Genomic Characterization of SARS-CoV-2 Isolated from Patients with Distinct Disease Outcomes in Mexico

Celia Boukadida,[a] Blanca Taboada,[b] Marina Escalera-Zamudio,[c] Pavel Isa,[b] José Ernesto Ramírez-González,[d]
Joel Armando Vazquez-Perez,[e] José Esteban Muñoz-Medina,[f] Concepción Grajales-Muñiz,[g] Carolina González-Torres,[h]
Francisco Javier Gaytán-Cervantes,[h] Alma Rincón-Rubio,[a] Margarita Matías-Florentino,[a] Héctor Esteban Paz-Juárez,[a]
Alejandro Sanchez-Flores,[i] Edgar Mendieta-Condado,[d] Gisela Barrera-Badillo,[d] Lucía Hernández-Rivas,[d] Susana López,[b]
Irma López-Martínez,[d] Santiago Ávila-Ríos,[a] Carlos F. Arias[b]

[a]Centro de Investigación en Enfermedades Infecciosas, Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Mexico City, Mexico
[b]Departamento de Genética del Desarrollo y Fisiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico
[c]Department of Zoology, Oxford University, Oxford, United Kingdom
[d]Instituto de Diagnóstico y Referencia Epidemiológicos, Dirección General de Epidemiología, Mexico City, Mexico
[e]Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Mexico City, Mexico
[f]División de Laboratorios de Vigilancia e Investigación Epidemiológica, Instituto Mexicano del Seguro Social, Mexico City, Mexico
[g]Coordinación de Control Técnico de Insumos, Instituto Mexicano del Seguro Social, Mexico City, Mexico
[h]División de Desarrollo de la Investigación, Instituto Mexicano del Seguro Social, Mexico City, Mexico
[i]Unidad Universitaria de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

Celia Boukadida and Blanca Taboada contributed equally to this article. Author order was determined by alphabet order of last names.

ABSTRACT The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), has shown a wide spectrum of clinical manifestations ranging from asymptomatic infections to severe disease and death. Pre-existing medical conditions and age have been mainly linked to the development of severe disease; however, the potential association of viral genetic characteristics with different clinical conditions remains unclear. SARS-CoV-2 variants with increased transmissibility were detected early in the pandemics, and several variants with potential relevance for public health are currently circulating around the world. In this study, we characterized 57 complete SARS-CoV-2 genomes during the exponential growth phase of the early epidemiological curve in Mexico, in April 2020. Patients were categorized under distinct disease severity outcomes: mild disease or ambulatory care, severe disease or hospitalized, and deceased. To reduce bias related to risk factors, the patients were less than 60 years old and with no diagnosed comorbidities A trait-association phylogenomic approach was used to explore genotype–phenotype associations, represented by the co-occurrence of mutations, disease severity outcome categories, and clusters of Mexican sequences. Phylogenetic results revealed a higher genomic diversity compared to the initial viruses detected during the early stage of the local epidemic. We identified a total of 90 single nucleotide variants compared to the Wuhan-Hu-1 genome, including 54 nonsynonymous mutations. We did not find evidence for the co-occurrence of mutations associated with specific disease outcomes. Therefore, in the group of patients studied, disease severity was likely mainly driven by the host genetic background and other demographic factors.

IMPORTANCE The genetic association of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) with different clinical conditions remains unclear and needs further investigation. In this study, we characterized 57 complete SARS-CoV-2 genomes from patients in Mexico with distinct disease severity outcomes: mild disease or ambulatory care, severe disease or hospitalized, and deceased. To reduce bias related to risk factors the patients were less than 60 years old and with no diagnosed comorbidities. We did not find evidence for the co-occurrence of mutations associated with specific disease outcomes.

Therefore, in the group of patients studied, disease severity was likely mainly driven by the host genetic background and other demographic factors.

KEYWORDS SARS-CoV-2, genomic diversity, Mexico, distinct clinical outcomes

The coronavirus disease 2019 (COVID-19), declared a pandemic by the WHO on March 11th 2020, is caused by a coronavirus of zoonotic origin that was recently introduced into the human population, named the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (1). SARS-CoV-2, first detected in December 2019, belongs to the *Sarbecovirus* subgenus, *Betacoronavirus* genus of the *Coronaviridae* family (1), together with SARS-CoV-1, which caused a short-lived outbreak of great medical importance in 2002 (2). As of October 2021, the COVID-19 pandemic had caused over 240 million cases and 4.9 million deaths globally, as reported by the World Health Organization (3). Genomic epidemiology and surveillance efforts worldwide have rendered more than 2.3 million full SARS-CoV-2 genome sequences available in public data repositories such as the Global Initiative on Sharing Avian Influenza Data (GISAID) platform (4, 5). Studying virus genome sequences in an evolutionary context has allowed to explore in real time the impact of genetic changes occurring in the circulating virus populations within different geographical regions (6–8). As a result, an increasing number of SARS-CoV-2 variants displaying different mutations have become of potential relevance for public health (9). Some of these mutations encode amino acid changes within the spike (S) protein of the virus, which mediates its attachment to the host cell receptor through its globular domain and is the main target for the neutralizing antibody response (10). Some of these variants seem to be associated with immune escape or increased transmissibility (11, 12).

As an example, the SARS-CoV-2 variant displaying the amino acid change D614G in the S protein emerged during late January to early February 2020 and became predominant on a global scale during March–April 2020. The mutation D614G is associated with increased transmissibility of the virus (13–16). At the molecular level, the D614G change was demonstrated to favor a slightly more open conformation of the spike protein with added interactions that prevent premature dissociation of the trimer and modulate structural rearrangements for membrane fusion, effectively enhancing infectivity (17, 18). Comparably, recent epidemiological reports suggest that viral variants belonging to the B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617.2 (Delta) expanding lineages could also be associated with an increased transmissibility and immune escape (19). Such variants are currently classified as variants of concern (VOC) (20), and share amino acid changes in the S protein that in some cases have emerged in parallel, such as the K417N/T (in B.1.351 and P.1), N501Y (in B.1.1.7, B.1.351, and P.1), E484K (in B.1.351 and P.1) or P681H/R (in B.1.1.7 and B.1.617.2) (9). The N501Y substitution introduces an additional interaction with the angiotensin converting enzyme 2 (ACE2) receptor, increasing the binding affinity, and abolishes the binding of a potent neutralizing antibody (21, 22). The E484K change was also demonstrated to increase the binding affinity of the spike receptor binding domain (RBD) for ACE2 (23). In addition, E484K confers resistance to several monoclonal antibodies and significantly reduces neutralization by convalescent plasma and sera obtained from vaccinated individuals, indicating that this substitution is located in an immunodominant epitope (11, 19, 24).

A wide spectrum of clinical manifestations have been observed in symptomatic SARS-CoV-2 infected individuals, ranging from asymptomatic infections or mild upper-respiratory tract symptoms treated under ambulatory care, to severe pneumonia and multiorgan failure, resulting in hospitalization and treatment within the intensive care unit that may ultimately lead to death (25). Demographic characteristics such as age and sex have been linked to the development of a severe disease and/or death in the context of COVID-19 (26). Additional identified risk factors are associated with underlying medical conditions such as obesity, cancer, cardiovascular disease, diabetes, immunosuppression, as well as other respiratory, kidney, liver and neurological conditions (26–30).

The potential role of viral mutations in disease severity remains unclear, and to date, there is insufficient evidence supporting that the genetic background of SARS-CoV-2 may play a role in disease outcome (31, 32). However, as the genetic diversity of the virus

TABLE 1 Characteristics of the study participants[a]

Disease outcome category	Ambulatory care (A) n = 24	Hospitalized (H) n = 14	Deceased (D) n = 19	P values
Age (yrs), median (range)	34 (24-50)	47 (22-59)	45 (27-59)	** *P* = **0.0013**
Sex (Male), n (%)	15 (62.5%)	10 (71.4%)	15 (78.9%)	NS, *P* = *0.5004*
Location				
Mexico City, n (%)	8 (33.3%)	11 (78.6%)	11 (57.9%)	*
Other states of Mexico, n (%)	16 (66.7%)	3 (21.4%)	8 (42.1%)	***P* = 0.0227**
Symptoms	20 (83.3%)	11 (78.6%)	16 (88.9%)	NS, *P* = 0.73
Fever, n (%)	21 (87.5%)	10 (71.4%)	17 (94.4%)	NS, *P* = 0.17
Cough, n (%)	22 (91.7%)	7 (50.0%)	8 (44.4%)	** *P* = **0.0020**
Headache, n (%)	15 (62.5%)	6 (42.9%)	4 (22.2%)	* ***P* = 0.0338**
Sore throat, n (%)	16 (66.7%)	7 (50.0%)	9 (50.0%)	NS, *P* = 0.46
Fatigue, n (%)	15 (62.5%)	7 (50.0%)	10 (55.6%)	NS, *P* = 0.74
Myalgia, n (%)	16 (66.7%)	8 (57.1%)	11 (61.1%)	NS, *P* = 0.83
Arthralgia, n (%)	12 (50.0%)	6 (42.9%)	4 (22.2%)	NS, *P* = 0.18
Rhinorrhea, n (%)	3 (12.5%)	1 (7.1%)	2 (11.1%)	NS, *P* = 0.87
Conjunctivitis, n (%)	1 (4.2%)	12 (85.7%)	17 (94.4%)	**** ***P* < 0.0001**
Dyspnea, n (%)	2 (8.3%)	5 (35.7%)	2 (11.1%)	NS, *P* = 0.07
Diarrhea, n (%)	8 (33.3%)	3 (21.4%)	6 (33.3%)	NS, *P* = 0.70
Chest pain, n (%)	0	0	1	
NA, n				
Time from symptom onset to diagnosis (days), median (interquartile range)	2.5 (1–4)	8 (2.5–13.25)	6 (4.75–8.25)	** *P* = **0.0018**

[a]Kruskal-Wallis and Chi-square tests were used to compare continuous and categorical variables, respectively (NS, not significant; *, $p \leq 0.05$; **, $p \leq 0.005$; ****, $p < 0.0001$). NA, not available. In bold, the values that were statistically significant.

increases, exploring the potential impact of emerging mutations related to virulence in the human host (i.e., by increasing transmissibility, or linked to a worse disease outcomes) remains an important task. In this study, we characterized 57 complete SARS-CoV-2 genomes generated from Mexican samples derived from patients categorized under distinct disease severity outcomes: *a*mbulatory care or mild disease (A); *h*ospitalized or severe disease (H); and *d*eceased (D). To reduce bias related to risk factors in the sampled population, patient inclusion criteria were limited to less than 60 years of age and no diagnosed comorbidities. A trait-association phylogenomic approach was used to explore genotype–phenotype associations, represented by the co-occurrence of mutations, disease severity outcome categories, and clusters of Mexican sequences.

RESULTS

Clinical sample distribution. The Mexican isolates studied here were collected during the exponential growth phase of the early epidemiological curve in Mexico, from 03/26/2020 to 04/30/202, (Table S1, Fig. S1 in the supplemental material). A total of 57 genomic sequences were obtained, and for their analysis the patients were categorized into three groups according to disease severity, based on the clinical history and the official disease outcome recorded for each patient: *a*mbulatory care (A, *n* = 24), severe disease or *h*ospitalized (H, *n* = 14), and fatal or *d*eceased (D, *n* = 19) (Table 1). The samples were selected from those patients who met the inclusion criteria at the time of the study; the median age of the patients was 49 years (range 22 to 59) (Table S3).

The samples were obtained from 7 out of 32 states in the country, with those from Mexico City being predominant (Table 1). This biased pattern is likely to reflect either a higher number of reported cases in the capital city or the geographical uneven stage of the epidemic in the country (33) (Fig. S1). Differences in genomic surveillance efforts at a regional level, with most surveillance occurring in Mexico City, are also expected to influence sample distribution. Significant differences in patient age and the sites where the sample were collected were observed across the three severity groups (A, H, and D). Sex was the only category with no significant difference observed across groups (Table 1). On average, patients were significantly younger within the group "A", compared to the "D" and "H" groups, while samples from Mexico City were overrepresented within the "H" and "D" groups.
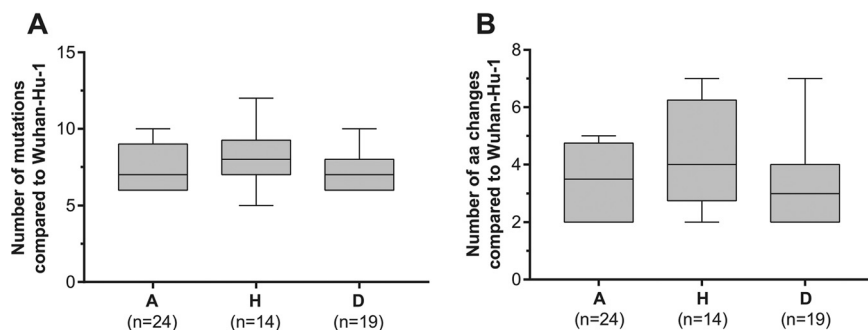
FIG 1 Genome variation compared to Wuhan-Hu-1 across severity groups. (A) Number of mutations and (B) amino acid changes, compared to the reference genome Wuhan-Hu-1 in the ambulatory care (A), hospitalized (H), and deceased (D) study groups. Boxes contain the 25th and 75th percentiles, and whiskers show the minimum and maximum values.

Differences in the clinical presentation were also observed across the three severity groups. As expected, dyspnea was markedly more frequent in hospitalized and deceased patients compared to individuals with asymptomatic/mild disease (Table 1). In contrast, at the time of diagnosis, patients in group "A" more frequently presented headache and sore throat compared to the "H" and "D" groups. Since the time from symptom onset to diagnosis was significantly longer for individuals with severe and fatal outcomes, the observed lower frequency of headache and sore throat in these patients might reflect the resolution of early COVID-19 symptoms as described previously (34, 35). Other symptoms were found at similar frequencies in the three severity groups.

Genomic and phylogenetic characterization of Mexican isolates. The 57 Mexican virus genome sequences displayed good quality, as shown by a median genome coverage of 99.9% (range: 98.5%–99.99%) and median depth of 1,274× (range: 33×–37,020×) (Table S1). The median pairwise genetic distance observed across the 57 genomes was of 6 nucleotides (range: 0–14), reflecting a low genetic diversity, consistent with observations made on viral diversity at a global scale during this period (36, 37). The Mexican sequences showed between 5 to 12 nucleotide and 2 to 7 amino acid changes compared to the reference genome (Fig. 1, Table 2, Table S1). No significant differences in the number of mutations or amino acid changes within the Mexican sequences were detected across the different groups (A, H, or D) (Fig. 1A and B, Kruskal-Wallis tests, $P = 0.19$ and $P = 0.17$ respectively).

Phylogenetic placement revealed that the 57 Mexican isolates fell within the PANGO lineage B.1 and the descending B.1.609, B.1.1, B.1.189, B.1.1.222, and B.1.111 lineages (Table S1). None of the isolates fell within the PANGO lineage A, identified in Mexico in early March 2020 (38). The lineage distribution observed, including the replacement of lineage A over B, is consistent with reports at a global scale for that particular time (13, 39, 40). No major differences in the distribution of different lineages across the different disease outcome groups (A, H, or D) were observed (Fig. 2, chi-square test, $P = 0.32$).

Phylogenetic analysis revealed a single well-supported cluster of Mexican sequences comprising 27/57 of the isolates sequenced in this study (Fig. 3). These 27 sequences belonged to PANGO lineages B.1.609 (25 isolates) and B.1.189 (2 isolates), first detected in March and April 2020, respectively, and estimated to have originated in USA/Mexico (40). Within this single cluster, four subclusters were identified (SC1-4, Fig. 3), all diverging from the same ancestral sequence (root: 17307_IMSS_29A_F_MEX_A_17/04/2020). The 30 remaining Mexican isolates sequenced for this study displayed a scattered position across the tree (Fig. 3). Additionally, only a small cluster of sequences corresponding to the earliest isolates from the country evidencing local transmission were observed to group within the PANGO lineage A, consistent with previous observations by our group (38). Our observations suggest that most of the Mexican sequences used for this study likely represent an early but established local transmission of the virus within the country, linked to the establishment of lineages B.1.609, B.1, and B.1.1 within the region (40).

Identification of single nucleotide variants (SNVs). Mutation analysis of the 57 Mexican sequences revealed 90 single nucleotide variants (SNVs) relative to the reference

TABLE 2 Description of full-length SARS-CoV-2 genomes[a]

Sample ID	Pango lineage	# of changes nt[b]	aa[b]	aa changes
25_INER_49A_M_MEX_D_26/03/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
42_INER_53A_M_CMX_D_30/03/2020	B.1.609	7	3	Orf1a:Q4189H,Orf1b:P314L,S:D614G
68_INER_40A_M_MEX_H_04/04/2020	B.1.609	8	2	Orf1b:P314L,S:D614G
75_INER_34A_M_CMX_A_04/04/2020	B.1.1.222	9	5	N:R203K,N:G204R,Orf1b:P314L,S:D614G,S:T732A
83_INER_39A_M_CMX_D_06/04/2020	B.1.609	7	3	Orf1b:P314L,S:A260S,S:D614G
85_INER_27A_M_CMX_A_06/04/2020	B.1	7	4	Orf1a:L3338F,Orf1b:P314L,Orf3a:Q57H,S:D614G
91_INER_45A_F_MEX_D_07/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
95_INER_56A_M_CMX_D_13/04/2020	B.1.609	9	3	Orf1a:I1276T,Orf1b:P314L,S:D614G
104_INER_43A_M_CMX_D_10/04/2020	B.1.111	7	4	Orf1b:P314L,Orf3a:Q57H,Orf7a:T14I,S:D614G
106_INER_36A_F_CMX_H_10/04/2020	B.1	5	2	Orf1b:P314L,S:D614G
109_INER_41A_F_MEX_D_08/04/2020	B.1	6	2	Orf1b:P314L,S:D614G
112_INER_28A_M_CMX_A_08/04/2020	B.1	7	5	Orf1a:T265I,Orf1a:D3972E,Orf1b:P314L,Orf3a:Q57H,S:D614G
118_INER_59A_M_CMX_D_08/04/2020	B.1.609	7	2	Orf1b:P314L,S:D614G
126_INER_56A_M_MEX_D_11/04/2020	B.1.1	9	5	N:R203K,N:G204R,Orf1a:A1049V,Orf1b:P314L,S:D614G
137_INER_58A_M_CMX_D_13/04/2020	B.1.1	10	7	N:R203K,N:G204R,Orf1a:I404T,Orf1a:H712Y,Orf1b:A176V,Orf1b:P314L,S:D614G
173_INER_34A_M_CMX_H_15/04/2020	B.1.1	12	6	M:A69S,N:R203K,N:G204R,Orf1b:P314L,Orf1b:L1629S,S:D614G
192_INER_39A_M_CMX_D_17/04/2020	B.1.609	8	4	Orf1a:Y1920H,Orf1a:A4285V,Orf1b:P314L,S:D614G
204_INER_52A_M_CMX_H_18/04/2020	B.1	7	5	Orf1a:T265I,Orf1b:P314L,Orf3a:Q57H,S:D614G,S:D936Y
205_INER_34A_M_CMX_A_18/04/2020	B.1	7	3	Orf1b:P314L,Orf1b:V1768I,S:D614G
207_INER_36A_M_MEX_H_17/04/2020	B.1	9	7	Orf1a:K141R,Orf1a:T265I,Orf1b:P314L,Orf1b:G2197D,Orf3a:Q57H,S:D614G, S:D936Y
221_INER_59A_M_CMX_H_18/04/2020	B.1.609	7	2	Orf1b:P314L,S:D614G
228_INER_53A_F_CMX_H_19/04/2020	B.1	9	7	Orf1a:K141R,Orf1a:T265I,Orf1b:P314L,Orf1b:G2197D,Orf3a:Q57H,S:D614G, S:D936Y
229_INER_50A_M_CMX_H_19/04/2020	B.1.609	10	5	N:S327L,Orf1a:G3072C,Orf1b:P314L,Orf3a:S166L,S:D614G
238_INER_59A_M_CMX_D_21/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
260_INER_43A_M_MEX_D_22/04/2020	B.1.609	7	2	Orf1b:P314L,S:D614G
275_INER_45A_M_CMX_D_25/04/2020	B.1	6	3	Orf1b:P314L,S:V143F,S:D614G
286_INER_48A_M_MEX_H_26/04/2020	B.1.609	8	4	Orf1a:A4285V,Orf1b:P314L,S:D614G,S:V1268I
296_INER_38A_F_CMX_H_23/04/2020	B.1	7	3	Orf1b:P314L,Orf3a:Q57H,S:D614G
317_INER_55A_F_CMX_H_27/04/2020	B.1	9	4	Orf1a:L1249F,Orf1a:N1995H,Orf1b:P314L,S:D614G
689_InDRE_38A_F_YUC_D_10/04/2020	B.1	6	3	Orf1a:T2952I,Orf1b:P314L,S:D614G
1498_InDRE_35A_M_CMX_D_15/04/2020	B.1.609	8	4	Orf1a:Y1920H,Orf1a:A4285V,Orf1b:P314L,S:D614G
5361_InDRE_27A_F_TLA_D_30/04/2020	B.1.609	8	3	Orf1a:A4285V,Orf1b:P314L,S:D614G
8279_IMSS_31A_M_CMX_D_01/04/2020	B.1	6	4	Orf1a:T265I,Orf1b:P314L,Orf3a:Q57H,S:D614G
9823_IMSS_55A_M_CMX_H_05/04/2020	B.1	6	3	N:A359S,Orf1b:P314L,S:D614G
10793_IMSS_26A_F_MEX_A_08/04/2020	B.1	8	4	Orf1a:E647G,Orf1a:S2132G,Orf1b:P314L,S:D614G
10932_IMSS_30A_M_CMX_A_08/04/2020	B.1.1	10	4	N:R203K,N:G204R,Orf1b:P314L,S:D614G
10938_IMSS_22A_M_CMX_H_08/04/2020	B.1	8	4	Orf1a:M3652T,Orf1b:P314L,Orf3a:Q57H,S:D614G
11296_IMSS_49A_M_CMX_A_07/04/2020	B.1	8	5	N:G243C,Orf1a:T4031I,Orf1b:P314L,S:D614G,S:V1122L
11398_IMSS_59A_M_GRO_D_08/04/2020	B.1.1	8	4	N:R203K,N:G204R,Orf1b:P314L,S:D614G
11433_IMSS_43A_F_BCS_A_07/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
11461_IMSS_32A_M_GRO_A_08/04/2020	B.1	10	4	Orf1a:T50I,Orf1a:S142L,Orf1b:P314L,S:D614G
11647_IMSS_33A_F_BCS_A_08/04/2020	B.1.609	7	3	Orf1a:S4398L,Orf1b:P314L,S:D614G
11688_IMSS_48A_M_BCS_A_07/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
11692_IMSS_50A_F_BCS_A_07/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
11697_IMSS_37A_M_BCS_A_07/04/2020	B.1.609	7	2	Orf1b:P314L,S:D614G
11712_IMSS_41A_F_BCS_A_08/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
11713_IMSS_35A_F_BCS_A_08/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
11768_IMSS_42A_M_BCN_A_08/04/2020	B.1.609	9	5	Orf1a:P777S,Orf1a:P3613S,Orf1b:P314L,Orf1b:S1003G,S:D614G
11769_IMSS_24A_M_BCN_A_08/04/2020	B.1.609	8	3	Orf1a:P3613S,Orf1b:P314L,S:D614G
12060_IMSS_26A_M_MEX_A_08/04/2020	B.1	8	4	E:L73F,Orf1a:T4355I,Orf1b:P314L,S:D614G
12542_IMSS_39A_M_BCS_A_06/04/2020	B.1.609	6	2	Orf1b:P314L,S:D614G
14214_IMSS_26A_M_CMX_A_16/04/2020	B.1.1	10	5	N:R203K,N:G204R,Orf1b:P314L,Orf1b:V1706I,S:D614G
14215_IMSS_47A_F_CMX_A_16/04/2020	B.1.189	10	5	Orf1a:V2047F,Orf1b:P314L,Orf7a:E95*,Orf9b:R32L,S:D614G
16144_IMSS_27A_M_MEX_A_19/04/2020	B.1.189	9	4	Orf1a:V2047F,Orf1b:P314L,Orf9b:R32L,S:D614G
16221_IMSS_46A_M_CMX_H_17/04/2020	B.1.1	11	7	N:R203K,N:G204R,Orf1a:H712Y,Orf1a:T3058I,Orf1a:L3606F,Orf1b:P314L,S:D614G
17307_IMSS_29A_F_MEX_A_17/04/2020	B.1	7	2	Orf1b:P314L,S:D614G
17360_IMSS_38A_F_GRO_A_17/04/2020	B.1	6	2	Orf1b:P314L,S:D614G

[a]nt, nucleotide; aa, amino acid.
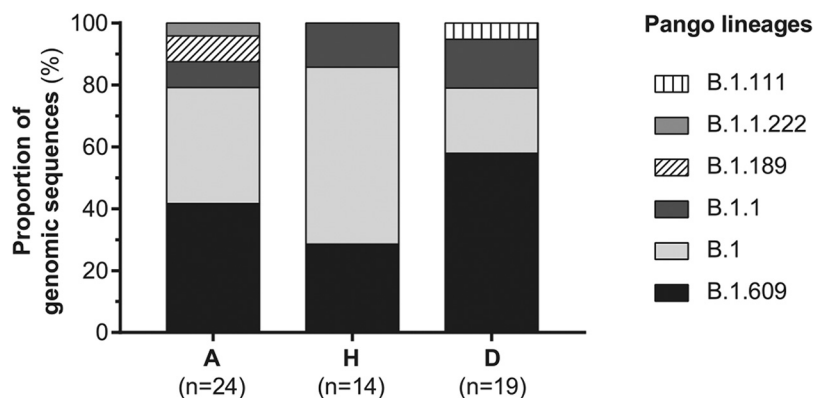[b]Compared to Wuhan-Hu-1 reference genome (GenBank MN908947).

FIG 2 Lineage distribution across severity groups. The PANGO lineages of genomic sequences belonging to the ambulatory care (A), hospitalized (H), and deceased (D) study groups were determined using the PANGOLIN tool v3.1.7 (lineages version 2021-07-09) (54).

genome. SNVs were mostly represented by 54 nonsynonymous mutations and in-Orf synonymous mutations, with a single in-frame deletion and a stop codon insertion detected (Table S4). Shared mutations between all Mexican sequences were observed mainly within Orf1ab and S (Fig. 4). Four nucleotide substitutions (C241T, C3037T, C14408T, A23403G), corresponding to lineage-defining mutations for PANGO lineage B (40), were present in all 57 Mexican virus genomes. The C14408T and A23403G nucleotide substitutions correspond to two non-synonymous mutations resulting in the P323L amino acid change within the viral RdRp (RNA-dependent RNA polymerase), and to the D614G amino acid change within S (spike protein). Nucleotide substitution C241T corresponds to a synonymous mutation located within the 5′ UTR (untranslated region), while C3037T corresponds to a synonymous substitution within Orf1ab/nsp3. Apart from these lineage-defining mutations, 74% of the remaining SNVs were observed within single genome sequences (Fig. 4, Table S4). Singletons observed are likely to reflect random evolutionary processes.

Amino acid changes in the spike protein have received special attention due to their potential effect on viral fitness and immune escape. In addition to the D614G substitution found in all 57 isolates, we identified 6 nonsynonymous substitutions within S (Fig. 4, Table S4), none of them located in the RBD. Two amino acid changes were detected in the N-terminal domain, three were located within the S2 subunit, and one in the cytoplasmic tail. However, these substitutions were identified in only one to three samples, precluding further analyses of their potential phenotypic impact.

Genotype–phenotype association analysis. Mutation analysis for the 27 Mexican isolates falling within the single cluster identified (PANGO lineages B.1.609 and B.1.189, Fig. 3) revealed six shared amino acid changes occurring across sequences: three of these within Orf1a (T265I/nsp2-85, L3606F/nsp6-38, and A4285V/nsp10-32), one within Orf1b (P314L Orf1b/RdRp-323), and two within S (D614G and D936Y) (Table S4). However, genotype–phenotype association analysis (see Materials and Methods section: "Cluster identification and genotype-phenotype association analysis") revealed that none of these mutations appear to have a structured distribution associated with the Mexican sequences, or with the complete sampled sequence data used (Fig. S2). A potential link between the sequence subclusters and the different disease outcome groups (A, H, or D) was identified (SC1 and SC2 associated with group "A", and SC3 and SC4 associated with groups "H" and "D") (Fig. 3). However, as the four subclusters descend from the same ancestor node, these cannot be considered as "independent," and any observation of genotype–phenotype associations for independent lineages is not sustained. Thus, no potential phenotype–genotype associations were detectable under a trait-association phylogenomic approach for our data set. The pattern of occurrence for shared mutations observed across the Mexican sequences suggests that most of these represent characters arising from a founder effect (41), with no apparent association with any disease outcome.
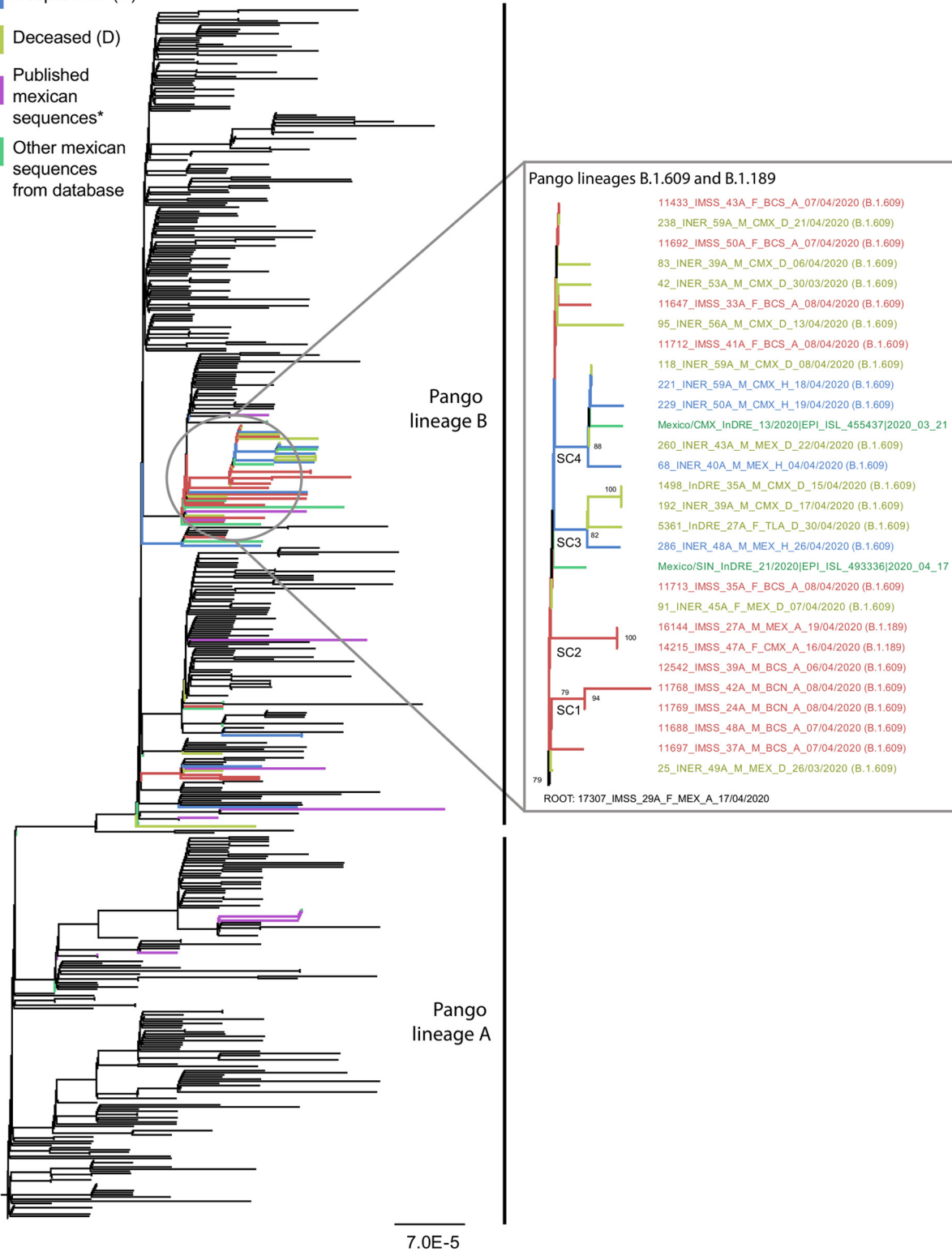
FIG 3 Phylogenetic placement of Mexican isolates. ML tree inferred from the Orf1ab+S alignment comprising 565 sequences, including the 57 Mexican virus genome sequences generated for this work, classified under the different disease outcome groups (A, indicated in red;
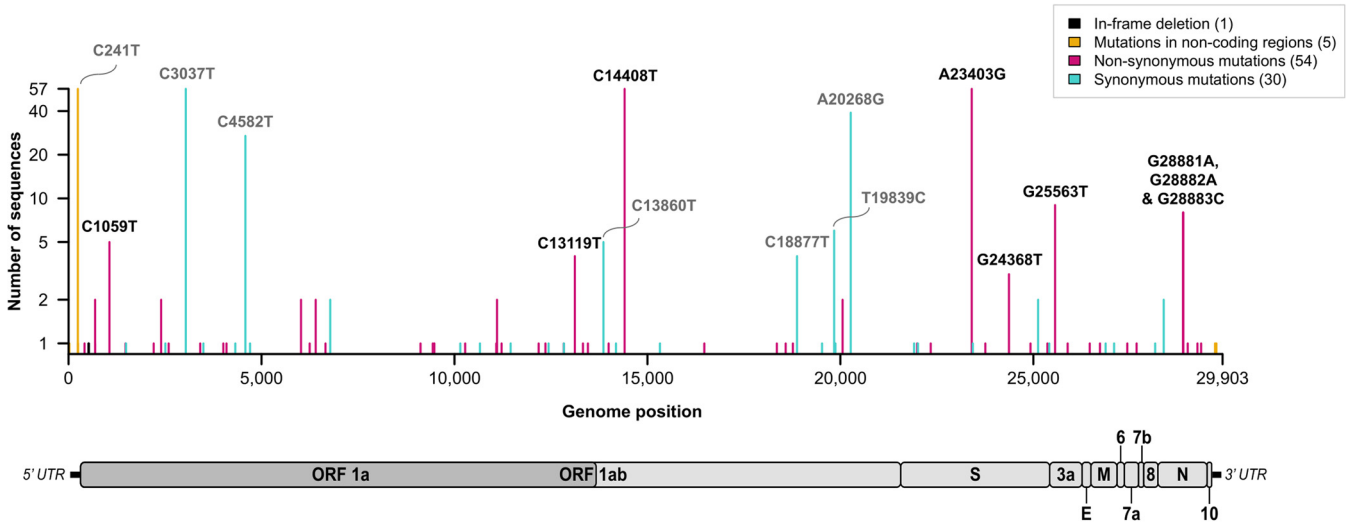
(Continued on next page)

FIG 4 Distribution of identified mutations in the SARS-CoV-2 genome. For each single nucleotide variant identified, the number of viral isolates was plotted as a function of the nucleotide position along the SARS-CoV-2 genome. Single nucleotide variants were divided into 4 categories: in-frame deletion (black bar), mutations in noncoding regions (dark yellow bars), nonsynonymous mutations (pink bars), and synonymous mutations (turquoise bars). The number of single nucleotide variants in each category is indicated in the legend box. The SARS-CoV-2 genome organization is represented below the plot.

DISCUSSION

Here we present the genomic and phylogenetic characterization of 57 SARS-CoV-2 virus genome sequences corresponding to isolates from patients with distinct disease outcomes. Our analysis revealed that genetic diversity of the characterized viruses, collected during March–April 2020 in Mexico, showed a gradual accumulation of mutations since the initial introduction of the virus until the early establishment of the epidemic in the country (38). This is shown by the local shift from the PANGO lineage A to B, and the establishment of the PANGO lineages B.1.609, B.1, and B.1.1 within the country, reflecting an extended local transmission of the virus.

We further explored any possible genotype-to-phenotype associations to detect the co-occurrence of distinct mutations with different disease outcomes. To minimize biases inherent to the host population that could impact disease severity (26), the samples used in this study were all derived from patients below 60 years of age, without any confirmed comorbidities considered to be risk factors for severe COVID-19. We found no evidence suggesting the co-occurrence of mutations associated with any particular disease outcome. Our results are consistent with previous studies from other geographical regions (28, 42), suggesting that even among individuals at low risk for severe COVID-19, the host genetic and medical background are the factors most likely determining patient clinical outcome (28). In agreement with this observation, we found that patients with mild COVID-19 (A group) were significantly younger compared to individuals belonging to the D or H severe outcome groups, in line with studies highlighting older age as a major risk factor associated with COVID-19 morbidity and mortality (26, 43).

In addition to demographic and clinical conditions, host genetic factors have also been suggested to be associated with disease severity during SARS-CoV-2 infections, increasing the risk of a severe outcome for patients who would otherwise be considered low risk based solely on demographic and/or medical condition. Some of these identified loci may be involved in regulating key host antiviral defense mechanisms or mediating inflammatory organ damage responses (such as the interferon receptor gene IFNAR2

FIG 3 Legend (Continued)

H, indicated in blue; and D, indicated in yellow). An additional 34 Mexican sequences available from the GISAID platform at the time of the analysis were also included (indicated in green), together with the 17 Mexican isolates characterized from the earliest introduction events of SARS-CoV-2 within the country (indicated in purple) (26). From the 57 isolates sequenced for this study, 27 formed a single well-supported cluster belonging to the PANGO lineages B.1.609 and B.1.189. Within this single cluster, four subclusters were identified (SC1-4), all diverging from the same ancestral sequence. The 30 remaining Mexican isolates sequenced for this study displayed a scattered positioning across the tree.

[44]), and/or may potentially involve the ABO blood-group system (45, 46). Additional studies have shown that limited access to health care systems, and other socioeconomic determinants related to marginalization and poverty, are associated with COVID-19 severity and mortality, and may confound the observed genetic associations (26, 30, 47).

Limitations of our study include a small sample size relative to the number of cases reported within the country to the date of the analysis, resulting in a heterogeneous representation of the geographical distribution of cases across the country. Moreover, a limited genetic diversity of the virus coupled with a lack of metadata related to disease severity for other sequences from Mexico and for those globally collected, limited the exhaustive exploration of the co-occurrence of associated convergent mutations linked to different disease outcomes within a more general scenario.

Genomic surveillance programs of SARS-CoV-2 have generated an exceptionally high number of viral sequences and have allowed to monitor the evolution of SARS-CoV-2 at regional and global levels. However, many questions remain concerning the potential phenotypic effects of the identified mutations. More multidisciplinary analyses integrating genomic, epidemiological, and clinical data as well as experimental virology studies using *in vitro* and *in vivo* models are needed to further characterize the sets of mutations observed in circulating variants and their potential effects on viral fitness, virulence, and antigenicity.

MATERIALS AND METHODS

Ethical statement and sample collection. The samples used in this study and their associated data are part of the national response to COVID-19, collected under Mexican Official NOM-017-SSA2-2012 for epidemiological surveillance. Clinical specimens were collected in laboratories and hospitals belonging to the Public Health Ministry of Mexico (Red Nacional de Laboratorios de Salud Publica (RNLSP); Instituto Nacional de Enfermedades Respiratorias, INER; and Instituto Mexicano del Seguro Social, IMSS). All samples were unlinked from any personal identifiers prior to commencement of the study. Oropharyngeal and/or naso-pharyngeal swabs and tracheal aspirates were placed in virus transport media following official procedures for diagnostics and sequencing processing (48). Samples were collected from PCR-confirmed cases, with molecular diagnosis carried out following protocols validated by the InDRE (Instituto de Diagnóstico y Referencia Epidemiológicos, Secretaría de Salud, Mexico, National Reference Laboratory for COVID-19), as recommended by the World Health Organization (49). All samples were obtained from individuals under 60 years of age and with no diagnosed medical conditions linked to severe COVID-19 (chronic respiratory disease, cardiovascular disease, diabetes, immunosuppressive condition, and obesity class II and III).

Sample processing, sequencing, and viral genome assembly. Viral nucleic acid extraction, amplification, and sequencing was done as previously described (38) with slight modifications. Clinical samples that failed to render full viral genome sequences using a high-throughput approach were further sequenced under an amplicon-based protocol with the POLAR method (50) (Table S1). Sequence quality control and assembly was performed as described (38). Only full virus genome sequences with a coverage >98% and a mean depth >30× were considered for further analysis (Table S1).

Sequence data collation. Approximately 44,500 SARS-2-CoV-2 full genomes sampled worldwide available in the GISAID platform (4, 5) collected as of May 8th, 2020 were downloaded and aligned as part of the public data sets used for the COVID-19 Genomics UK (COG-UK) Consortium (51, 52) (Table S2). Thirty-four additional sequences from Mexico also available in the GISAID platform were included, as well as 17 Mexican virus genomes characterized from the earliest introductions of SARS-CoV-2 into the country (38). A sequence subsampling scheme from the original COG data set was done to include all sequences collected globally within the time frame of our study and up to a week after the collection date of our samples (April 30th, 2020). Subsampling was done using a random approach, since we were not looking into mutations occurring within a global context. For this, the global COG sequence alignment (excluding Mexican sequences) was randomly subsampled to 2% and sequences with the following characteristics were removed: (i) >1000-nt shorter than the full genome length, (ii) identical to each other, or (iii) containing >10% ambiguities (N or X). A total of 508 sequences retained in addition to the 57 genomes generated in this study, were used for the final data set.

Phylogenetic placement and tree inference. Phylogenetic placement and genetic characterization of the 57 Mexican virus genome sequences was done using the PANGOLIN tool v3.1.7 lineages version 2021-07-09 (53, 54) (Table S1). For local phylogenomic analysis, the main viral ORFs (Orf1ab and S) were extracted and aligned. Individual ORF alignments were concatenated to generate an Orf1ab+S alignment, excluding UTRs and noncoding intergenic regions. The final data set was composed of 565 sequences with a total length of 24,821 nucleotides. A maximum likelihood (ML) tree was then estimated using RA×ML v8 under a general time reversible nucleotide substitution model with gamma-distributed among-site rate variation (GTR+G) (55). Branch support was assessed using a bootstrap analysis with 100 replicates. The resulting ML tree was rooted using the Wuhan-Hu-1 reference genome. The general clustering pattern observed for the Mexican sequences was verified by comparing it to available global trees (56, 57) and to a tree generated in previous study by our group (38).

Cluster identification and genotype–phenotype association analysis. To search for mutations occurring at a minimum frequency of 0.01 within the sampled sequence population, variable sites across the alignment were identified by comparing across sequences and to a general consensus generated under a 95% threshold using the "Find Variations/SNPs" function in Geneious Prime v2020.0.4 (58). Sites corresponding to nonsynonymous shared mutations across the 57 Mexican sequences were coded as discrete taxon traits and subjected to ancestral state reconstruction on the ML tree under a maximum likelihood framework (RAS-ML) (59). Ancestral reconstructions were performed at a node level, representing changes occurring within nodes and not those occurring on terminal branches likely to correspond to mutations that have not been fixed in the circulating virus population.

Clusters inferred to have originated within Mexico were identified in the ML tree by detecting a minimum of two sister sequences diverging from an "ancestral" Mexican isolate, and a bootstrap support value of >50 for the joining node. Following cluster identification, the corresponding clinical metadata available for the 57 Mexican sequences were used to identify virus clusters associated with different disease outcome categories (A, D, or H groups) (Table S3, Fig. S2). Using a trait-association phylogenomic approach, the ancestral reconstruction trees were visually inspected to identify the co-occurrence of mutations (virus genotypes) within the Mexican sequence clusters, and the phenotype traits of interest (A, D, or H groups). This method allows detection of mutations that co-occur across independent sequence clusters (i.e., lineages) and that may be linked to a given phenotype (in this case disease outcome) (60–62). Any structured genotype-to-genotype association observed within the tree can be formally tested using a phylogenetically informed statistical approach (60–62), instead of using a standard χ^2 test that does not take into consideration possible phylogenetic correlation of the sequence data.

Data availability. The generated sequences of SARS-CoV-2 used in this study have been publicly shared through the Global Initiative on Sharing All Influenza Data (GISAID) repository and have also been deposited in the NCBI database. Accession numbers are listed in Table S3.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 6.2 MB.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3.
2. Anonymous. 2012. Family—Coronaviridae, p 806–828. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), Virus taxonomy. https://doi.org/10.1016/B978-0-12-384684-6.00068-9. Elsevier, San Diego, CA.
3. World Health Organization. 2021. Coronavirus disease (COVID-19) pandemic. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed June 2021.
4. GISAID. EpiCoV. https://www.epicov.org. Accessed June 2021.
5. Shu Y, McCauley J. 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. Euro Surveill 22:30494. https://doi.org/10.2807%2F1560-7917.ES.2017.22.13.30494.
6. Rambaut ALN, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T, Robertson DL, Volz E, on behalf of COVID-19 Genomics Consortium UK (CoG-UK). 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Virologicalorg 563.
7. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao M, Korsman S, Davies M-A, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Pond SLK, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2

(SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv https://doi.org/10.1101/2020.12.21.20248640%.

8. Faria N, Candido D, Moyses Franco LA, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, Camilo CC, Fraiji NA, Esashika Crispim MA, Carvalho MSS, Rambaut A, Loman N, Pybus OG, Sabino CS, on behalf of CADDE Genomic Network. 2021. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virologicalorg 586.

9. O'Toole Á, Hill V, Pybus OG, Watts A, Bogoch II, Khan K, Messina JP; COVID-19 Genomics UK (COG-UK) Consortium; Network for Genomic Surveillance in South Africa (NGS-SA); Brazil-UK CADDE Genomic Network, Tegally H, Lessells RR, Giandhari J, Pillay S, Tumedi KA, Nyepetsi G, Kebabonye M, Matsheka M, Mine M, Tokajian S, Hassan H, Salloum T, Merhi G, Koweyes J, Geoghegan JL, de Ligt J, Ren X, Storey M, Freed NE, Pattabiraman C, Prasad P, Desai AS, Vasanthapuram R, Schulz TF, Steinbrück L, Stadler T, Swiss Viollier Sequencing Consortium, Parisi A, Bianco A, García de Viedma D, Buenestado-Serrano S, Borges V, Isidro J, Duarte S, Gomes JP, Zuckerman NS, Mandelboim M, Mor O, Seemann T, Arnott A, Draper J, Gall M, Rawlinson W, Deveson I, Schlebusch S, McMahon J, Leong L, Lim CK, Chironna M, Loconsole D, Bal A, Josset L, Holmes E, St George K, Lasek-Nesselquist E, Sikkema RS, Oude Munnink B, Koopmans M, Brytting M, Sudha Rani V, Pavani S, Smura T, Heim A, Kurkela S, Umair M, Salman M, Bartolini B, Rueca M, Drosten C, Wolff T, Silander O, Eggink D, Reusken C, Vennema H, Park A, Carrington C, Sahadeo N, Carr M, Gonzalez G, SEARCH Alliance San Diego, National Virus Reference Laboratory, SeqCOVID-Spain, Danish Covid-19 Genome Consortium (DCGC), Communicable Diseases Genomic Network (CDGN), Dutch National SARS-CoV-2 Surveillance Program, Division of Emerging Infectious Diseases (KDCA), de Oliveira T, Faria N, Rambaut A, Kraemer MUG. 2021. Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. Wellcome Open Res 6:121. https://doi.org/10.12688/wellcomeopenres.16661.2.

10. Yoshimoto FK. 2020. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. Protein J 39:198–216. https://doi.org/10.1007/s10930-020-09901-4.

11. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe 29:463–476.e6. https://doi.org/10.1016/j.chom.2021.02.003.

12. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Luchsinger L, Hillyer CD, Caskey M, Robbiani DF, Rice CM, Nussenzweig MC, Hatziioannou T, Bieniasz PD. 2020. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife 9:e61312. https://doi.org/10.7554/eLife.61312.

13. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, Sheffield C-GG, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC, Sheffield COVID-19 Genomics Group. 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182:812–827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

14. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, Zhang X, Muruato AE, Zou J, Fontes-Garfias CR, Mirchandani D, Scharton D, Bilello JP, Ku Z, An Z, Kalveram B, Freiberg AN, Menachery VD, Xie X, Plante KS, Weaver SC, Shi PY. 2021. Spike mutation D614G alters SARS-CoV-2 fitness. Nature 592:116–121. https://doi.org/10.1038/s41586-020-2895-3.

15. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A, Diehl WE, Dauphin A, Carbone C, Veinotte K, Egri SB, Schaffner SF, Lemieux JE, Munro JB, Rafique A, Barve A, Sabeti PC, Kyratsous CA, Dudkina NV, Shen K, Luban J. 2020. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell 183:739–751.e8. https://doi.org/10.1016/j.cell.2020.09.032.

16. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole A, Southgate J, Johnson R, Jackson B, Nascimento FF, Rey SM, Nicholls SM, Colquhoun RM, da Silva Filipe A, Shepherd J, Pascall DJ, Shah R, Jesudason N, Li K, Jarrett P, Pacchiarini N, Bull M, Geidelberg L, Siveroni I, Consortium C-U, Goodfellow I, Loman NJ, Pybus OG, Robertson DL, Thomson EC, Rambaut A, Connor TR, COG-UK Consortium. 2021. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 184:64–75.e11. https://doi.org/10.1016/j.cell.2020.11.020.

17. Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, Walsh RM, Jr, Rits-Volloch S, Zhu H, Woosley AN, Yang W, Sliz P, Chen B. 2021. Structural impact on SARS-CoV-2 spike protein by D614G substitution. Science 372:525–530. https://doi.org/10.1126/science.abf2303.

18. Mansbach RA, Chakraborty S, Nguyen K, Montefiori DC, Korber B, Gnanakaran S. 2021. The SARS-CoV-2 spike variant D614G favors an open conformational state. Sci Adv 7:eabf3671. https://doi.org/10.1126/sciadv.abf3671.

19. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong PD, Graham BS, Mascola JR, Chang JY, Yin MT, Sobieszczyk M, Kyratsous CA, Shapiro L, Sheng Z, Huang Y, Ho DD. 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593:130–135. https://doi.org/10.1038/s41586-021-03398-2.

20. World Health Organization. 2021. COVID-19 weekly epidemiological update - 4 May 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—4-may-2021.

21. Yang TJ, Yu PY, Chang YC, Liang KH, Tso HC, Ho MR, Chen WY, Lin HT, Wu HC, Hsu SD. 2021. Effect of SARS-CoV-2 B.1.1.7 mutations on spike protein structure and function. Nat Struct Mol Biol 28:731–739. https://doi.org/10.1038/s41594-021-00652-z.

22. Zhu X, Mannar D, Srivastava SS, Berezuk AM, Demers JP, Saville JW, Leopold K, Li W, Dimitrov DS, Tuttle KS, Zhou S, Chittori S, Subramaniam S. 2021. Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. PLoS Biol 19:e3001237. https://doi.org/10.1371/journal.pbio.3001237.

23. Barton MI, MacGowan SA, Kutuzov MA, Dushek O, Barton GJ, van der Merwe PA. 2021. Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. Elife 10:e70658. https://doi.org/10.7554/eLife.70658.

24. Chen RE, Zhang X, Case JB, Winkler ES, Liu Y, VanBlargan LA, Liu J, Errico JM, Xie X, Suryadevara N, Gilchuk P, Zost SJ, Tahan S, Droit L, Turner JS, Kim W, Schmitz AJ, Thapa M, Wang D, Boon ACM, Presti RM, O'Halloran JA, Kim AHJ, Deepak P, Pinto D, Fremont DH, Crowe JE, Jr, Corti D, Virgin HW, Ellebedy AH, Shi PY, Diamond MS. 2021. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. Nat Med 27:717–726. https://doi.org/10.1038/s41591-021-01294-w.

25. WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection. 2020. A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect Dis 20:e192–e197. https://doi.org/10.1016/S1473-3099(20)30483-7.

26. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, McDonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong AYS, Grieve R, Harrison D, Forbes H, Schultze A, Croker R, Parry J, Hester F, Harper S, Perera R, Evans SJW, Smeeth L, Goldacre B. 2020. Factors associated with COVID-19-related death using OpenSAFELY. Nature 584:430–436. https://doi.org/10.1038/s41586-020-2521-4.

27. Liu K, Chen Y, Lin R, Han K. 2020. Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. J Infect 80:e14–e18. https://doi.org/10.1016/j.jinf.2020.03.005.

28. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, Jia X, Wu M, Shi B, Xu S, Chen J, Wang W, Chen B, Jiang L, Yu S, Lu J, Wang J, Xu M, Yuan Z, Zhang Q, Zhang X, Zhao G, Wang S, Chen S, Lu H. 2020. Viral and host factors related to the clinical outcome of COVID-19. Nature 583:437–440. https://doi.org/10.1038/s41586-020-2355-0.

29. Kalligeros M, Shehadeh F, Mylona EK, Benitez G, Beckwith CG, Chan PA, Mylonakis E. 2020. Association of obesity with disease severity among patients with coronavirus disease 2019. Obesity (Silver Spring) 28:1200–1204. https://doi.org/10.1002/oby.22859.

30. Bello-Chavolla OY, Gonzalez-Diaz A, Antonio-Villa NE, Fermin-Martinez CA, Marquez-Salinas A, Vargas-Vazquez A, Bahena-Lopez JP, Garcia-Pena C, Aguilar-Salinas CA, Gutierrez-Robledo LM. 2020. Unequal impact of structural health determinants and comorbidity on COVID-19 severity and lethality in older Mexican adults: Considerations beyond chronological aging. J Gerontol A Biol Sci Med Sci 76:e52–e59. https://doi.org/10.1093/gerona/glaa163.

31. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. 2020. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat Commun 11:5986. https://doi.org/10.1038/s41467-020-19818-2.

32. Grubaugh ND, Hanage WP, Rasmussen AL. 2020. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. Cell 182:794–795. https://doi.org/10.1016/j.cell.2020.06.040.

33. Secretaría de c. 2021. Datos Abiertos Dirección General de Epidemiología. https://www.gob.mx/salud/documentos/datos-abiertos-152127. Accessed 2020.

34. Nehme M, Braillard O, Alcoba G, Aebischer Perone S, Courvoisier D, Chappuis F, Guessous I. 2021. COVID-19 symptoms: longitudinal evolution and persistence in outpatient settings. Ann Intern Med 174:723–725. https://doi.org/10.7326/M20-5926.

35. Tenforde MW, Kim SS, Lindsell CJ, Billig Rose E, Shapiro NI, Files DC, Gibbs KW, Erickson HL, Steingrub JS, Smithline HA, Gong MN, Aboodi MS, Exline MC, Henning DJ, Wilson JG, Khan A, Qadir N, Brown SM, Peltan ID, Rice TW, Hager DN, Ginde AA, Stubblefield WB, Patel MM, Self WH, Feldstein LR, IVY Network Investigators, CDC COVID-19 Response Team. 2020. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network—United States, March–June 2020. MMWR Morb Mortal Wkly Rep 69:993–998. https://doi.org/10.15585/mmwr.mm6930e1.

36. Rausch JW, Capoferri AA, Katusiime MG, Patro SC, Kearney MF. 2020. Low genetic diversity may be an Achilles heel of SARS-CoV-2. Proc Natl Acad Sci U S A 117:24614–24616. https://doi.org/10.1073/pnas.2017726117.

37. MacLean OA, Lytras S, Weaver S, Singer JB, Boni MF, Lemey P, Kosakovsky Pond SL, Robertson DL. 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol 19:e3001115. https://doi.org/10.1371/journal.pbio.3001115.

38. Taboada B, Vazquez-Perez JA, Munoz-Medina JE, Ramos-Cervantes P, Escalera-Zamudio M, Boukadida C, Sanchez-Flores A, Isa P, Mendieta-Condado E, Martinez-Orozco JA, Becerril-Vargas E, Salas-Hernandez J, Grande R, Gonzalez-Torres C, Gaytan-Cervantes FJ, Vazquez G, Pulido F, Araiza-Rodriguez A, Garces-Ayala F, Gonzalez-Bonilla CR, Grajales-Muniz C, Borja-Aburto VH, Barrera-Badillo G, Lopez S, Hernandez-Rivas L, Perez-Padilla R, Lopez-Martinez I, Avila-Rios S, Ruiz-Palacios G, Ramirez-Gonzalez JE, Arias CF. 2020. Genomic analysis of early SARS-CoV-2 variants introduced in Mexico. J Virol 94:e01056-20.

39. Alm E, Broberg EK, Connor T, Hodcroft EB, Komissarov AB, Maurer-Stroh S, Melidou A, Neher RA, O'Toole A, Pereyaslov D, The WHO European Region Sequencing Laboratories and GISAID EpiCoV Group. 2020. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill 25:2001410. https://doi.org/10.2807/1560-7917.es.2020.25.32.2001410.

40. O'Toole A, Hill V, McCrone J, Scher E, Rambaut A. PANGO lineages: lineage B. https://cov-lineages.org/lineages/lineage_B.html. Accessed July 2021.

41. Page RDM, Holmes EC. 2009. Molecular evolution: a phylogenetic approach. Wiley, New York, NY.

42. Everett J, Hokama P, Roche AM, Reddy S, Hwang Y, Kessler L, Glascock A, Li Y, Whelan JN, Weiss SR, Sherrill-Mix S, McCormick K, Whiteside SA, Graham-Wooten J, Khatib LA, Fitzgerald AS, Collman RG, Bushman F. 2021. SARS-CoV-2 genomic variation in space and time in hospitalized patients in Philadelphia. mBio 12:e03456-20.

43. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 395:1054–1062. https://doi.org/10.1016/S0140-6736(20)30566-3.

44. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, Wang B, Wu Y, Meynert A, Griffiths F, Oosthuyzen W, Kousathanas A, Moutsianas L, Yang Z, Zhai R, Zheng C, Grimes G, Beale R, Millar J, Shih B, Keating S, Zechner M, Haley C, Porteous DJ, Hayward C, Yang J, Knight J, Summers C, Shankar-Hari M, Klenerman P, Turtle L, Ho A, Moore SC, Hinds C, Horby P, Nichol A, Maslove D, Ling L, McAuley D, Montgomery H, Walsh T, et al. 2021. Genetic mechanisms of critical illness in Covid-19. Nature 591:92–98. https://doi.org/10.1038/s41586-020-03065-y.

45. Severe Covid GG, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernandez J, Prati D, Baselli G, Asselta R, Grimsrud MM, Milani C, Aziz F, Kassens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez AE, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani

AL, Peschuck A, Julia A, Pesenti A, Voza A, Jimenez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, et al. 2020. Genome-wide association study of severe Covid-19 with respiratory failure. N Engl J Med 383:1522–1534.

46. Zietz M, Zucker J, Tatonetti NP. 2020. Associations between blood type and COVID-19 infection, intubation, and death. Nat Commun 11:5761. https://doi.org/10.1038/s41467-020-19623-x.

47. Li SL, Pereira RHM, Prete CA, Zarebski AE, Emanuel L, Alves PJ, Peixoto PS, Braga CK, de Souza Santos AA, de Souza WM, Barbosa RJ, Buss LF, Mendrone A, de Almeida-Neto C, Ferreira SC, Salles NA, Marcilio I, Wu C-H, Gouveia N, Nascimento VH, Sabino EC, Faria NR, Messina JP. 2020. Social and racial inequalities in COVID-19 risk of hospitalisation and death across São Paulo state, Brazil. medRxiv. https://doi.org/10.1101/2020.12.09.20246207%.

48. Secretaría de Salud. 2021. Lineamiento Estandarizado para la Vigilancia Epidemiológica y por Laboratorio de la enfermedad respiratoria viral. https://www.gob.mx/salud/documentos/lineamiento-estandarizado-para-la-vigilancia-epidemiologica-y-por-laboratorio-de-la-enfermedad-respiratoria-viral.

49. World Health Organization. 2020. Diagnostic testing for SARS-CoV-2. https://www.who.int/publications/i/item/diagnostic-testing-for-sars-cov-2.

50. Quick J. 2020. nCoV-2019 sequencing protocol v3 (LoCost) V.3. https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye?step=7. Accessed 2020.

51. Covid-Genomics UK Consortium. 2020. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe 1:e99–e100. https://doi.org/10.1016/S2666-5247(20)30054-9.

52. COG-UK Consortium. 2020. COVID-19 genomics UK consortium. https://www.cogconsortium.uk/.

53. O'Toole A, Hill V, McCrone J, Scher E, Rambaut A. Pangolin COVID-19 Lineage Assigner. https://github.com/cov-lineages/pangolin. Accessed February 2021.

54. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, Du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5:1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

55. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

56. Aksamentov I, Neher R. Nextclade. https://clades.nextstrain.org/. Accessed February 2021.

57. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34:4121–4123. https://doi.org/10.1093/bioinformatics/bty407.

58. Biomatters. Geneious Prime v2020.0.4. https://www.geneious.com Accessed March 2021.

59. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528. https://doi.org/10.1093/bioinformatics/bty633.

60. Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, Carlson J, Yusim K, McMahon B, Gaschen B, Mallal S, Mullins JI, Nickle DC, Herbeck J, Rousseau C, Learn GH, Miura T, Brander C, Walker B, Korber B. 2007. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. Science 315:1583–1586. https://doi.org/10.1126/science.1131528.

61. Escalera-Zamudio M, Golden M, Gutierrez B, Theze J, Keown JR, Carrique L, Bowden TA, Pybus OG. 2020. Parallel evolution in the emergence of highly pathogenic avian influenza A viruses. Nat Commun 11:5511. https://doi.org/10.1038/s41467-020-19364-x.

62. Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceeding of the Royal Society Biological Sciences. https://doi.org/10.1098/rspb.1994.0006.