wang and collaborators (63), who advocate for an opposing interpretation. They argue that the NSP12:G671S mutation potentially poses a destabilizing influence on the replication complex (64). Consequently, there arises a compelling need for empirical investigations aimed at understanding whether these mutations also bear the capacity to influence the incidence of nucleotide deletions. Deletions have been present since January 2020, appearing independently and periodically in different sublines, and, when occurring in the protein responsible for cell entry, the Spike protein, could alter the relationship between the virus and the host by causing more severe disease, acting in a different site of the body (37, 38).

Conclusion

Deletions represent a mechanism of genetic diversity little explored in the literature, but it has the potential to help understand the function of genes considered at first to be non-essential for virus replication. Deletions in a viral genome can result in multiple effects. Accessory proteins, such as ORF3a and ORF8, have been linked to lower virulence and milder infections in SARS-CoV-2. Although not essential for virus replication, accessory genes are reported to play a role in virus spread and infectivity. These regions are characterized by highly flexible evolutionary processes, indirectly indicating the plasticity of the virus to continue adapting and acquiring new functions. Conversely, fixed deletions in structural proteins, such as the Spike protein, are typically found in specific regions, such as the non-RBD S1 domain, and are often associated with greater transmission capability and outbreaks. In SARS-CoV-2, this type of mutation is recurrent and appears to be related to the ability of the RNA-dependent RNA polymerase to remain coupled to the genetic material. In this case, deletions in accessory genes are more evident due to the lower associated adaptive cost. Punctual deletion patterns are part of the evolutionary history of SARS-CoV-2 and influenced the generation of the main variants of concern, which reinforces the need for more comprehensive studies on the tendency of viral proteins to lose elements, maintaining or improving their function.

ACKNOWLEDGMENTS

We are thankful to the GISAID database for making this work possible.

This work was supported by the Oswaldo Cruz Foundation under the Genomics Network umbrella and the Ministério da Saúde, Brazil.

P.M.C.J., C.F.A., and I.O.D. wrote the paper and designed the figures. All authors read the final manuscript, and F.M. and R.D.L. approved it for publication.

AUTHOR AFFILIATIONS

[1]Fiocruz Genomic Network, Oswaldo Cruz Foundation (FIOCRUZ), branch Ceara, Eusebio, Brazil
[2]Fiocruz Genomic Network, Oswaldo Cruz Foundation (FIOCRUZ), branch Pernambuco, Recife, Brazil

AUTHOR ORCIDs

Pedro Miguel Carneiro Jeronimo http://orcid.org/0000-0002-2509-5822
Cleber Furtado Aksenen http://orcid.org/0000-0002-5297-7505
Fabio Miyajima http://orcid.org/0000-0002-1347-4825

AUTHOR CONTRIBUTIONS

Pedro Miguel Carneiro Jeronimo, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft | Cleber Furtado Aksenen, Formal analysis, Writing – original draft, Writing – review and editing | Igor Oliveira Duarte, Conceptualization, Writing – original draft | Roberto D. Lins, Supervision | Fabio Miyajima, Investigation, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The data set supporting the conclusions of this article is available in the GISAID repository, https://doi.org/10.55876/gis8.230510zx.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Table S1 (JVI01404-23-s0001.pdf). Genomic data availability and snapshot (EPI_SET_230510zx).
Table S2 (JVI01404-23-s0002.xlsx). Overview of key deletions investigated.
Table S3 (JVI01404-23-s0003.xlsx). Frequency of mutations of VOCs Alpha, Delta, and Omicron in non-structural proteins 7, 8, and 12.

REFERENCES

1. Chavez S, Long B, Koyfman A, Liang SY. 2021. Coronavirus disease (COVID-19): a primer for emergency physicians. Am J Emerg Med 44:220–229. https://doi.org/10.1016/j.ajem.2020.03.036
2. WHO: World Health Organization. 2020. Coronavirus disease (COVID-19).
3. Saberi A, Gulyaeva AA, Brubacher JL, Newmark PA, Gorbalenya AE. 2018. A planarian nidovirus expands the limits of RNA genome size. PLoS Pathog 14:e1007314. https://doi.org/10.1371/journal.ppat.1007314
4. Hasöksüz M, Kiliç S, Saraç F. 2020. Coronaviruses and SARS-CoV-2. Turk J Med Sci 50:549–556. https://doi.org/10.3906/sag-2004-127
5. Boehm E, Kronig I, Neher RA, Eckerle I, Vetter P, Kaiser L, Geneva Centre for Emerging Viral Diseases. 2021. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin Microbiol Infect 27:1109–1117. https://doi.org/10.1016/j.cmi.2021.05.022
6. Naveca FG, Nascimento V, de Souza VC, Corado A de L, Nascimento F, Silva G, Costa Á, Duarte D, Pessoa K, Mejía M, et al. 2021. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. Nat Med 27:1230–1238. https://doi.org/10.1038/s41591-021-01378-7
7. Aguilar Rangel M, Dolan PT, Taguwa S, Xiao Y, Andino R, Frydman J. 2023. High-resolution mapping reveals the mechanism and contribution of genome insertions and deletions to RNA virus evolution. Proc Natl Acad Sci U S A 120:e2304667120. https://doi.org/10.1073/pnas.2304667120
8. Cheynier R, Kils-Hütten L, Meyerhans A, Wain-Hobson S. 2001. Insertion/deletion frequencies match those of point mutations in the hypervariable regions of the simian immunodeficiency virus surface envelope

gene. J Gen Virol 82:1613–1619. https://doi.org/10.1099/0022-1317-82-7-1613
9. Weng S, Zhou H, Ji C, Li L, Han N, Yang R, Shang J, Wu A. 2022. Conserved pattern and potential role of recurrent deletions in SARS-CoV-2 evolution. Microbiol Spectr 10:e0219121. https://doi.org/10.1128/spectrum.02191-21
10. Mahmoudabadi G, Milo R, Phillips R. 2017. Energetic cost of building a virus. Proc Natl Acad Sci U S A 114:E4324–E4333. https://doi.org/10.1073/pnas.1701670114
11. Wang Y, Chen XY, Yang L, Yao Q, Chen KP. 2022. Human SARS-CoV-2 has evolved to increase U content and reduce genome size. Int J Biol Macromol 204:356–363. https://doi.org/10.1016/j.ijbiomac.2022.02.034
12. Panzera Y, Calleros L, Goñi N, Marandino A, Techera C, Grecco S, Ramos N, Frabasile S, Tomás G, Condon E, Cortinas MN, Ramas V, Coppola L, Sorhouet C, Mogdasy C, Chiparelli H, Arbiza J, Delfraro A, Pérez R. 2022. Consecutive deletions in a unique Uruguayan SARS-CoV-2 lineage evidence the genetic variability potential of accessory genes. PLoS ONE 17:e0263563. https://doi.org/10.1371/journal.pone.0263563
13. Nagwa Ali S, Mohamed Ahmed R, Eslam Mansour S, Sara Ahmed R. 2022. Genetic variants of COVID-19 and vaccination. Is there a correlation? Open J Proteom Genom 7:001–005. https://doi.org/10.17352/ojpg.000011
14. Lauring AS, Frydman J, Andino R. 2013. The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol 11:327–336. https://doi.org/10.1038/nrmicro3003
15. Duffy S. 2018. Why are RNA virus mutation rates so damn high? PLoS Biol 16:e3000003. https://doi.org/10.1371/journal.pbio.3000003
16. Bull JJ, Sanjuán R, Wilke CO. 2007. Theory of lethal mutagenesis for viruses. J Virol 81:2930–2939. https://doi.org/10.1128/JVI.01624-06
17. Peck KM, Lauring AS, Sullivan CS. 2018. Complexities of viral mutation rates. J Virol 92:e01031-17. https://doi.org/10.1128/JVI.01031-17
18. Domingo E, Perales C. 2019. Viral quasispecies. PLoS Genet 15:e1008271. https://doi.org/10.1371/journal.pgen.1008271
19. Domingo E, Sheldon J, Perales C. 2012. Viral quasispecies evolution. Microbiol Mol Biol Rev 76:159–216. https://doi.org/10.1128/MMBR.05023-11
20. Lauring AS, Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. PLoS Pathog 6:e1001005. https://doi.org/10.1371/journal.ppat.1001005
21. Smith EC, Denison MR. 2013. Coronaviruses as DNA wannabes: a new model for the regulation of RNA virus replication fidelity. PLoS Pathog 9:e1003760. https://doi.org/10.1371/journal.ppat.1003760
22. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res 117:17–37. https://doi.org/10.1016/j.virusres.2006.01.017
23. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. 2021. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol 19:155–170. https://doi.org/10.1038/s41579-020-00468-6
24. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, Duprex WP. 2021. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science 371:1139–1142. https://doi.org/10.1126/science.abf6950
25. Ellis J. 2021. All in the family: a comparative look at coronaviruses. Can Vet J 62:825–833.
26. Redondo N, Zaldívar-López S, Garrido JJ, Montoya M. 2021. SARS-CoV-2 accessory proteins in viral pathogenesis: knowns and unknowns. Front Immunol 12:708264. https://doi.org/10.3389/fimmu.2021.708264
27. Nguyen TT, Pathirana PN, Nguyen T, Nguyen QVH, Bhatti A, Nguyen DC, Nguyen DT, Nguyen ND, Creighton D, Abdelrazek M. 2021. Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus). Sci Rep 11:3487. https://doi.org/10.1038/s41598-021-83105-3
28. Sanjuán R, Domingo-Calap P. 2016. Mechanisms of viral mutation. Cell Mol Life Sci 73:4433–4448. https://doi.org/10.1007/s00018-016-2299-6
29. Chrisman BS, Paskov K, Stockham N, Tabatabaei K, Jung J-Y, Washington P, Varma M, Sun MW, Maleki S, Wall DP. 2021. Indels in SARS-CoV-2 occur at template-switching hotspots. BioData Min 14:20. https://doi.org/10.1186/s13040-021-00251-0
30. Pereira F. 2020. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. Infect Genet Evol 85:104525. https://doi.org/10.1016/j.meegid.2020.104525
31. Pathak VK, Temin HM. 1992. 5-azacytidine and RNA secondary structure increase the retrovirus mutation rate. J Virol 66:3093–3100. https://doi.org/10.1128/JVI.66.5.3093-3100.1992
32. Pita JS, de Miranda JR, Schneider WL, Roossinck MJ. 2007. Environment determines fidelity for an RNA virus replicase. J Virol 81:9072–9077. https://doi.org/10.1128/JVI.00587-07
33. Lau S-Y, Wang P, Mok BW-Y, Zhang AJ, Chu H, Lee AC-Y, Deng S, Chen P, Chan K-H, Song W, Chen Z, To KK-W, Chan JF-W, Yuen K-Y, Chen H. 2020. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. Emerg Microbes Infect 9:837–842. https://doi.org/10.1080/22221751.2020.1756700
34. Simas MC da C, Costa SM, Gomes P da SFC, Cruz NVG da, Corrêa IA, de Souza MRM, Dornelas-Ribeiro M, Nogueira TLS, Santos CGMD, Hoffmann L, Tanuri A, Moura-Neto RS de, Damaso CR, Costa LJ da, Silva R. 2023. Evaluation of SARS-CoV-2 ORF7a deletions from COVID-19-positive individuals and its impact on virus spread in cell culture. Viruses 15:801. https://doi.org/10.3390/v15030801
35. Andreano E, Piccini G, Licastro D, Casalino L, Johnson NV, Paciello I, Dal Monego S, Pantano E, Manganaro N, Manenti A, Manna R, Casa E, Hyseni I, Benincasa L, Montomoli E, Amaro RE, McLellan JS, Rappuoli R. 2021. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. Proc Natl Acad Sci U S A 118:36. https://doi.org/10.1073/pnas.2103154118
36. Gribble J, Stevens LJ, Agostini ML, Anderson-Daniels J, Chappell JD, Lu X, Pruijssers AJ, Routh AL, Denison MR. 2021. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. PLoS Pathog 17:e1009226. https://doi.org/10.1371/journal.ppat.1009226
37. Licitra BN, Millet JK, Regan AD, Hamilton BS, Rinaldi VD, Duhamel GE, Whittaker GR. 2013. Mutation in spike protein cleavage site and pathogenesis of feline coronavirus. Emerg Infect Dis 19:1066–1073. https://doi.org/10.3201/eid1907.121094
38. Chen F, Knutson TP, Rossow S, Saif LJ, Marthaler DG. 2019. Decline of transmissible gastroenteritis virus and its complex evolutionary relationship with porcine respiratory coronavirus in the United States. Sci Rep 9:1–11. https://doi.org/10.1038/s41598-019-40564-z
39. Dolskiy AA, Gudymo AS, Taranov OS, Grishchenko IV, Shitik EM, Prokopov DY, Soldatov VO, Sobolevskaya EV, Bodnev SA, Danilchenko NV, Moiseeva AA, Torzhkova PY, Bulanovich YA, Onhonova GS, Ivleva EK, Kubekina MV, Belykh AE, Tregubchak TV, Ryzhikov AB, Gavrilova EV, Maksyutov RA, Deykin AV, Yudkin DV. 2022. The tissue distribution of SARS-CoV-2 in transgenic mice with inducible ubiquitous expression of hACE2. Front Mol Biosci 8:1339. https://doi.org/10.3389/fmolb.2021.821506
40. Papanikolaou V, Chrysovergis A, Ragos V, Tsiambas E, Katsinis S, Manoli A, Papouliakos S, Roukas D, Mastronikolis S, Peschos D, Batistatou A, Kyrodimos E, Mastronikolis N. 2022. From delta to Omicron: S1-RBD/S2 mutation/deletion equilibrium in SARS-CoV-2 defined variants. Gene 814:146134. https://doi.org/10.1016/j.gene.2021.146134
41. Meng B, Kemp SA, Papa G, Datir R, Ferreira I, Marelli S, Harvey WT, Lytras S, Mohamed A, Gallo G, et al. 2021. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. Cell Rep 35:109292. https://doi.org/10.1016/j.celrep.2021.109292
42. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, et al. 2021. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 596:276–280. https://doi.org/10.1038/s41586-021-03777-9
43. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, Beltramello M, Chen A, Liu Z, Zatta F, et al. 2021. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell 184:2332–2347. https://doi.org/10.1016/j.cell.2021.03.028
44. Rahman MS, Islam MR, Alam ASMRU, Islam I, Hoque MN, Akter S, Rahaman MM, Sultana M, Hossain MA. 2021. Evolutionary Dynamics of SARS-Cov-2 Nucleocapsid protein and its consequences. Journal of medical virology 93:2177–2195. https://doi.org/10.1002/jmv.26626
45. Sun Y-S, Sun H, Zhu H-P, Li G-L, Xu F, Lu H-J, Tang A, Wu B-B, Li Y-D, Yao P-P, Jiang J-M. 2022. Comparative transcriptomic analyzes of human lung epithelial cells infected with wild-type SARS-CoV-2 and its variant

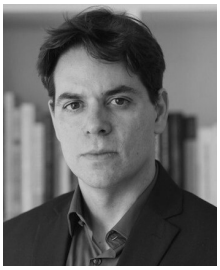
with a 12-bp missing in the E gene. Front Microbiol 13:1079764. https://doi.org/10.3389/fmicb.2022.1079764
46. Sun Y-S, Xu F, An Q, Chen C, Yang Z-N, Lu H-J, Chen J-C, Yao P-P, Jiang J-M, Zhu H-P. 2020. A SARS-CoV-2 variant with the 12-bp deletion at E gene. Emerg Microbes Infect 9:2361–2367. https://doi.org/10.1080/22221751.2020.1837017
47. Benedetti F, Snyder GA, Giovanetti M, Angeletti S, Gallo RC, Ciccozzi M, Zella D. 2020. Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. J Transl Med 18:329. https://doi.org/10.1186/s12967-020-02507-5
48. Lin J-W, Tang C, Wei H-C, Du B, Chen C, Wang M, Zhou Y, Yu M-X, Cheng L, Kuivanen S, et al. 2021. Genomic monitoring of SARS-CoV-2 uncovers an Nsp1 deletion variant that modulates type I interferon response. Cell Host Microbe 29:489–502. https://doi.org/10.1016/j.chom.2021.01.015
49. Gupta M, Azumaya CM, Moritz M, Pourmal S, Diallo A, Merz GE, Jang G, Bouhaddou M, Fossati A, Brilot AF, et al. 2021. CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. bioRxiv. https://doi.org/10.1101/2021.05.10.443524
50. Bal A, Destras G, Gaymard A, Bouscambert-Duchamp M, Valette M, Escuret V, Frobert E, Billaud G, Trouillet-Assant S, Cheynet V, Brengel-Pesce K, Morfin F, Lina B, Josset L. 2020. Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del). Clin Microbiol Infect 26:960–962. https://doi.org/10.1016/j.cmi.2020.03.020
51. Silvas JA, Vasquez DM, Park J-G, Chiem K, Allué-Guardia A, Garcia-Vilanova A, Platt RN, Miorin L, Kehrer T, Cupic A, Gonzalez-Reiche AS, Bakel H van, García-Sastre A, Anderson T, Torrelles JB, Ye C, Martinez-Sobrido L. 2021. Contribution of SARS-CoV-2 accessory proteins to viral pathogenicity in K18 human ACE2 transgenic mice. J Virol 95:e0040221. https://doi.org/10.1128/JVI.00402-21
52. Young BE, Fong S-W, Chan Y-H, Mak T-M, Ang LW, Anderson DE, Lee C-P, Amrun SN, Lee B, Goh YS, et al. 2020. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. Lancet 396:603–611. https://doi.org/10.1016/S0140-6736(20)31757-8
53. Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, Gloza-Rausch F, Balboni A, Battilani M, Rihtarič D, Toplak I, Ameneiros RS, Pfeifer A, Thiel V, Drexler JF, Müller MA, Drosten C. 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep 8:15177. https://doi.org/10.1038/s41598-018-33487-8
54. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, Zhuang Y, Kalimuddin S, Low JGH, Tan CW, et al. 2020. Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. mBio 11:e01610-20. https://doi.org/10.1128/mBio.01610-20
55. Gamage AM, Tan KS, Chan WOY, Liu J, Tan CW, Ong YK, Thong M, Andiappan AK, Anderson DE, Wang DY, Wang L-F. 2020. Infection of human nasal epithelial cells with SARS-CoV-2 and a 382-nt deletion isolate lacking OEF8 reveals similar viral kinetics and host transcriptional profiles. PLoS Pathog 16:e1009130. https://doi.org/10.1371/journal.ppat.1009130
56. Quéromès G, Destras G, Bal A, Regue H, Burfin G, Brun S, Fanget R, Morfin F, Valette M, Trouillet-Assant S, Lina B, Frobert E, Josset L. 2021. Characterization of SARS-CoV-2 ORF6 deletion variants detected in a nosocomial cluster during routine genomic surveillance, Lyon, France. Emerg Microbes Infect 10:167–177. https://doi.org/10.1080/22221751.2021.1872351
57. Addetia A, Xie H, Roychoudhury P, Shrestha L, Loprieno M, Huang M-L, Jerome KR, Greninger AL. 2020. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. J Clin Virol 129:104523. https://doi.org/10.1016/j.jcv.2020.104523
58. Mazur-Panasiuk N, Rabalski L, Gromowski T, Nowicki G, Kowalski M, Wydmanski W, Szulc P, Kosinski M, Gackowska K, Drweska-Matelska N, Grabowski J, Piotrowska-Mietelska A, Szewczyk B, Bienkowska-Szewczyk K, Swadzba J, Labaj P, Grzybek M, Pyrc K. 2021. Expansion of a SARS-CoV-2 Delta variant with an 872 nt deletion encompassing ORF7a, ORF7b, and ORF8. Euro Surveill 26:2100902. https://doi.org/10.2807/1560-7917.ES.2021.26.39.2100902
59. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, et al. 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science 368:779–782. https://doi.org/10.1126/science.abb7498
60. Kannan SR, Spratt AN, Quinn TP, Heng X, Lorson CL, Sönnerborg A, Byrareddy SN, Singh K. 2020. Infectivity of SARS-CoV-2: there is something more than D614G? J Neuroimmune Pharmacol 15:574–577. https://doi.org/10.1007/s11481-020-09954-3
61. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D, Ippodrino R. 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med 18:179. https://doi.org/10.1186/s12967-020-02344-6
62. Kim S-M, Kim E-H, Casel MAB, Kim Y-I, Sun R, Kwack M-J, Yoo J-S, Yu M-A, Yu K-M, Jang S-G, Rollon R, Choi JH, Gil J, Eun K, Kim H, Ensser A, Hwang J, Song M-S, Kim MH, Jung JU, Choi YK. 2022. SARS-CoV-2 variants show temperature-dependent enhanced polymerase activity in the upper respiratory tract and high transmissibility. bioRxiv. https://doi.org/10.1101/2022.09.27.509689
63. Wang C, Elghobashi-Meinhardt N, Balch WE. 2022. Covariant fitness clusters reveal structural evolution of SARS-CoV-2 polymerase across the human population. bioRxiv. https://doi.org/10.1101/2022.01.07.475295
64. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. 2020. Structure of replicating SARS-CoV-2 polymerase. Nature 584:154–156. https://doi.org/10.1038/s41586-020-2368-8

AUTHOR BIOS

Pedro Miguel Carneiro Jeronimo is a Bioinformatics analyst who completed his undergraduate studies at the Federal University of Ceará, Brazil. Since 2019, he has been working in the field of bioinformatics, focusing on molecular biology, data analysis, and machine learning. His interest in this field stems from the ability of programming in biology to analyze genomic data and drive innovation. Pedro is currently pursuing a Master's degree in Health Sciences at the Federal University of Ceará while simultaneously working as a Bioinformatics Analyst at both the Federal University of Ceará and Fundação Oswaldo Cruz (Fiocruz, Brazil). His work primarily revolves around bioinformatics and its application in understanding genomic data, which he has been actively exploring for over 2 years. His current pursuits reflect his ongoing commitment to leveraging bioinformatics for genomic analysis.

Cleber Furtado Aksenen is a biotechnologist who completed his undergraduate studies at the Federal University of Ceará, Brazil, from 2017 to 2022. Currently, he is pursuing a master's degree in pharmacology at the same institution, with an expected completion date in 2024. He works as a Bioinformatics analyst at the Genomic Surveillance Network of the Oswaldo Cruz Foundation (Fiocruz, Brazil), with genomic sequencing of new SARS-CoV-2 variants and other respiratory viruses, including influenza, RSV, and arboviruses. He has experience and interest in biotechnology, bioinformatics, molecular biology, next generation sequencing, data science, and statistics. He has been actively contributing to this field for the past 3 years.

Igor Oliveira Duarte has a bachelor's and a master's degree in biotechnology from the Federal University of Ceará, in Fortaleza, Brazil. He has worked as a Bioinformatics analyst and scientific contributor at the Genomic Surveillance Network of the Oswaldo Cruz Foundation (Fiocruz, Brazil), and there he worked with SARS-CoV-2 genomic monitoring for two years. In the meantime, he also contributed as an analyst to the Wastewater SARS-CoV-2 Surveillance Network, which published weekly reports on the pathogen levels found in the urban wastewaters of Fortaleza, Brazil. Currently, he is a Ph.D. student at Kiel University, in Kiel, Germany. His interests include evolution, genomics, symbiosis, bioinformatics, and data science.



Roberto D. Lins obtained his B.Sc. degree in biological sciences in 1994 and his Ph.D. in chemistry in 1999 from UFPE (Brazil). During his Ph.D., he worked under the supervision of Professors Ricardo Ferreira (UFPE, Brazil) and J. Andrew McCammon (UCSD, EUA). After a short stay at the University of Houston (EUA) as a Visiting Assistant Professor (2000), he pursued postdoctoral research in Switzerland at the ETH-Zurich (2000–2003) and at EPFL (2003–2005). In 2005, he became a Senior Research Scientist at PNNL (EUA), and in 2009 Professor of Chemistry at UFPE (Brazil). In 2014, he became a Research Scientist in Public Health at the Oswaldo Cruz Foundation (Fiocruz, Brazil), where he coordinates a theoretical-experimental laboratory with a focus on engineering novel protein-based vaccines, biopharmaceuticals, and diagnostic markers for viral infectious diseases. He is also a member of the Fiocruz Genomics Surveillance Network and acts as the coordinator of a proteomics facility at Fiocruz.

Fabio Miyajima is a graduate in dentistry from the University of São Paulo and holds a Master's degree from Unicamp with an emphasis on Forensic Genetics. He obtained a doctorate from the University of Manchester in Epidemiology and Health Sciences. He has a postdoctoral degree in clinical research of nosocomial diseases from the Royal Liverpool and Broadgreen University Hospitals and in Translational Medicine from the University of Liverpool. He was a Special Visiting Professor of the Postgraduate Program in Pharmacology at the Federal University of Ceará. Since 2018, he has been a Specialist in Science, Technology, Production, and Innovation in Public Health at Fiocruz. His research interests include emerging diseases, dysbiosis, intestinal microbiome, mental health, diagnostic/prognostic tools, drug repositioning, and population genetics/immunogenetics.