Check for updates

Molecular transition of SARS-CoV-2 from critical patients during the first year of the COVID-19 pandemic in Mexico City

Aldo Hugo De La Cruz-Montoya[1,2],
Clara Estela Díaz Velásquez[1,2], Héctor Martínez-Gregorio[1,2],
Miguel Ruiz-De La Cruz[1,2,3], José Bustos-Arriaga[2],
Tannya Karen Castro-Jiménez[2],
Jonadab Efraín Olguín-Hernández[1], Miriam Rodríguez-Sosa[2],
Luis Ignacio Terrazas-Valdes[1,2],
Luis Armando Jiménez-Alvarez[4],
Nora Elemi Regino-Zamarripa[4,5], Gustavo Ramírez-Martínez[4],
Alfredo Cruz-Lagunas[4], Irlanda Peralta-Arrieta[6],
Leonel Armas-López[2], Belinda Maricela Contreras-Garza[6],
Gabriel Palma-Cortés[7], Carlos Cabello-Gutierrez[7],
Renata Báez-Saldaña[6], Joaquín Zúñiga[4,5,6]*,
Federico Ávila-Moreno[2,6,8]* and Felipe Vaca-Paniagua[1,2,9]*

[1]Laboratorio Nacional en Salud, Diagnóstico Molecular y Efecto Ambiental en Enfermedades Crónico-Degenerativas, Facultad de Estudios Superiores Iztacala, Tlalnepantla, Mexico, [2]Unidad de Investigación en Biomedicina, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla, Mexico, [3]Departamento de Infectómica y Patogénesis Molecular, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN), Avenida Instituto Politécnico Nacional, Colonia San Pedro Zacatenco, Delegación Gustavo A. Madero, Ciudad de México, Mexico, [4]Laboratorio de Inmunobiología y Genética y Departamento de Virología, Instituto Nacional de Enfermedades Respiratorias (INER) Ismael Cosio Villegas, Ciudad de México, Mexico, [5]Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Ciudad de México, Mexico, [6]Instituto Nacional de Enfermedades Respiratorias (INER) Ismael Cosio Villegas, Ciudad de México, Mexico, [7]Department of Research in Virology and Mycology, Instituto Nacional de Enfermedades Respiratorias (INER) Ismael Cosio Villegas, Ciudad de México, Mexico, [8]Laboratorio 12 de Enfermedades Pulmonares y Epigenómica del Cáncer, Unidad de Investigación en Biomedicina (UBIMED), Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Tlalnepantla, Mexico, [9]Subdirección de Investigación Básica, Instituto Nacional de Cancerología, Ciudad de México, Mexico

Background: The SARS-CoV-2 virus has caused unprecedented mortality since its emergence in late 2019. The continuous evolution of the viral genome through the concerted action of mutational forces has produced distinct variants that became dominant, challenging human immunity and vaccine development.

Aim and methods: In this work, through an integrative genomic approach, we describe the molecular transition of SARS-CoV-2 by analyzing the viral whole

genome sequences from 50 critical COVID-19 patients recruited during the first year of the pandemic in Mexico City.

Results: Our results revealed differential levels of the evolutionary forces across the genome and specific mutational processes that have shaped the first two epidemiological waves of the pandemic in Mexico. Through phylogenetic analyses, we observed a genomic transition in the circulating SARS-CoV-2 genomes from several lineages prevalent in the first wave to a dominance of the B.1.1.519 variant (defined by T478K, P681H, and T732A mutations in the spike protein) in the second wave.

Conclusion: This work contributes to a better understanding of the evolutionary dynamics and selective pressures that act at the genomic level, the prediction of more accurate variants of clinical significance, and a better comprehension of the molecular mechanisms driving the evolution of SARS-CoV-2 to improve vaccine and drug development.

KEYWORDS

SARS-CoV-2, genome, COVID-19, mutational signatures, evolutionary forces

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified as the causal agent of pneumonia in a cluster of patients in Wuhan, China in late December 2019 and led off the pandemic. SARS-CoV-2 is an enveloped virus with a linear positive-sense single-stranded RNA genome that belongs to the *Coronaviridae* family and causes the infectious coronavirus disease 2019 (COVID-19) (Zhu et al., 2020). Its genome is composed of 29,903 bases distributed across several open reading frames (ORF), starting with a 5'-untranslated region (5'-UTR) followed by two-overlapping ORFs that comprise almost three-quarters of the genome, composed by ORF1a and ORF1b which code for 16 nonstructural proteins (Nsp) that take part in viral replication and transcription. Then, various sequences that are expressed as subgenomic RNAs that code for accessory proteins such as ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10, four coding for structural proteins S, E, M and N, all flanked by a 3'-UTR (Michel et al., 2020; Naqvi et al., 2020).

Mexico has experienced five COVID-19 waves, with a total of 7,541,591 confirmed cases and 333,531 deaths as to April 3, 2023 (Worldometers.info, 2023). During the first year, in low and middle-income countries, factors such as poverty, access to healthcare, and increased population risk of several metabolic and cardiovascular diseases made the pandemic particularly difficult to handle. This was reflected in a higher mortality rate due to COVID-19 in Latin America, specifically in countries like Mexico where 1,992,794 cases and 196,525 deaths were reported only in the first year (Worldometers.info, 2023), compared with high-income countries like those from western Europe (Ortiz-Hernández and Pérez-Sastré, 2020; Palacio-Mejía et al., 2022).

SARS-CoV-2 has a high replication fidelity mainly dependent on the RNA-dependent on the 3' to 5' exoribonuclease activity of the nonstructural protein 14 (Nsp14) and therefore a low mutation rate compared with other viruses (Denison et al., 2011). Nsp14 has been associated with proofreading activity, reducing the mutation rate in the viral genome during its replication (Ma et al., 2015). Even so, many mutations have been identified throughout the SARS-CoV-2 genomes sequenced so far, which means that there are other mechanisms such as oxidative stress and RNA editing which could be shaping the SARS-CoV-2 mutational profile (Di Giorgio et al., 2020; Simmonds, 2020; Graudenzi et al., 2021; Graudenzi et al., 2021; Mourier et al., 2021; Ratcliff and Simmonds, 2021).

The first SARS-CoV-2 genome reported from a Mexican patient was published in May 2020 and belonged to the subclade A2a, lineage G (Garcés-Ayala et al., 2020). An initial genomic analysis from 17 whole genome SARS-CoV-2 sequences isolated from Mexican patients reported the presence of the H49Y aa substitution in the Spike protein and the phylogenetic analysis showed that the viruses sequenced belonged to the A2/G and B/S lineages, two of the three lineages circulating at that time (Taboada et al., 2020). The rapid availability of vast SARS-CoV-2 genomic data has been a key factor in the response to the pandemic; from tracking, diagnosing, and following transmission across populations to the design and development of vaccines and antivirals (Amanat and Krammer, 2020; Zhang and Holmes, 2020; Dellicour et al., 2021). The study of these genomic data is crucial to understand the selective pressures and the evolutionary dynamics acting on the SARS-CoV-2 genome, predicting more accurate variants of clinical significance, better understanding the molecular mechanisms driving SARS-CoV-2 evolution, and improving vaccine and drug development.

In this study, we extensively describe the molecular transition of SARS-CoV-2 from critical COVID-19 patients from the first year of the pandemic in Mexico City through an integrative genomic approach.

Materials and methods

Population of study

Nasopharyngeal swab samples were obtained from 200 patients from the National Institute of Respiratory Diseases (INER) in Mexico City from June 2020 to February 2021 (Figure 1, panel 1), within the approved protocols B28-16 and B09-20. Patients were grouped into mild, severe, and critical according to the COVID-19 Treatment Guidelines. The National Institutes of Health guidelines for treatment categorize mild COVID as patients with symptoms such as fever, cough, sore throat, malaise, headache, muscle pain, nausea, vomiting, diarrhea, and loss of taste and smell. According to the Infectious Diseases Society of America, severe COVID is defined by SpO2 ≤94% on room air, including patients on supplemental oxygen and symptoms like shortness of breath, dyspnea, and abnormal chest imaging. Critical COVID is characterized by the presence of symptoms previously mentioned and often the need for mechanical ventilation and extracorporeal membrane oxygenation (ECMO), and the existence of some of the following: respiratory

failure, septic shock, and/or multiple organ dysfunction (Chan, 2022).

RNA extraction and sample selection

Total RNA was extracted using the QIAmp Viral RNA Mini kit (QIAGEN, Hilden, Germany) in the presence of 5 µg/mL of carrier RNA and quantified by fluorometry in a Quantus Fluorometer (Promega Corporation, Madison, Wisconsin, USA). Fifty nanograms from each sample were analyzed by qPCR using the GeneFinder COVID-19 Plus RealAmp Kit (OSANG Healthcare Co., Ltd, South Korea) and all samples with a Ct <35 were selected for library preparation (Figure 1, panel 2).

Library preparation and SARS-CoV-2 whole genome sequencing

Library preparation was performed employing the CleanPlex SARS-CoV-2 Research and Surveillance Panel (Paragon Genomics, Inc., Hayward, California, USA) according to the manufacturer's instructions. Briefly, total RNA was used for cDNA synthesis and purification. Then, a multiplex PCR with target-specific primers to amplify targets of interest was carried out, covering the entire SARS-CoV-2 genome with a 2-pool design, followed by a
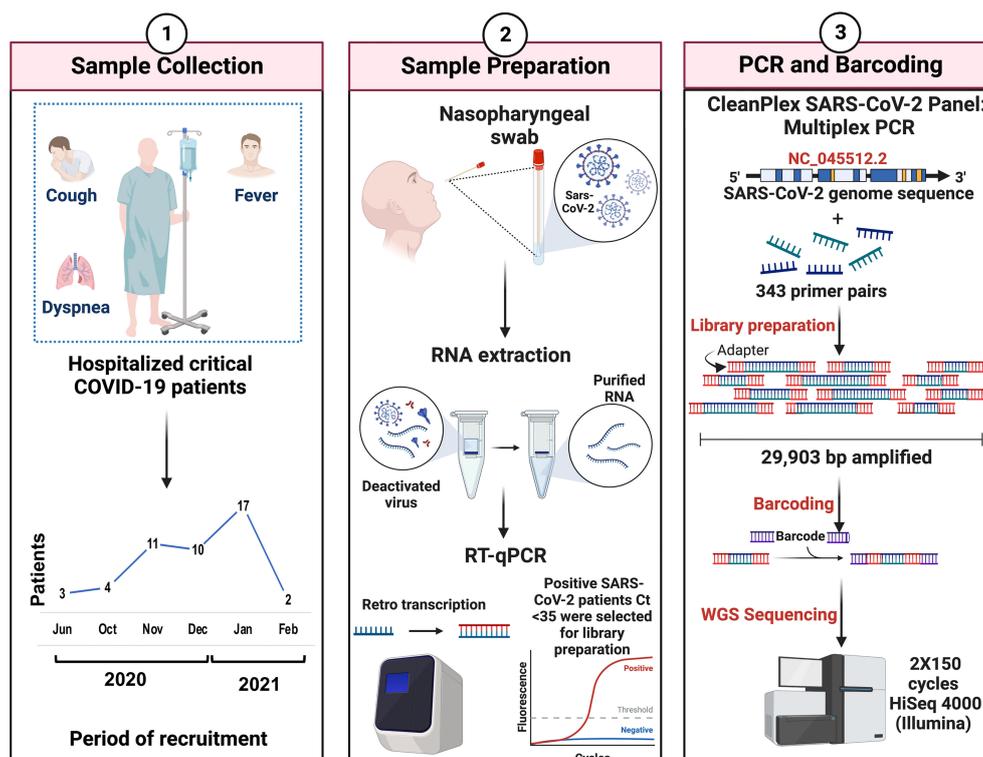
FIGURE 1
Overview of the study workflow. Critical COVID-19 patients were selected according to the COVID-19 Treatment Guidelines (Panel 1). Nasopharyngeal samples were collected, and the viral RNA extraction of SARS-CoV-2-positive patients was carried out. Criteria of <35 ct in RT-qPCR were required for the preparation of genomic libraries (Panel 2). Library preparation was performed using multiplex PCR and WGS was performed in a 2X150 cycles format on an Illumina HiSeq (Panel 3).

digestion reaction that achieves background cleaning by removing nonspecific PCR products. Finally, a PCR reaction with CleanPlex Indexed PCR Primers to amplify and add indexes to the generated libraries was conducted. After library preparation, the amplification of viral RNA was verified by automated electrophoresis using a 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, California, USA). The presence of a peak at ~275 bp indicated the successful amplification of the targeted regions. The samples were normalized by molar concentration, pooled, and sequenced for 2X150 cycles in a HiSeq 4000 (Illumina, San Diego, California, USA) at Novogene (Sacramento, California, USA) (Figure 1, panel 3).

Bioinformatic analysis

SOPHiA DDM software (SOPHiA Genetics Inc., Boston, Massachusetts, USA) was used to align reads to the reference genome (NC_045512.2), read filtering, and variant calling. First, read mapping adaptors were trimmed, mispriming events were eliminated, read softclipped regions were realigned and primer sequences trimmed. Reads shorter than 21 bp were excluded, and the resulting alignment was used for variant calling using SOPHiA DDM pipeline. FASTA and annotated files were downloaded for downstream analysis. Mutations were considered from samples with a coverage of at least 98% and a mutation fraction of at least 70% (Kubik et al., 2021). To visualize the molecular transition of SARS-CoV-2 mutations were plotted and segmented by year using the ComplexHeatmap R package (Gu et al., 2016).

Molecular characterization and evolutionary analysis

For the identification of molecular characteristics of SARS-CoV-2, mutations from 50 critical patients were classified as single nucleotide variants (SNV), synonymous, and nonsynonymous. Mutations were analyzed by genomic region, normalized to each region length and plotted in the whole genome using R. The overall transition/transversion ratio was calculated with MEGA software V.11.0 (Tamura et al., 2021).

579 complete SARS-CoV-2 genome sequences (>29,000 nt) with high coverage uploaded from Mexico City to the Global Initiative on Sharing Avian Influenza Data (GISAID) database within 2020-06-01 to 2021-02-28 were selected. FASTA files were downloaded to study the evolutionary forces acting on SARS-CoV-2 genomes along with the 50 whole genome sequences from this study. Using these 629 sequences, we evaluated the transition/transversion ration for each genomic region and assessed neutrality with Tajima's D test using DNASP6 software (Rozas et al., 2017) and positive selection with Datamonkey tool (Weaver et al., 2018).

Mutational signatures

For the mutation signature analysis, we used Alexandrov´s strategy (Alexandrov et al., 2013). We evaluated all changes in a trinucleotide context where the mutated nucleotide is in the center. In this approach, there are 192 (12*16) possible changes. For the ZAP mutational signature, we used a dinucleotide context because this enzyme acts in CpG nucleotide pairs (16 different dinucleotide combinations). We used a 12- and 6-substitution pattern to present these data (Graudenzi et al., 2021). Then, we analyzed single nucleotide changes in every SARS-CoV-2 region and normalized them to region length to identify specific mutational signatures in each region. All mutational signatures were plotted in R using ggplot2 package (Villanueva and Chen, 2019).

Protein modeling

PyMOL (DeLano, 2002) was used to model the SARS-CoV-2 spike glycoprotein and identify the mutated sites in our samples. The structure of the spike protein (id:7DZW) was downloaded from the Protein Data Bank (PDB) (Kouranov et al., 2006) and modeled in two ways: i) using all the mutations identified in our cohort and ii) using only the mutations identified in the January-February 2021 period.

Phylogenetic analysis

Phylogenetic trees were constructed using Nextstrain and visualized with Auspice v2.37.3. The tool Nexctlade v.2.5.0 was used to compare the 50 whole-genome sequences to SARS-CoV-2 reference sequences and determine their location in the phylogenetic tree (Hadfield et al., 2018).

Results

Patient characteristics

For this study, we collected nasopharyngeal swab samples over eight months (June 2020 to February 2021) from a total of 200 COVID-19-positive patients. After qPCR analysis only samples with a Ct <35 were selected for library preparation and whole genome sequencing (Figure 1). After this filter, 50 patients were selected from which 35 (70%) are men and 13 (26%) are women. Most of them were older than 30 years with mean ages of 59.1 and 57.2, respectively. Only 4 patients were diagnosed with mild disease (8%) and the other 46 patients were diagnosed with critical disease (92%) (Figure 2).
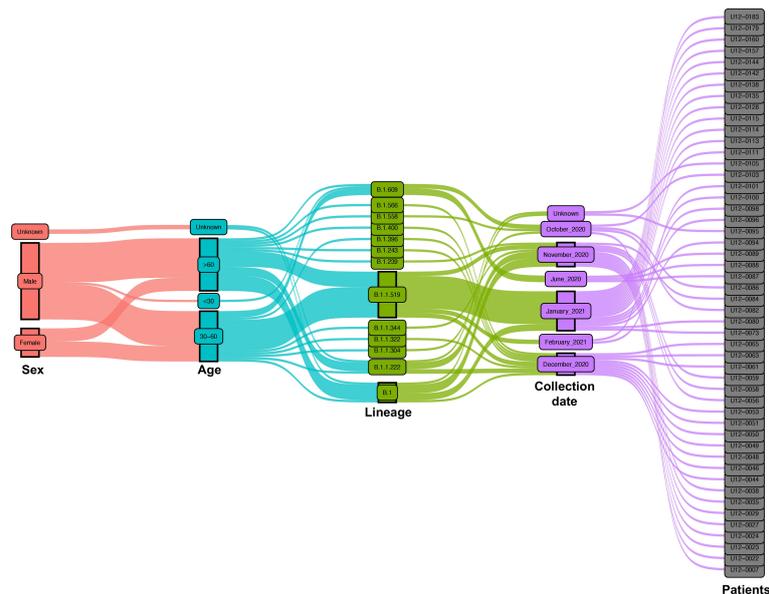
FIGURE 2
Epidemiological and virological characteristics of critical COVID-19 Mexican patients. Demographic parameters, nasopharyngeal swab collection dates, and virologic parameters of the patients are represented in an alluvial diagram, each one depicted in a different color. The distribution of the main lineages of SARS-CoV-2 is sketched during the 8-month study period from June 2020 to February 2021.

Global mutational analysis of SARS-CoV-2

From the 50 SARS-CoV-2 whole genome sequences, we identified a total of 923 mutations (mean = 18.46, range 5 - 28) across 12 different ORFs, 5', and 3'-UTRs. Of these mutations, 48.2% were synonymous and 51.8% were nonsynonymous (Figure 3B). The highest percentage of synonymous mutations are present in the 5' untranslated region (UTR) and sequences that code for Nsp3 and Nsp15, ORF1a and ORF1b, respectively. While the S, ORF1b, and N regions had the largest number of nonsynonymous mutations distributed across them (Figure 3A). Among the sequences that code for structural proteins, the two with more mutations were S and N. In the S region we identified 159 nonsynonymous (17.22%) and 21 synonymous mutations (2.27%), while in N we identified 79 nonsynonymous (8.55%) and 56 synonymous mutations (6.06%) (Figure 3A). We performed a mutational analysis and found diverse mutational transition and transversion patterns dominated by a majority of C>T (35.09%), followed by T>C (20.05%), G>A (11.51%), and A>G (7.57%) transitions (Figure 3C).

Evolutionary analysis of SARS-CoV-2

To evaluate positive selection across the whole viral genome, we calculated the dN/dS (nonsynonymous/synonymous ratio) from 629 complete SARS-CoV-2 genomic sequences. Four amino acid changes showed significant positive selection: P141S (Nsp3), T492I (Nsp4), P681H (S), and T732A (S) (Table 1). In contrast, when we evaluated Tajima's D test, there was a tendency for purifying selection in all genomic regions. Ti/Tv showed higher values in Nsp4, Nsp5, Nsp9, Nsp10, Nsp15, and ORF6 (Table 2).

Mutational signature analysis

We performed a mutational signature analysis in the context of one, two, and three nucleotides. In a one-nucleotide context, we identified that most changes were C>T/G>A (59.6%), followed by T>C/A>G (23.2%), and C>A/G>T (12.4%) (Figure 4C). Then, we normalized the number of mutations to sequence length to determine the mutational signature distribution along the SARS-CoV-2 genome. We found that most C>T/G>A changes were overrepresented mainly in the 5'UTR, followed by the N region. While the T>C/A>G changes were mainly found in the Nsp15 followed by Nsp6, and S. Although C>G/G>C changes were not the most prevalent across the genome, these changes were found distributed only between N and the 3'UTR (Figure 4B). We also identified that the most common single nucleotide changes in the context of three nucleotides were those of C>T followed by A>G (Figure 4A). Finally, in the dinucleotide context, where the first nucleotide is the mutation and the second is the next nucleotide in the 5'>3' direction, the most prevalent dinucleotide changes were CpT, CpA, CpG, and GpG (Figure 4D).

Genomic landscape of SARS-CoV-2

The S:D614G and ORF1b:P323L amino acid (aa) substitutions in the spike protein and the RNA-dependent RNA polymerase (RdRp), respectively, were present in all 50 SARS-CoV-2 genomes,

FIGURE 3
Mutational profile of SARS-CoV-2 in critical COVID-19 patients. **(A)** Distribution of nonsynonymous, synonymous mutations, and single nucleotide variants across the SARS-CoV-2 genome. **(B)** Percentage of nonsynonymous, synonymous mutations, and single nucleotide variants. **(C)** Percentage of transitions and transversions.
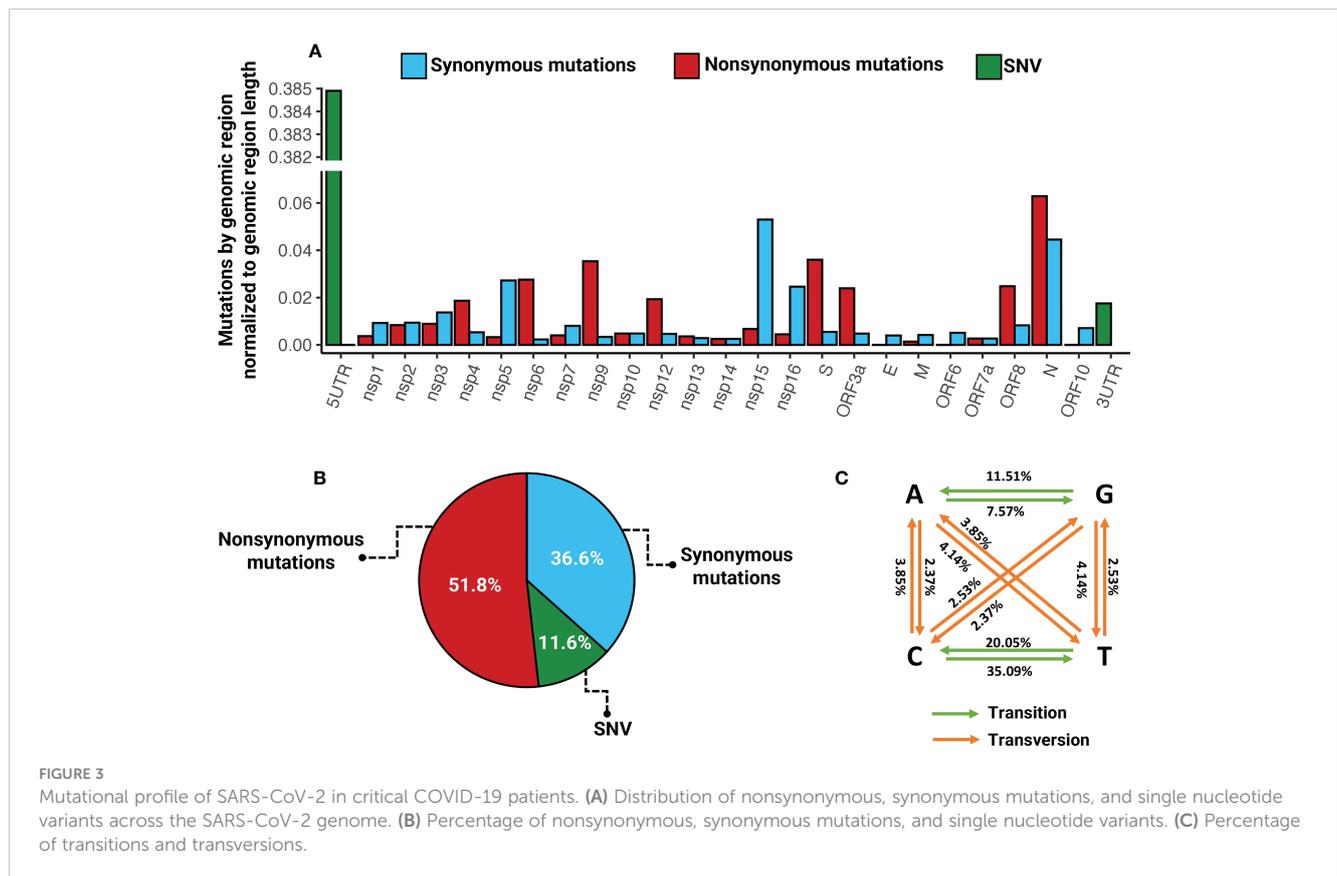
co-occurring with two synonymous mutations; C241T in the 5'-UTR and C3037T in the Nsp3. Other frequent aa changes in S across samples were T732A (28%), P681H (22%), and T478K (20%). These three aa substitutions in the S sequence define the B.1.1.519 Phylogenetic Assignment of Named Global Outbreak (PANGO) lineage, which we found in 21 (42%) of our samples, followed by B.1 (9/50, 18%), B.1.1.222 (6/50, 12%), and B.1.609 (5/50,10%). More aa substitutions that characterized the B.1.1.519 lineage that were present in our samples are P959S, T3255I, I3618V, and T4175I in ORF1a and R203K, and G204R in N (Figure 5A, B).

Protein alterations by modeling analysis

We modeled the SARS-CoV-2 spike glycoprotein structure and identified the mutated sites represented in our samples. In the left panel of Figure 6, we can observe the location of the different

structural domains along the spike protein. In the middle panel, all the mutations identified in our cohort in the whole study are shown and the D614G amino acid substitution which is present in all samples is indicated. In the right panel, only the mutations identified in the January-February 2021 period are represented. The distribution of amino acid changes represented in the primary structure of the SARS-CoV-2 spike protein is depicted in Figure 6C.

Phylogenetic analysis

The phylogenetic tree depicted in Figure 7 shows the phylogenetic analysis of 50 whole-genome SARS-CoV-2 sequences from critical COVID-19 patients circulating during the first two waves in Mexico. The samples are grouped through clades 20A, 20C, and 20B, in ascending order regarding the number of mutations compared with the reference genome.

TABLE 1 Nonsynonymous/Synonymous mutation ratio from 629 complete SARS-CoV-2 genome sequences from Mexico in the period from June 2020 to February 2021.

Nucleotide change	Amino Acid change	Genomic location	dN/dS (p-value)
C421T	P141S	ORF1a (Nsp3)	7.00 (p=0.003)
C1475T	T492I	ORF1a (Nsp4)	3.00 (p=0.088)
C2042A	P681H	S	3.34 (p=0.082)
A2194G	T732A	S	6.00 (p=0.039)

TABLE 2 Transition/transversion ratio and Tajima's D neutrality test per genomic region from 629 complete SARS-CoV-2 genome sequences from Mexico in the period from June 2020 to February 2021.

Genomic location	Tajima's D	Ti/Tv
Nsp1	-2.6	6.3
Nsp2	-2.8	2.4
Nsp3	-2.8	3.5
Nsp4	-2.7	13.9
Nsp5	-2.5	19.4
Nsp6	-2.5	5.0
Nsp7	-2.3	7.2
Nsp8	-2.5	7.2
Nsp9	-2.0	44.0
Nsp10	-2.3	13.8
Nsp12	-2.8	2.3
Nsp13	-2.7	3.7
Nsp14	-2.8	3.1
Nsp15	-2.5	12.0
Nsp16	-2.5	9.1
S	-2.8	1.1
ORF3a	-2.7	1.3
E	-2.2	1.0
M	-2.6	1.5
ORF6	-2.2	11.7
ORF7a	-2.6	4.1
ORF8	-2.6	2.3
N	-2.5	2.4
ORF10	-2.1	4.9

Discussion

In this work, we reported the genomic landscape of whole-genome sequences from critical COVID-19 patients within the first and second COVID-19 waves during the first year of the pandemic in Mexico City (Figure 8).

We recruited 200 COVID-19 patients from January 2020 to February 2021 and collected nasopharyngeal swab samples. All samples were evaluated by qPCR, and 50 samples that passed a Ct <35 quality filter were selected for library preparation and whole genome sequencing. Most patients in this study were men (70%), and older than 30 years (mean age=58.1), comparable to previous works performed in Mexican (Taboada et al., 2021) and other populations (Jin et al., 2020). Most critical patients (40%) harbored the B.1.1.519 variant. Unfortunately, and due to privacy matters, in this study we were not able to collect clinical characteristics. However, it has been reported that patients harboring the

B.1.1.519 variant showed an increase in respiratory symptoms unlike those harboring other variants. Additionally, in previous studies, logistic regression showed an association with increased dyspnea, chest pain, and cyanosis (Cedro-Tanda et al., 2021), which could explain the critical condition of most patients in this work.

Mutations in the SARS-CoV-2 genome can be neutral or deleterious, and some mutations provide evolutionary advantages at the molecular level, and may be associated with enhanced fitness, increased infectivity, and immune evasion (Korber et al., 2020; Li et al., 2021; Plante et al., 2021). To understand the evolution of the SARS-CoV-2 genome, we assessed the evolutionary forces that act on its loci. Overall, we found a higher proportion of nonsynonymous than synonymous mutations, 51.8% versus 48.2%, respectively. Some regions, such as Nsp6, Nsp9, Nsp12, S, ORF3a, ORF8, and N exhibited more nonsynonymous than synonymous mutations. On the other hand, we detected a higher rate of transversion than transitions in all the genome, however, transversion was not detected in some regions such as Nsp9, Nsp10, E, ORF6, and ORF7a.

To evaluate the neutrality and a positive selection in SARS-CoV-2 we performed Tajima's D test and calculated the dN/dS ratio. Neutrality analysis of Tajima's D showed that all genomic regions are under purifying selection. On the other hand, P141S (Nsp3), T492I (Nsp4), P681H (S), and T732A (S) showed significance in the dN/dS ratio, indicating that they are under positive selection. In addition, our results show that some regions with a high rate of nonsynonymous mutations are not positively selected. These data suggest that different selection pressures may act together on the evolution of SARS-CoV-2 improving fitness or establishing purifying selection.

Notably, the higher rate of nonsynonymous mutations in Nsp12, Nsp4, Nsp6, Nsp3, and Nsp9, where all key components of the replication machinery of the virus relay, may impact the replication efficiency, the acquisition of novel mutation, and fitness. Nsp12 is the N-terminal region of the RNA-dependent RNA polymerase; missense mutations in this region of the genome could influence the mutation rate of the polymerase (Hillen et al., 2020). Nsp4 and Nsp6 are predicted to participate in the anchoring of the replication complex to double membrane vesicles in the cytoplasm (Prentice et al., 2004). Nsp3 interacts with other Nsp proteins and with the viral RNA to form the replication complex (Ziebuhr et al., 2000). Nsp9 is an RNA-binding protein that has been proposed to participate in the replication complex (Egloff et al., 2004). Therefore, the aggregation of mutations in the essential regions of viral replication may facilitate adaptability and transmission by accelerating mutation acquisition (Yao et al., 2020). We also identified several mutations in the 5' and 3' UTRs. It is known that there are implications of the genetic variability in these regions that could impact the stability of predicted secondary structures that have been proposed to regulate the translation of viral proteins and compactness of the viral genome of other viruses and are molecular targets of antiviral candidates (Tubiana et al., 2015; Zhao et al., 2020; Miao et al., 2021).

Molecular mimicry is an effective evolutionary adaptation for viral infections to evade immune response mechanisms (Lucas et al., 2008; Strumillo et al., 2021). Most genomic SARS-CoV-2

FIGURE 4
Mutational patterns identified in the SARS-CoV-2 genome. **(A)** Mutational signature analysis in the context of three nucleotides. **(B)** Mutational signature analysis in a one-nucleotide context normalized to region length. **(C)** Percentage of one-nucleotide changes. **(D)** Mutational signature analysis in a dinucleotide context.
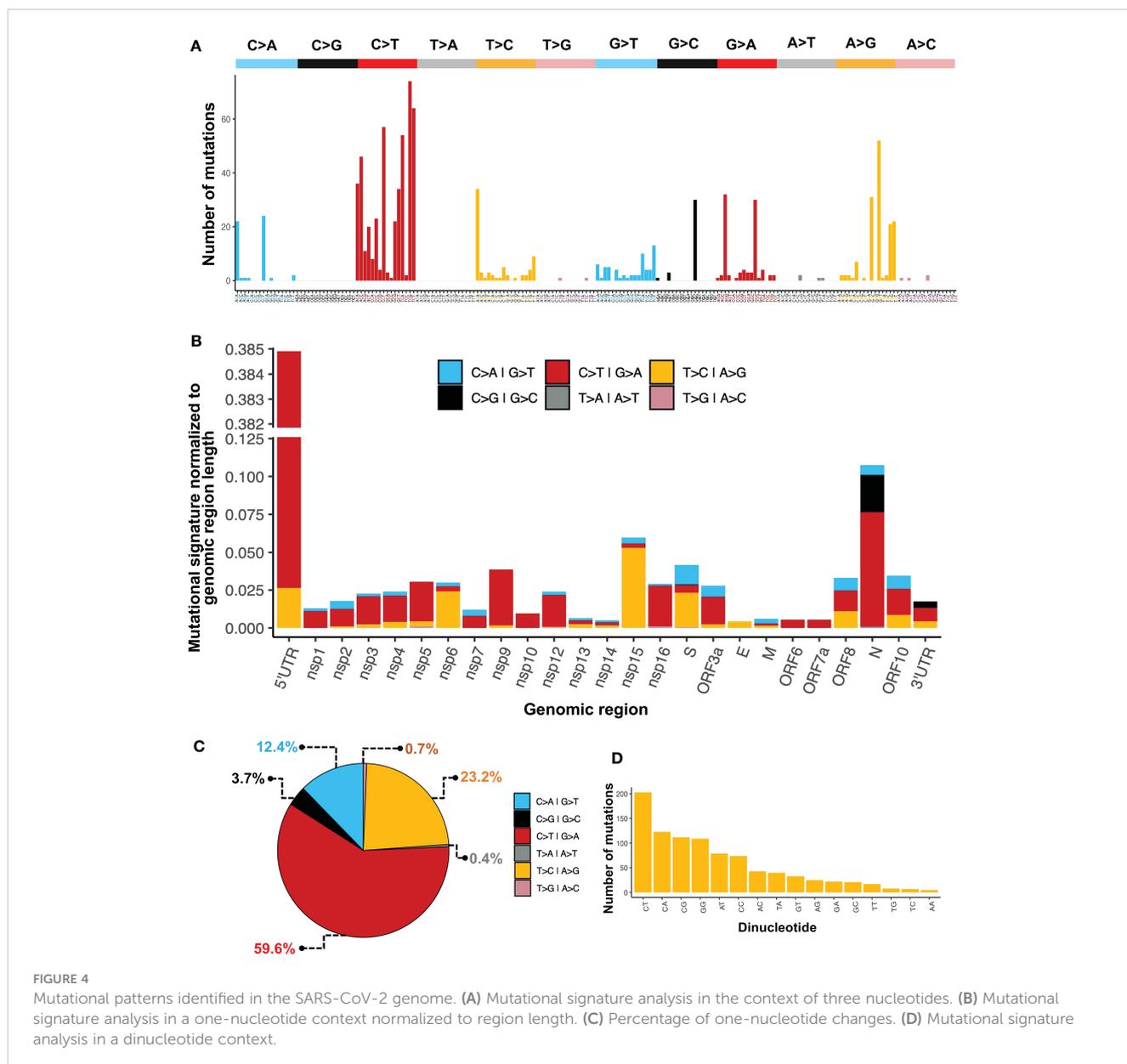
studies have focused on the analysis of nonsynonymous mutations. However, it has been reported that synonymous changes account for approximately 30% of all SARS-CoV-2 mutations. As a result of genomic large-scale analysis, it has been shown that synonymous mutations that increase the similarity of viral codons to human codons are more likely to become fixed in the SARS-CoV-2 genome throughout time, thereby adapting to the human host and improving the usage of the human aminoacyl-tRNA set (Ramazzotti et al., 2022).

The SARS-CoV-2 mutational signatures of our study revealed that C>T, T>C, and C>A are the predominant single-base substitutions. This mutational asymmetry phenomenon has been observed in SARS-CoV and SARS-CoV-2 (Simmonds, 2020). In addition, the RNA viruses HCV, FMDV, MNV, HPgV-1, and rubella virus have shown an excess of C>T changes over the

reverse (T>C) and complementary (G>A) transitions (Simmonds and Ansari, 2021).

The mutational processes and the subjacent mechanisms that seem to be shaping the SARS-CoV-2 genome through an increment in the C>T change frequency in RNA viruses remain to be elucidated, although there is previous evidence that supports three main mutational mechanisms that may cause mutational asymmetry: i) RNA transcription, ii) reactive oxidative species (ROS), and iii) RNA editing (Simmonds, 2020) (Figure 8).

RNA transcription is an important source of mutations due to transcription errors introduced by polymerases during the replication of RNA viruses. Nevertheless, biochemical characterizations of poliovirus RNA-dependent RNA polymerase (RdRp) and HIV reverse transcriptase did not identify an increase in C>T mutations over other changes (Ward et al., 1988; Kati et al., 1992; Freistadt et al.,
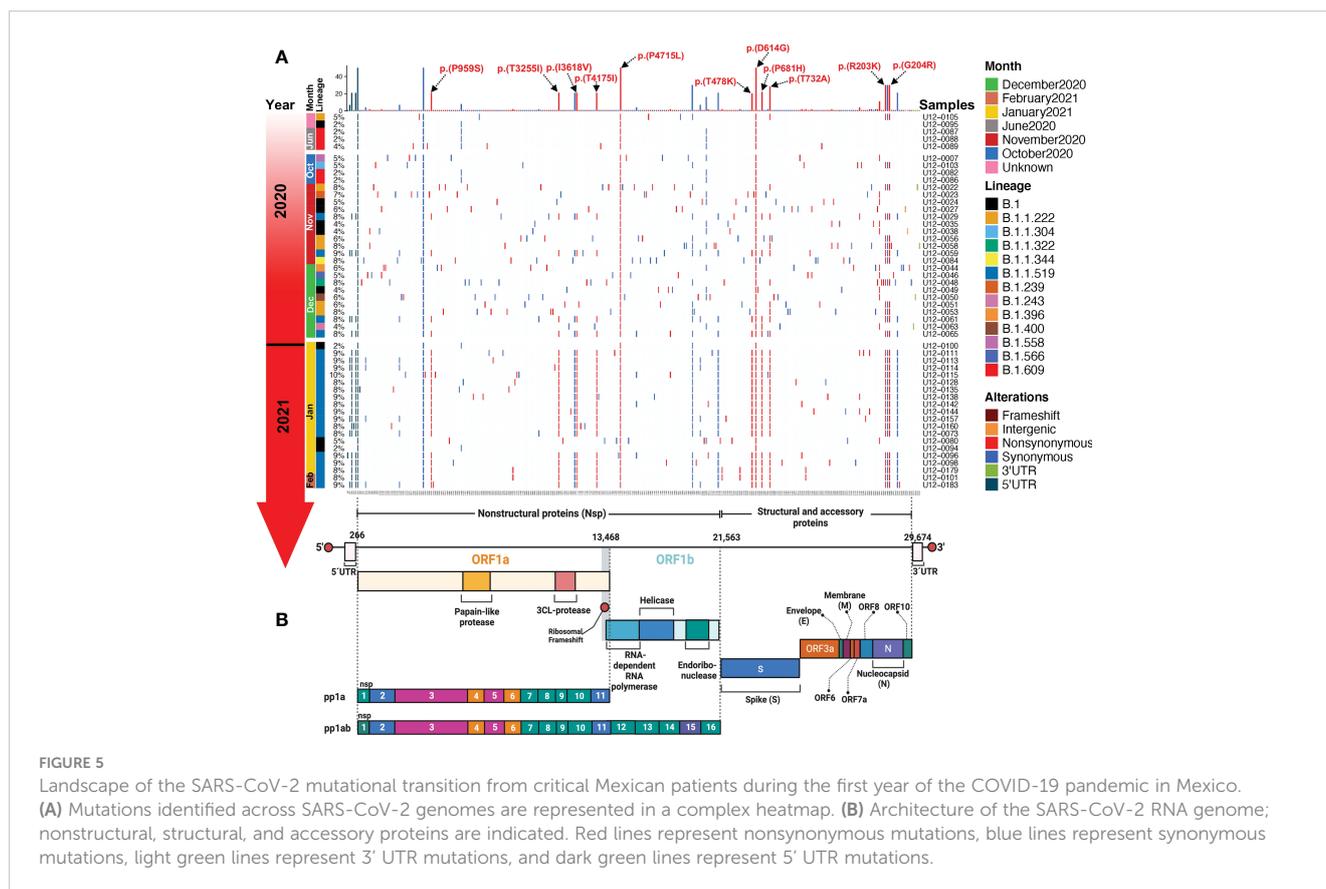
FIGURE 5
Landscape of the SARS-CoV-2 mutational transition from critical Mexican patients during the first year of the COVID-19 pandemic in Mexico.
(A) Mutations identified across SARS-CoV-2 genomes are represented in a complex heatmap. **(B)** Architecture of the SARS-CoV-2 RNA genome; nonstructural, structural, and accessory proteins are indicated. Red lines represent nonsynonymous mutations, blue lines represent synonymous mutations, light green lines represent 3' UTR mutations, and dark green lines represent 5' UTR mutations.

2007). Therefore, the quantitative role of RNA transcription in SARS-CoV-2 mutation accumulation requires further studies, particularly in the context of the pandemic situation.

Oxidative stress inside the cells is another proposed mutational process of the SARS-CoV-2 genome. Recent work suggests that G>T transversions across the SARS-CoV-2 genome might be the consequence of RNA mutational damage in periods of oxidative stress (Graudenzi et al., 2021; Mourier et al., 2021). Reactive oxidative species (ROS) are broadly associated with DNA mutations that produce 8-oxoguanine from guanines. 8-oxoguanine bases do not block DNA synthesis and are copied as adenines, resulting in G>T changes (David et al., 2007). ROS are known to target single-stranded RNA (Li et al., 2006) and their production during viral infections could generate G>T transversions in SARS-CoV-2 genomes. Consistent with our findings, there is evidence that supports higher frequencies of G>T than T>G or C>A changes are present in coronaviruses rather than other RNA viruses (Simmonds, 2020).

A third mutational process affecting the SARS-CoV-2 genome can be RNA editing. During replication, RNA viruses are susceptible to acquiring mutations due to RNA editing antiviral mechanisms in vertebrate cells. Among the most studied mechanisms are adenosine deaminase acting on RNA type 1 (ADAR1) and the apolipoprotein B mRNA editing enzyme catalytic polypeptide (APOBEC) family of enzymes. ADAR1 catalyzes adenine deamination to form inosine which successively is copied as guanine by RNA polymerases and generates A>G/T>C mutations (Samuel, 2011; Harris and Dudley, 2015). It has been

reported that the interferon-inducible isoform of ADAR1 may be involved in the excess of T>C and A>C changes in some measles viruses (Bass, 2002) and that APOBEC members can produce HIV-1 hypermutated proviral DNA copies (Vartanian et al., 1994). Members of the APOBEC family introduce mutations of cytidine to thymidine in single-stranded DNA during reverse transcription in the hepatitis B virus and some retroviruses (Suspène et al., 2005; Harris and Dudley, 2015). Whether the excess of C>T changes along SARS-CoV-2 genomes is a consequence of APOBEC-driven mechanisms or not has been widely discussed (Di Giorgio et al., 2020; Simmonds, 2020; Graudenzi et al., 2021; Ratcliff and Simmonds, 2021). Different studies have reported APOBEC3A-mediated mRNA editing and a 1.5 to 3-fold increment in C>T transitions with a 5'U context compared to other upstream bases within the SARS-CoV-2 genome (Sharma et al., 2015; Sharma et al., 2017). More recently, it was reported that APOBEC3A (A3A) cytidine deaminase has a pivotal role in the excess of C>T changes in the SARS-CoV-2 genome, which strongly suggest that A3A is the primary host player that shapes the mutation profile in the SARS-CoV-2 genome (Nakata et al., 2023).

Another antiviral mechanism is the one carried out by zinc finger antiviral proteins (ZAP), which are cytoplasmic proteins that target RNA viruses to protect cells from infections through the recognition of CG dinucleotides (Meagher et al., 2019). The consequences of purifying selection on synonymous mutations have been discussed in previous studies; the high mutation rate in the CG dinucleotide context in the SARS-CoV-2 genome has been interpreted as an evasion mechanism to the activity of the antiviral
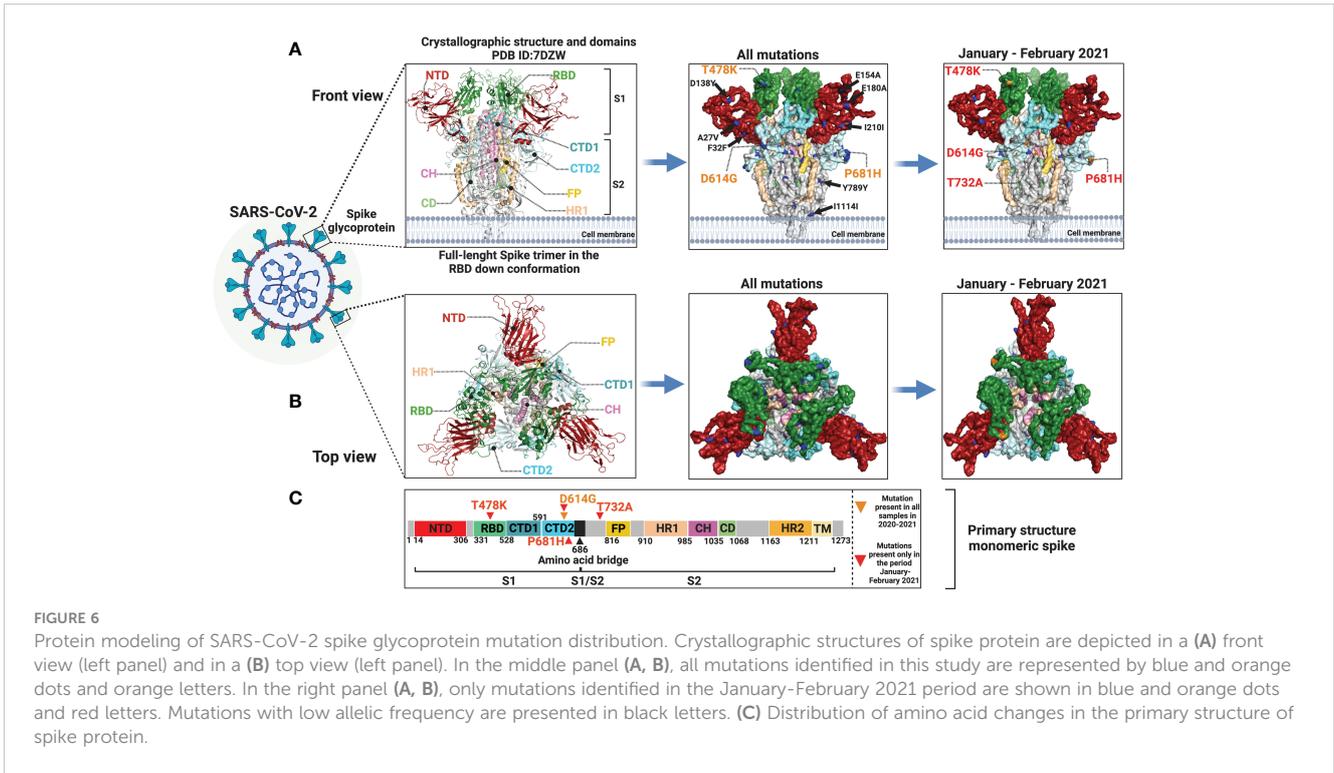
FIGURE 6
Protein modeling of SARS-CoV-2 spike glycoprotein mutation distribution. Crystallographic structures of spike protein are depicted in a **(A)** front view (left panel) and in a **(B)** top view (left panel). In the middle panel **(A, B)**, all mutations identified in this study are represented by blue and orange dots and orange letters. In the right panel **(A, B)**, only mutations identified in the January-February 2021 period are shown in blue and orange dots and red letters. Mutations with low allelic frequency are presented in black letters. **(C)** Distribution of amino acid changes in the primary structure of spike protein.

proteins ZAP of the host and has been reported in other coronaviruses (Woo et al., 2007; Xia, 2020; Mourier et al., 2021). Our results showed an increase in mutations in the context of the CG dinucleotide which can be interpreted as a decrease in CG dinucleotides that can be targeted by ZAP, evading degradation.

In addition, a report showed that genome-scale RNA secondary structure was significantly associated with C>T/T>C transitions in SARS-CoV-2, which could be an indication that C>T changes may be influenced by RNA structure (Simmonds and Ansari, 2021). All

FIGURE 7
Phylogenetic tree of a SARS-CoV-2 strains from COVID-19 critical patients in Mexico City. The root of the phylogenetic tree is the reference genome isolated in Wuhan-Hu-1 (NC_045512). Strains are distributed by clade, and samples sequenced in this study are represented by red dots. The set of these mutations precedes the origin of the omicron variant. The phylogenetic tree was generated and modified for display purposes from Nextstrain (https://nextstrain.org/ncov). Emergence dates from Delta and Omicron variants are shown.

these mutational asymmetry mechanisms still need to be fully characterized but evidence suggests that the increase in C>T changes across the SARS-CoV-2 genome is relevant in nucleotide variability and has an important role in viral sequence diversification and evolution.

From the perspective of molecular epidemiology, we observed two general genomic patterns of variant distribution in our data: a heterogeneous pattern detected in samples from 2020 and a more homogeneous arrangement from January 2021 onwards. The genetic composition of the viral genomes showed a reduction in the number of genetic mutations that were dispersed in the whole genome towards the fixation of mutations in the ORF1a, ORF1b, S, and N regions. This reduction of genetic variation leads to the appearance and eventual dominance of the B.1.1.519 variant and the loss of other genetic variants that may have lower biological fitness.

The B.1.1.519 SARS-CoV-2 lineage was reported to harbor 20 mutations, from which 11 are nonsynonymous and four are within the spike protein. This variant was first detected in Mexico in November 2020, and in August 2021 was reported as a potential variant of interest (VOI). This lineage was derived from B.1.1.222 which was characterized by the presence of the spike mutation T478K. In addition, the B.1.1.519 lineage also presents P681H and T732A spike mutations. By May 2021 it had spread to 31 countries and was predominantly found in Mexico, followed by the USA, Canada, and Germany (Rodríguez-Maldonado et al., 2021).

The S protein is a key element of the viral-host molecular interaction. Therefore, any structural changes in this protein are crucial for the infectivity, evolution, and immunological detection

FIGURE 8
Landscape of the SARS-CoV-2 molecular transition from critical patients during the first year of the COVID-19 pandemic in Mexico.

of SARS-CoV-2 (Huang et al., 2020). The S aa changes could induce neutralization escape by antibodies generated by vaccines (Garcia-Beltran et al., 2021; Koyama et al., 2022). We detected the S:D614G substitution in all our samples. This mutation was first described in January 2020 and originated the B.1 lineage (Hughes et al., 2022). In May 2020, its prevalence started to rise, and by June 2020 was reported in approximately 74% of all SARS-CoV-2 published sequences (Yurkovetskiy et al., 2020). The D614G mutation is known to enhance SARS-CoV-2 infectivity, competitive fitness,

transmission, and viral replication in animal models and human cells (Hou et al., 2020; Korber et al., 2020; Plante et al., 2021). This mutation emerged because of natural selection in the spike protein in the SARS-CoV-2 which caused the global pandemic (Zhang et al., 2020). Later, additional mutations appeared in several domains of the spike protein providing an increased fitness to the SARS-CoV-2, resulting in an increased threat to human health (Magazine et al., 2022). The T478K mutation had a high prevalence in Mexico and a rapid expansion (Di Giacomo et al., 2021). This mutation is located

in the ACE2 interaction domain and increases its positive charge. It has been reported that this mutation enhanced RBD–ACE2 complex stabilization (Cherian et al., 2021). Other amino acid change detected in S was P681H, which has been described to reduce endosomal cathepsins dependency and to augment cell surface access (Lista et al., 2022). However, other study concluded that this mutation does not significantly affect viral cellular entry (Lubinski et al., 2022). Although the T732A mutation has been reported in Mexico, no functional studies have been done to determine its biological effects (Barona-Gómez et al., 2021; Cedro-Tanda et al., 2021; Rodríguez-Maldonado et al., 2021).

During the pandemic, SARS-CoV-2 variants have undergone extensive evolutionary transitions and ultimately achieved global dominance (Telenti et al., 2022). In México, the first case of SARS-CoV-2 was reported on 28 February 2020 (Garcés-Ayala et al., 2020) while the first patient carrying variant B.1.1.519 was detected in Mexico City on 3 November 2020. Our analysis of SARS-CoV-2 evolution in Mexico between June 2020 and February 2021 suggests a transition of variants from B.1.1.222 to B.1.1.519. The variant B.1.1.222 is characterized by the S:D614G mutation, whereas B.1.1.519 has the additional aa changes T478K, P681H, and T732A in the spike protein. In November and December 2020, we detected 27.3% (3/11) and 30% (3/10) of patients with B.1.1.222, respectively, but none of the patients from 2021 had this variant. The prevalence of B.1.1.519 increased from 18.2% (2/11) and 20% (2/10) in November and December 2020, respectively, to 82.35% (14/17) in January 2021. Although our sample is scarce, these findings are aligned with previous reports from Mexico City where B.1.1.222 prevalence decreased in the later months of 2020, while B.1.1.519 increased in the first months of 2021, corresponding to the second COVID-19 wave in Mexico (Cedro-Tanda et al., 2021; Rodríguez-Maldonado et al., 2021; Taboada et al., 2021).

Our phylogenetic analysis showed that the 50 genomes of SARS-CoV-2 analyzed were grouped in three main clusters, according to NextClade. The B.1.1.519 and B.1.1.222 variants were grouped in independent clades, with B.1.1.222 derived from the 20A clade and B.1.1.519 from the 20B clade. These variants shared a most recent common ancestor with the mutation D614G in the spike protein, however, B.1.1.519 acquired the three additional aa changes previously described.

Pathogenic viruses have shown genetic evolution inside the host. Some reports have described intrapatient heterogeneity in Ebola, Middle East respiratory syndrome coronavirus, and SARS-CoV-2, characterized by the coexistence of several viral subclones (Ni et al., 2016; Park et al., 2016; Wang et al., 2021). In our dataset, the mean allelic fraction was 98.29% (range 72.3-100%; data not shown), and mutation sequencing depth mean was (15,867X; range 14-150,328; data not shown). In addition, with this analysis every patient only presented a single group of mutations, an indication of one single viral variant per patient. Therefore, even by sequencing at a high depth there was no clear evidence of subclonal variants, which suggests that all viruses might be clonal in the patients. However, since our samples were obtained from critical patients that have a long disease course, we cannot discard the possibility of SARS-CoV-2 evolving withing the host, especially in the patients

from the first wave (2020), where there was significant variability in the circulating viral variants.

We are aware that our study has some limitations that we believe do not affect the quality or the results obtained, which include (i) the lack of clinical variables and therefore their association with the disease, (ii) the sample size, which may appear to be small when compared with global epidemiological works but is similar to other SARS-CoV-2 genomic studies (Taboada et al., 2020; Elizondo et al., 2021; Yavarian et al., 2022), and (iii) the temporality of the samples.

The evolutionary behavior of SARS-CoV-2 throughout the pandemic has challenged vaccine development and has complicated the landscape for current treatments due to the adaptive mechanisms of the virus in response to the host immune system. This work on the mutational profile of the SARS-CoV-2 genome may contribute to a better understanding of the evolutionary dynamics and selective pressures acting on the SARS-CoV-2 genome, the prediction of more accurate variants of clinical significance, and a better understanding of the molecular mechanisms driving SARS-CoV-2 evolution to improve vaccine and drug development.

Data availability statement

The data presented in the study are deposited in the Sequence Read Archive repository, accession number PRJNA967733.

Ethics statement

The studies involving human participants were reviewed and approved by National Institute of Respiratory diseases Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Funding

19 contingency (CONACyT-COVID-19), with the project number 00311999 to FA-M.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500 (7463), 415–421. doi: 10.1038/nature12477

Amanat, F., and Krammer, F. (2020). SARS-CoV-2 vaccines: status report. *Immunity* 52, 583–589. doi: 10.1016/j.immuni.2020.03.007

Barona-Gómez, F., Delaye, L., Díaz-Valenzuela, E., Plisson, F., Cruz-Pérez, A., Díaz-Sánchez, M., et al. (2021). Phylogenomics and population genomics of SARS-CoV-2 in Mexico during the pre-vaccination stage reveals variants of interest B.1.1.28.4 and B.1.1.222 or B.1.1.519 and the nucleocapsid mutation S194L associated with symptoms. *Microb. Genomics* 7, 684. doi: 10.1099/mgen.0.000684

Bass, B. L. (2002). RNA Editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846. doi: 10.1146/annurev.biochem.71.110601.135501

Cedro-Tanda, A., Gómez-Romero, L., Alcaraz, N., de Anda-Jauregui, G., Peñaloza, F., Moreno, B., et al. (2021). The evolutionary landscape of SARS-CoV-2 variant B.1.1.519 and its clinical impact in Mexico city. *Viruses* 13 (11), 2182. doi: 10.3390/v13112182

Chan, C. (2022). NIH, IDSA update COVID-19 treatment guidelines as omicron makes comeback. *Pharm. Today* 28, 22–23. doi: 10.1016/j.ptdy.2022.04.006

Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., et al. (2021). SARS-CoV-2 spike mutations, L452R, T478K, E484Q and P681R, in the second wave of COVID-19 in maharashtra, India. *Microorganisms* 9, 1542. doi: 10.3390/microorganisms9071542

David, S. S., O'Shea, V. L., and Kundu, S. (2007). Base-excision repair of oxidative DNA damage. *Nature* 447, 941–950. doi: 10.1038/nature05978

DeLano, W. L. (2002) *Pymol: an open-source molecular graphics tool. CCP4 newsl. protein crystallogr.* Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.231.5879&rep=rep1&type=pdf#page=44.

Dellicour, S., Durkin, K., Hong, S. L., Vanmechelen, B., Martí-Carreras, J., Gill, M. S., et al. (2021). A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *Mol. Biol. Evol.* 38, 1608–1613. doi: 10.1093/molbev/msaa284

Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D., and Baric, R. S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* 8, 270–279. doi: 10.4161/rna.8.2.15013

Di Giacomo, S., Mercatelli, D., Rakhimov, A., and Giorgi, F. M. (2021). Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike mutation T478K. *J. Med. Virol.* 93, 5638–5643. doi: 10.1002/jmv.27062

Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G., and Conticello, S. G. (2020). Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci. Adv.* 6, eabb5813. doi: 10.1126/sciadv.abb5813

Egloff, M.-P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., et al. (2004). The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3792–3796. doi: 10.1073/pnas.0307877101

Elizondo, V., Harkins, G. W., Mabvakure, B., Smidt, S., Zappile, P., Marier, C., et al. (2021). SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. *Emerg. Microbes Infect.* 10, 51–65. doi: 10.1080/22221751.2020.1863747

Freistadt, M. S., Vaccaro, J. A., and Eberle, K. E. (2007). Biochemical characterization of the fidelity of poliovirus RNA-dependent RNA polymerase. *Virol. J.* 4, 44. doi: 10.1186/1743-422x-4-44

Garcés-Ayala, F., Araiza-Rodríguez, A., Mendieta-Condado, E., Rodríguez-Maldonado, A. P., Wong-Arámbula, C., Landa-Flores, M., et al. (2020). Full genome sequence of the first SARS-CoV-2 detected in Mexico. *Arch. Virol.* 165, 2095–2098. doi: 10.1007/s00705-020-04695-3

Garcia-Beltran, W. F., Lam, E. C., St Denis, K., Nitido, A. D., Garcia, Z. H., Hauser, B. M., et al. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* 184 (9), 2372–2383.e9. doi: 10.1016/j.cell.2021.03.013

Graudenzi, A., Maspero, D., Angaroni, F., Piazza, R., and Ramazzotti, D. (2021). Mutational signatures and heterogeneous host response revealed *via* large-scale characterization of SARS-CoV-2 genomic diversity. *iScience* 24, 102116. doi: 10.1016/j.isci.2021.102116

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi: 10.1093/bioinformatics/btw313

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., et al. (2018). Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34, 4121–4123. doi: 10.1093/bioinformatics/bty407

Harris, R. S., and Dudley, J. P. (2015). APOBECs and virus restriction. *Virology* 479-480, 131–145. doi: 10.1016/j.virol.2015.03.012

Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. *Nature* 584 (7819), 154–156. doi: 10.2210/pdb6yyt/pdb

Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K. H.3rd, et al. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission *in vivo*. *Science* 370, 1464–1468. doi: 10.1126/science.abe8499

Huang, Y., Yang, C., Xu, X.-F., Xu, W., and Liu, S.-W. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41, 1141–1149. doi: 10.1038/s41401-020-0485-4

Hughes, L., Gangavarapu, K., Latif, A. A., Mullen, J., Alkuzweny, M., Hufbauer, E., et al (2022). Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Res. Sq.* rs.3.rs-1723829 doi: 10.21203/rs.3.rs-1723829/v1

Jin, J.-M., Bai, P., He, W., Wu, F., Liu, X.-F., Han, D.-M., et al. (2020). Gender differences in patients with COVID-19: focus on severity and mortality. *Front. Public Health* 8, 152. doi: 10.3389/fpubh.2020.00152

Kati, W. M., Johnson, K. A., Jerva, L. F., and Anderson, K. S. (1992). Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* 267, 25988–25997. doi: 10.1016/s0021-9258(18)35706-5

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827.e19. doi: 10.1016/j.cell.2020.06.043

Kouranov, A., Xie, L., de la Cruz, J., Chen, L., Westbrook, J., Bourne, P. E., et al. (2006). The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.* 34, D302–D305. doi: 10.1093/nar/gkj120

Koyama, T., Miyakawa, K., Tokumasu, R., Jeremiah, S. S., Kudo, M., and Ryo, A. (2022). Evasion of vaccine-induced humoral immunity by emerging sub-variants of SARS-CoV-2. *Future Microbiol.* 17, 417–424. doi: 10.2217/fmb-2022-0025

Kubik, S., Marques, A. C., Xing, X., Silvery, J., Bertelli, C., De Maio, F., et al. (2021). Recommendations for accurate genotyping of SARS-CoV-2 using amplicon-based sequencing of clinical samples. *Clin. Microbiol. Infect.* 27, 1036.e1–1036.e8. doi: 10.1016/j.cmi.2021.03.029

Li, Q., Nie, J., Wu, J., Zhang, L., Ding, R., Wang, H., et al. (2021). SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape. *Cell* 184, 2362–2371.e9. doi: 10.1016/j.cell.2021.02.042

Li, Z., Wu, J., and Deleo, C. J. (2006). RNA Damage and surveillance under oxidative stress. *IUBMB Life* 58, 581–588. doi: 10.1080/15216540600946456

Lista, M. J., Winstone, H., Wilson, H. D., Dyer, A., Pickering, S., Galao, R. P., et al. (2022). The P681H mutation in the spike glycoprotein of the alpha variant of SARS-CoV-2 escapes IFITM restriction and is necessary for type I interferon resistance. *J. Virol.* 96, e0125022. doi: 10.1128/jvi.01250-22

Lubinski, B., Fernandes, M. H. V., Frazier, L., Tang, T., Daniel, S., Diel, D. G., et al. (2022). Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience* 25, 103589. doi: 10.1016/j.isci.2021.103589

Lucas, M., Karrer, U. R. S., Lucas, A., and Klenerman, P. (2008). Viral escape mechanisms - escapology taught by viruses. *Int. J. Exp. Pathol.* 82, 269–286. doi: 10.1046/j.1365-2613.2001.00204.x

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. *Proc. Natl. Acad. Sci.* 112, 9436–9441. doi: 10.1073/pnas.1508686112

Magazine, N., Magazine, N., Zhang, T., Wu, Y., McGee, M. C., Veggiani, G., et al. (2022). Mutations and evolution of the SARS-CoV-2 spike protein. *Viruses* 14, 640. doi: 10.3390/v14030640

Meagher, J. L., Takata, M., Gonçalves-Carneiro, D., Keane, S. C., Rebendenne, A., Ong, H., et al. (2019). Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proc. Natl. Acad. Sci.* 116, 24303–24309. doi: 10.1073/pnas.1913232116

Miao, Z., Tidu, A., Eriani, G., and Martin, F. (2021). Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol.* 18, 447–456. doi: 10.1080/15476286.2020.1814556

Michel, C. J., Mayer, C., Poch, O., and Thompson, J. D. (2020). Characterization of accessory genes in coronavirus genomes. *Virol. J.* 17, 131. doi: 10.1186/s12985-020-01402-1

Mourier, T., Sadykov, M., Carr, M. J., Gonzalez, G., Hall, W. W., and Pain, A. (2021). Host-directed editing of the SARS-CoV-2 genome. *Biochem. Biophys. Res. Commun.* 538, 35–39. doi: 10.1016/j.bbrc.2020.10.092

Nakata, Y., Ode, H., Kubota, M., Kasahara, T., Matsuoka, K., Sugimoto, A., et al. (2023). Cellular APOBEC3A deaminase drives mutations in the SARS-CoV-2 genome. *Nucleic Acids Res.* 51, 783–795. doi: 10.1093/nar/gkac1238

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., et al. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. *Biochim. Biophys. Acta Mol. Basis Dis.* 1866, 165878. doi: 10.1016/j.bbadis.2020.165878

Ni, M., Chen, C., Qian, J., Xiao, H.-X., Shi, W.-F., Luo, Y., et al. (2016). Intra-host dynamics of Ebola virus during 2014. *Nat. Microbiol.* 1, 16151. doi: 10.1038/nmicrobiol.2016.151

Ortiz-Hernández, L., and Pérez-Sastré, M. A. (2020). Social inequalities in the progression of COVID-19 in the Mexican population. *Rev. Panam. Salud Publica* 44, e106. doi: 10.26633/RPSP.2020.106

Palacio-Mejía, L. S., Hernández-Ávila, J. E., Hernández-Ávila, M., Dyer-Leal, D., Barranco, A., Quezada-Sánchez, A. D., et al. (2022). Leading causes of excess mortality in Mexico during the COVID-19 pandemic 2020–2021: a death certificates study in a middle-income country. *Lancet Regional Health - Americas* 13, 100303. doi: 10.1016/j.lana.2022.100303

Park, D., Huh, H. J., Kim, Y. J., Son, D.-S., Jeon, H.-J., Im, E.-H., et al. (2016). Analysis of intrapatient heterogeneity uncovers the microevolution of middle East respiratory syndrome coronavirus. *Cold Spring Harb. Mol. Case Stud.* 2, a001214. doi: 10.1101/mcs.a001214

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592 (7852), 116–121. doi: 10.1038/s41586-020-2895-3

Prentice, E., Jerome, W. G., Yoshimori, T., Mizushima, N., and Denison, M. R. (2004). Coronavirus replication complex formation utilizes components of cellular autophagy. *J. Biol. Chem.* 279 (11), 10136–10141. doi: 10.1074/jbc.M306124200

Ramazzotti, D., Angaroni, F., Maspero, D., Mauri, M., D'Aliberti, D., Fontana, D., et al. (2022). Large-Scale analysis of SARS-CoV-2 synonymous mutations reveals the adaptation to the human codon usage during the virus evolution. *Virus evolution* 8 (1), veac026. doi: 10.1093/ve/veac026

Ratcliff, J., and Simmonds, P. (2021). Potential APOBEC-mediated RNA editing of the genomes of SARS-CoV-2 and other coronaviruses and its impact on their longer term evolution. *Virology* 556, 62–72. doi: 10.1016/j.virol.2020.12.018

Rodríguez-Maldonado, A. P., Vázquez-Pérez, J. A., Cedro-Tanda, A., Taboada, B., Boukadida, C., Wong-Arámbula, C., et al. (2021). Emergence and spread of the potential variant of interest (VOI) B.1.1.519 of SARS-CoV-2 predominantly present in Mexico. *Arch. Virol.* 166, 3173–3177. doi: 10.1007/s00705-021-05208-6

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of Large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Samuel, C. E. (2011). Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* 411, 180–193. doi: 10.1016/j.virol.2010.12.004

Sharma, S., Patnaik, S. K., Kemer, Z., and Baysal, B. E. (2017). Transient overexpression of exogenous APOBEC3A causes c-to-U RNA editing of thousands of genes. *RNA Biol.* 14, 603–610. doi: 10.1080/15476286.2016.1184387

Sharma, S., Patnaik, S. K., Taggart, R. T., Kannisto, E. D., Enriquez, S. M., Gollnick, P., et al. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat. Commun.* 6, 6881. doi: 10.1038/ncomms7881

Simmonds, P. (2020). Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories. *mSphere* 5 (3), e00408-20. doi: 10.1128/msphere.00408-20

Simmonds, P., and Ansari, M. A. (2021). Extensive c->U transition biases in the genomes of a wide range of mammalian RNA viruses; potential associations with

transcriptional mutations, damage- or host-mediated editing of viral RNA. *PloS Pathog.* 17, e1009596. doi: 10.1371/journal.ppat.1009596

Strumillo, S. T., Kartavykh, D., Carvalho, F. F., Cruz, N. C., Souza Teodoro, A. C., Diaz, R., et al. (2021). Host–virus interaction and viral evasion. *Cell Biol. Int* 45 (6), 1124–1147. doi: 10.1002/cbin.11565

Suspène, R., Guétard, D., Henry, M., Sommer, P., Wain-Hobson, S., and Vartanian, J.-P. (2005). Extensive editing of both hepatitis b virus DNA strands by APOBEC3 cytidine deaminases *in vitro* and *in vivo. Proc. Natl. Acad. Sci.* 102, 8321–8326. doi: 10.1073/pnas.0408223102

Taboada, B., Vazquez-Perez, J. A., Muñoz-Medina, J. E., Ramos-Cervantes, P., Escalera-Zamudio, M., Boukadida, C., et al. (2020). Genomic analysis of early SARS-CoV-2 variants introduced in Mexico. *J. Virol.* 94 (18), e01056-20. doi: 10.1128/JVI.01056-20

Taboada, B., Zárate, S., Iša, P., Boukadida, C., Vazquez-Perez, J. A., Muñoz-Medina, J. E., et al. (2021). Genetic analysis of SARS-CoV-2 variants in Mexico during the first year of the COVID-19 pandemic. *Viruses* 13 (11), 2161. doi: 10.3390/v13112161

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Telenti, A., Hodcroft, E. B., and Robertson, D. L. (2022). The evolution and biology of SARS-CoV-2 variants. *Cold Spring Harbor Perspect. Med.* 12, a041390. doi: 10.1101/cshperspect.a041390

Tubiana, L., Božič, A. L., Micheletti, C., and Podgornik, R. (2015). Synonymous mutations reduce genome compactness in icosahedral ssRNA viruses. *Biophys. J.* 108, 194–202. doi: 10.1016/j.bpj.2014.10.070

Vartanian, J. P., Meyerhans, A., Sala, M., and Wain-Hobson, S. (1994). G->A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc. Natl. Acad. Sci.* 91, 3092–3096. doi: 10.1073/pnas.91.8.3092

Villanueva, R. A. M., and Chen, Z. J. (2019). "ggplot2: elegant graphics for data analysis *2nd ed*," in *Measurement: Interdisciplinary Research and Perspectives* 17 (3), 160–167. doi: 10.1080/15366367.2019.1565254

Wang, Y., Wang, D., Zhang, L., Sun, W., Zhang, Z., Chen, W., et al. (2021). Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. *Genome Med.* 13, 30. doi: 10.1186/s13073-021-00847-5

Ward, C. D., Stokes, M. A., and Flanegan, J. B. (1988). Direct measurement of the poliovirus RNA polymerase error frequency *in vitro. J. Virol.* 62, 558–562. doi: 10.1128/jvi.62.2.558-562.1988

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777. doi: 10.1093/molbev/msx335

Woo, P. C. Y., Wong, B. H. L., Huang, Y., Lau, S. K. P., and Yuen, K.-Y. (2007). Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in coronaviruses. *Virology* 369, 431–442. doi: 10.1016/j.virol.2007.08.010

Worldometers.info (2023) *COVID-19 in Mexico.* Available at: https://www.worldometers.info/coronavirus/country/mexico/ (Accessed April 3, 2023).

Xia, X. (2020). Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. *Mol. Biol. Evol.* 37, 2699–2705. doi: 10.1093/molbev/msaa094

Yao, H., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, M., et al. (2020). Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity *in vitro* and with clinical implications *in vivo. Cell Discovery* 6, 76. doi: 10.1038/s41421-020-00226-1

Yavarian, J., Nejati, A., Salimi, V., Shafiei Jandaghi, N. Z., Sadeghi, K., Abedi, A., et al. (2022). Whole genome sequencing of SARS-CoV2 strains circulating in Iran during five waves of pandemic. *PloS One* 17, e0267847. doi: 10.1371/journal.pone.0267847

Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., et al. (2020). Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* 183, 739–751.e8. doi: 10.1016/j.cell.2020.09.032

Zhang, Y.-Z., and Holmes, E. C. (2020). A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell* 181, 223–227. doi: 10.1016/j.cell.2020.03.035

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Rangarajan, E. S., Izard, T., et al. (2020). The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* 2020.06.12.148726. doi: 10.1101/2020.06.12.148726

Zhao, J., Qiu, J., Aryal, S., Hackett, J., and Wang, J. (2020). The RNA architecture of the SARS-CoV-2 3'-untranslated region. *Viruses* 12, 1473. doi: 10.3390/v12121473

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017

Ziebuhr, J., Gorbalenya, A. E., and Snijder, E. J. (2000). Virus-encoded proteinases and proteolytic processing in the nidovirales. *J. Gen. Virol.* 81, 853–879. doi: 10.1099/0022-1317-81-4-853