International Journal of Infectious Diseases 116 (2022) 91–100

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Genomic characterization of SARS-CoV-2 and its association with clinical outcomes: a 1-year longitudinal study of the pandemic in Colombia

Ángela María Ruiz-Sternberg[a,*], Henry Mauricio Chaparro-Solano[a,b,c], Ludwig L. Albornóz[d,e], Ángela María Pinzón-Rondón[a], Juan Mauricio Pardo-Oviedo[a,c], Nicolás Molano-González[a], Diego Andrés Otero-Rodríguez[b], Fabio Andrés Zapata-Gómez[b], Jubby Marcela Gálvez[b]

[a] Clinical Investigation Group, Universidad del Rosario, Bogotá, Colombia
[b] Genuino Research Group, Gencell Pharma, Bogotá, Colombia
[c] Hospital Universitario Mayor – Méderi, Bogotá, Colombia
[d] Departamento de Patología y Medicina de Laboratorio, Fundación Valle del Lili, Cali, Colombia
[e] Facultad de Ciencias de la Salud, Universidad ICESI, Cali, Colombia

ARTICLE INFO

Article history:
Received 23 August 2021
Revised 10 November 2021
Accepted 8 December 2021

Keywords:
SARS-CoV-2
SARS-CoV-2 variants
COVID-19
Mortality
Hospitalization
High-throughput nucleotide sequencing

ABSTRACT

Objectives: This study aimed to explore associations between the molecular characterization of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and disease severity in ambulatory and hospitalized patients in two main Colombian epicentres during the first year of the coronavirus disease 2019 pandemic.

Methods: In total, 1000 patients with SARS-CoV-2 infection were included in this study. Clinical data were collected from 997 patients, and 678 whole-genome sequences were obtained by massively parallel sequencing. Bivariate, multi-variate, and classification and regression tree analyses were run between clinical and genomic variables.

Results: Age >88 years, and infection with lineages B.1.1, B.1.1.388, B.1.523 or B.1.621 for patients aged 71–88 years were associated with death [odds ratio (OR) 6.048036, 95% confidence interval (CI) 1.346567–32.92521; *P*=0.01718674]. The need for hospitalization was associated with higher age and comorbidities. The hospitalization rate increased significantly for patients aged 38–51 years infected with lineages A, B, B.1.1.388, B.1.1.434, B.1.153, B.1.36.10, B.1.411, B.1.471, B.1.558 or B.1.621 (OR 8.368427, 95% CI 2.573145–39.10672, *P*=0.00012). Associations between clades and clinical outcomes diverged from previously reported data.

Conclusions: Infection with lineage B.1.621 increased the hospitalization and mortality rates. These findings, plus the rapidly increasing prevalence in Colombia and other countries, suggest that lineage B.1.621 should be considered as a 'variant of interest'. If associated disease severity is confirmed, possible designation as a 'variant of concern' should be considered.

Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), an RNA virus from the coronavirus family with a genome of 29.8 kb, has emerged as a new viral pathogen that causes coronavirus disease 2019 (COVID-19). Due to its important transmission capabilities, this virus spread rapidly around the world, and the World Health Organization (WHO) declared a pandemic situation in March 2020. By 29 July 2021, there had been >196 million cases and approximately 4 million reported deaths worldwide (Johns Hopkins Coronavirus Resource Center, 2021). In Colombia (estimated population in 2020 of 50.3 million), there had been 4,877,323 cases and 123,781 reported deaths at the time of submission of this article. By August 2021, Colombia had faced three COVID-19 waves: July–August 2020, January 2021, and April - August 2021. The latter was the most aggressive,

* Corresponding author. Clinical Investigation Group, Universidad del Rosario, Carrera 24 # 63C – 69, Bogotá, Colombia. Tel.: +57 1 2970200.
E-mail address: angela.ruiz@urosario.edu.co (Á.M. Ruiz-Sternberg).

https://doi.org/10.1016/j.ijid.2021.12.326

representing by that time the second-highest worldwide number of daily new cases and deaths since May 2021 (Instituto Nacional de Salud, 2021).

SARS-CoV-2 infection can have a wide spectrum of clinical outcomes, from asymptomatic infection to severe disease and death. Although the sociodemographic and clinical risk factors related to COVID-19 clinical presentation are well known, the influence of the viral mutational profile on infectivity and disease severity is yet to be fully elucidated (Richardson et al., 2020; Huang and Wang, 2021). All viruses undergo genomic changes as they spread, but such variations do not generally imply a structural or functional impact on protein translation (Peacock et al., 2021).

Since publication of the complete genomic sequence in December 2019, the SARS-CoV-2 genome has been thoroughly characterized, leading to the description of genes and regions that are important for host recognition and cellular entry, as well as evasion of the immune response. Different nomenclature systems based on the identification of mutation markers, such as the Global Initiative on Sharing All Influenza Data (GISAID) which defines eight major clades (S, L, V, G, GH, GR, GV and GRY), and/or genetic, epidemiological and geographical characteristics, such as the Phylogenetic Assignment Named Global Outbreak (PANGO) lineages, have been proposed (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017; Rambaut et al., 2020). These systems are useful for tracking pandemic viral spread, allowing investigation of possible associations between novel genetic variants, lineages or clades, and disease severity, despite the fact that these genetic variations may not, in themselves, explain viral phenotypic characteristics.

As a result of implementation of the PANGO system and establishment of genomic surveillance programmes in countries around the world, an increasing number of lineages and variants have been described. Despite important efforts and investments made for continuous sequencing of the SARS-CoV-2 genome, an insignificant proportion of variants have been recognized as epidemiologically or clinically relevant. These variants, called 'variants of concern' (VOCs), 'variants of interest' (VOIs), and 'variants of high consequence' (VOHs), demand greater interest from governments and public health agencies as they contain changes that modify viral transmissibility, disease severity, and response to therapeutic and diagnostic tools (Janik et al., 2021). It has been hypothesized that these variants are the result of selective pressure due to changes in host immune characteristics, as well as the development of new drugs, immunotherapy and vaccines. The first VOC was the Alpha variant (B.1.1.7 lineage), identified in September 2020 in England (Galloway et al., 2021). Since then, other VOCs and VOIs have been reported, some of which have displayed convergent mutations that confer the functional adaptive characteristics of the virus. These variants have become predominant, and exhibit higher transmissibility and/or a significant impact on immunity and disease severity (World Health Organization, 2021). Few longitudinal studies have been undertaken to determine possible associations between the mentioned lineages, clades or other classification systems, and clinical outcomes (Hamed et al., 2021; Lamptey et al., 2021; Nakamichi et al., 2021; Young et al., 2021).

This study aimed to explore associations between the molecular characterization of SARS-CoV-2 and disease severity in ambulatory and hospitalized patients from two main cities in Colombia during the first year of the pandemic.

Methods

Biological specimen collection, nucleic acid inactivation, and extraction

Informed consent was obtained from eligible patients. Prospective specimens were confirmed SARS-CoV-2 respiratory tract specimens [nasopharyngeal aspirate or swab; positive result on quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) assay] collected from patients recruited from tertiary care university hospitals and a molecular diagnostics laboratory from two main pandemic epicentres in Colombia. Retrospective specimens were RNA eluate or primary nasopharyngeal swabs/aspirates obtained from biorepositories at the participating research centre. RT-qPCR-negative samples were excluded. Demographic and clinical characteristics were collected in CASPIO (Caspio, Inc., Sunnyvale, CA, USA).

Viral RNA inactivation and extraction were performed on 0.2-mL aliquots of viral transport medium (primary sample swab specimens), or on 1-mL aliquots in sterile isotonic saline solution (primary aspirate samples). All specimens were heat-inactivated (56 °C for 30 min) and manipulated under biosafety level 2 conditions.

RNA extraction (cell lysis, bead binding to magnetic rods, RNA binding to beads, washing and elution) was performed to obtain 0.1 mL of the RNA eluate. Automated RNA extraction methods comprised ExiPrep 96 Viral DNA/RNA Kit on ExiPrep 96 Lite instrument (Bioneer Corp., Daejeon, Republic of Korea), MGIEasy Nucleic Acid Extraction Kit on MGISP-960 (MGI Tech Co. Ltd, Shenzhen, People's Republic of China), NucliSENS Nucleic Acid Extraction Reagents on NucliSENS easyMAG (bioMérieux SA, Marcy l'Etoile, France), or MagNA Pure Compact Nucleic Acid Isolation Kit I on MagNA Pure Compact (Roche Diagnostics GmbH, Mannheim, Germany).

Statistical methods

Qualitative variables were reported as frequencies and percentages. Quantitative variables were reported as mean and standard deviation or median and interquartile range, depending on normality distribution.

The Kruskall–Wallis test and Chi-squared test were used to assess associations between viral genome characteristics (presence or absence of genetic variants, total number of variants per sample, total number of variants discriminated by gene and impact of the variant in the protein, PANGOLIN lineage, and GISAID clade) and death and need for hospitalization, respectively. In a second approach, the Classification and Regression Trees (Breiman et al., 2017) algorithm was used to find the relevant variables associated with death and hospitalization. This algorithm is useful as the number of covariates that can be included in the model is unlimited, unlike more traditional approaches such as logistic regression. Sex, age, number of comorbidities, asymptomatic status, body mass index, and the aforementioned genetic characteristics were included as covariates.

The overall significance level was set at 5%. R Version 4.0.2 was used for all statistical analyses.

Phylogenetic analysis

All SARS-CoV-2 genomes were downloaded from SOPHiA DDM bioinformatics software (SOPHiA Genetics Inc., Saint Sulpice, Switzerland). Fasta files were aligned to the reference genome, NC_045512, using MAFFT v7 software (Katoh et al., 2002). Next, a nucleotide substitution model was predicted using jModelTest v2.1.10 (Posada, 2008). Later, a maximum likelihood tree was constructed with IQ-TREE 2 software (Minh et al., 2020) using the GTR + Γ model and 1000 bootstrap replicates. Finally, each genome had a lineage assigned using the PANGOLIN webserver (Rambaut et al., 2020). Additionally, the CoVsurver online server (CoVsurver, 2021) was used for assignment of GISAID clades.

This study was approved by the institutional review boards of Universidad del Rosario and participating hospital research centres.

Table 1
Baseline demographic characteristics of the cohort (n=997).

	Category	n (%)
Age (years)	<40	349 (35%)
	40–60	312 (31.2%)
	>60	336 (33.7%)
Sex	Female	485 (48.6%)
	Male	512 (51.3%)
Ethnicity	Ethnic majority (Mestizo/White)	764 (76.6%)
	Afro	22 (2.8%)
	Indigenous	5 (0.5%)
	Mulato	12 (1.4%)
	Other	46 (4.6%)
	NA	148 (17.4%)
Nationality	Colombian	987 (98.9%)
	Venezuelan	7 (0.7%)
	Other	3 (0.4%)
Residence	Bogotá	621 (62.2%)
	Cali	289 (29%)
	Cundinamarca	68 (6.9%)
	Other	19 (1.9%)
Age (years)	Mean (SD)	50.6 (18.4)

SD, standard deviation.

Table 2
Baseline clinical characteristics of the cohort (n=997).

Variable	Category	
Clinical outcomes, n (%)	Ambulatory	508 (50.9%)
	In-hospital	328 (32.8%)
	Intensive care	92 (9.2%)
	Deceased	69 (6.9%)
Symptomatology, n (%)	Asymptomatic	93 (9.3%)
	Symptomatic	904 (90.7%)
Complications, n (%)	Cardiac	48 (4.8%)
	Respiratory	293 (29.4%)
	Renal	79 (7.9%)
	Neurological	36 (3.6%)
	Thromboembolic	28 (2.8%)
Symptoms, n (%)	Fever	469 (47.1%)
	Cough	544 (54.5%)
	Fatigue	523 (52.4%)
	Dyspnea	415 (41.6%)
	Diarrhoea	173 (17.3%)
	Sore throat	235 (23.6)
	Anosmia	271 (27.2%)
	Dysgeusia	232 (23.3%)
	Chest pain	110 (11%)
	Nasal congestion	177 (17.7)
	Abdominal pain	62 (6.2%)
	Cyanosis	7 (0.7%)
	Headache	324 (32.4)
Body mass index (kg/m^2)	Mean (SD)	26.4 (4.4)

SD, standard deviation.

All international and national bioethical principles and regulations for clinical investigation in human subjects were followed.

Results

Characteristics of the study population

Demographic characteristics

Clinical and demographic information was obtained from 997 patients. The mean age was 50.6 years, 35% of participants were aged <40 years, and 33.7% of participants were aged >60 years. The sex distribution in the cohort was homogeneous, and 76.6% were Mestizo/White (ethnic majority). Patients resided in Bogotá (62.2%) and Cali (29%) (Table 1).

Clinical characteristics

At diagnosis, 90.7% of the patients had symptoms. Overall, 50.8% were outpatients, 33% were hospitalized, 9.2% received intensive care unit (ICU) support, and 6.9% died. The most common complications were respiratory (29.4%); symptoms were cough (54.5%), fatigue (52.4%) and fever (47.1%) (Table 2).

Table 3
Phylogenetic Assignment Named Global Outbreak (PANGO) lineages detected and their frequencies.

PANGO lineage	WHO label for VOCs/VOIs	n	%
B.1	-	305	45.0
B.1.111	-	77	11.4
B.1.1.348	-	66	9.7
B.1.153	-	39	5.8
B.1.1	-	39	5.8
B.1.420	-	31	4.6
A	-	16	2.4
B	-	11	1.6
B.1.383[a]	-	9	1.3
B.1.523	-	8	1.2
B.1.621	Mu	7	1.0
B.1.1.388	-	6	0.9
B.1.36.10	-	5	0.7
B.1.293[a]	-	5	0.7
B.1.1.413[a]	-	5	0.7
B.1.1.291[a]	-	5	0.7
B.1.1.28	-	4	0.6
B.1.177.86[a]	-	2	0.3
B.1.1.1	-	2	0.3
B.1.416	-	2	0.3
B.1.1.434[b]	-	2	0.3
B.1.411[a]	-	2	0.3
B.1.36.31[a]	-	2	0.3
B.1.1.213[a]	-	2	0.3
B.1.389[a]	-	1	0.1
B.1.165	-	1	0.1
B.1.1.100	-	1	0.1
B.1.319[b]	-	1	0.1
B.59[b]	-	1	0.1
B.1.456[a]	-	1	0.1
B.1.485[a]	-	1	0.1
B.1.1.409[b]	-	1	0.1
B.1.1.37[a]	-	1	0.1
B.1.505[a]	-	1	0.1
B.1.606[a]	-	1	0.1
B.1.529[a]	-	1	0.1
B.1.565[b]	-	1	0.1
B.1.1.272[a]	-	1	0.1
A.2	-	1	0.1
B.1.281[a]	-	1	0.1
B.1.533[a]	-	1	0.1
B.1.471[a]	-	1	0.1
B.1.324[a]	-	1	0.1
A.2.4[a]	-	1	0.1
B.1.1.7[c]	Alpha	1	0.1
B.1.575[b]	-	1	0.1
B.1.558[a]	-	1	0.1
P.1[c]	Gamma	1	0.1
B.1.1.371[a]	-	1	0.1
B.1.404[a]	-	1	0.1

WHO, World Health Organization; VOC, variant of concern; VOI, variant of interest.
[a] New in South America.
[b] New in Colombia.
[c] VOC.

SARS-CoV-2 genetic characterization

Seven hundred and sixty-three SARS-CoV-2 sequences were obtained; genomic coverage was >95% in most samples (63.4%). In 10.2% of cases, coverage was 75–95%; 25–75% and <25% coverage occurred in 13.4% and 13% of sequences, respectively.

In total, 2715 single variants were identified: mis-sense variants (54.1%), synonymous variants (37.75%) and loss of function variants (3.6%). The remaining 4.7% of variants included those in the

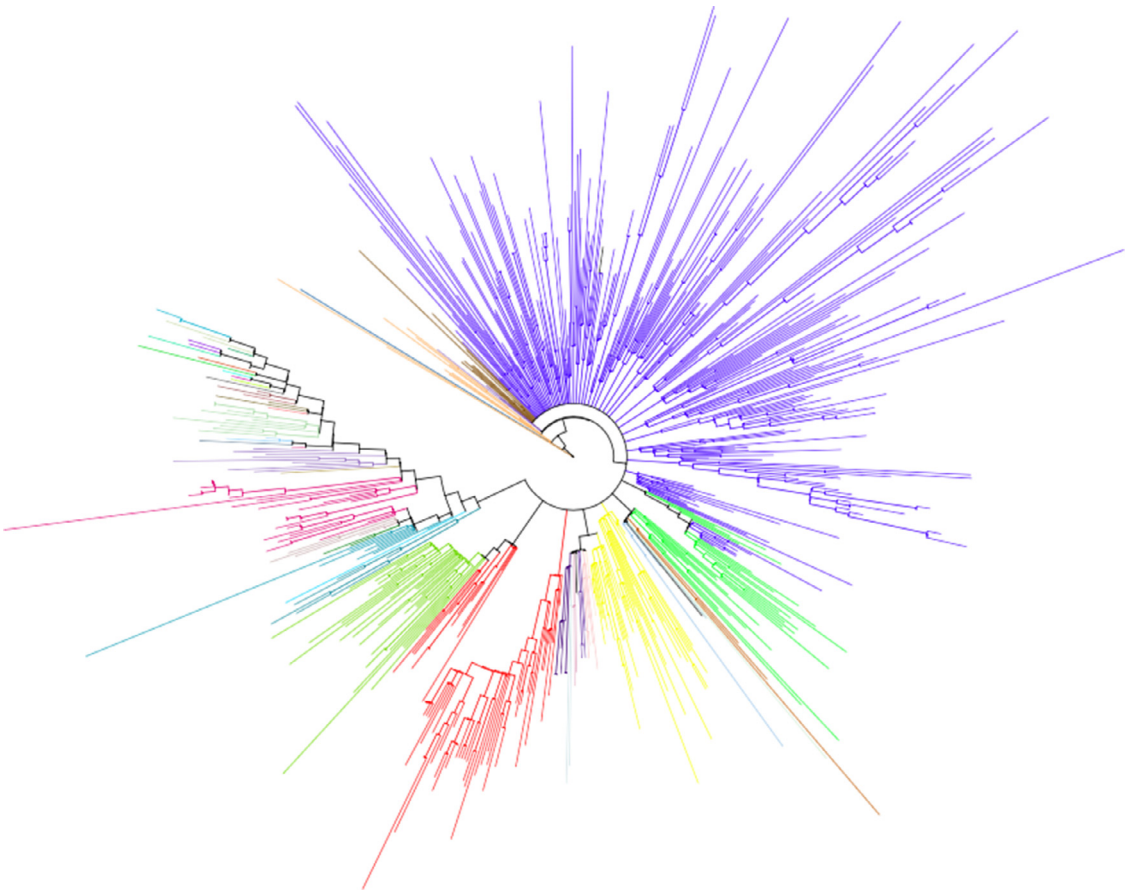

Figure 1. Phylogenetic tree of 673 Colombian samples. Most sequences clearly cluster together while the remaining 15 sequences lie in different branches more closely related to the Wuhan strain. The most common lineages are labelled with the following colours: B.1, blue; B.1.111, red; B.1.1.348, yellow; B.1.1, green; B.1.153, olive green; B.1.420, pink. Other colours are described in HTML code in Table S1 (see online supplementary material).

untranslated (3'UTR and 5'UTR) and intergenic regions, as well as in-frame, InDels, loss-of-start and loss-of-stop codon variants.

Most genetic changes occurred in *ORF1ab* (63.5%), *S* (13.3%) and *N* (6.2%). When adjusted for kb, the highest rates were found in *ORF8* (357.51 variants/kb), 3'UTR (218.34 variants/kb) and *N* (183.92 variants/kb); and the lowest rates were found in *ORF1ab* (80.97 variants/kb), S (94.19 variants/kb) and M (101.64/kb).

Phylogenetic analysis

Due to poor genomic coverage, 85 sequences were discarded from phylogenetic analysis. In addition, five samples did not pass the Chi-square test performed by IQ-TREE and were thus excluded. The maximum likelihood tree constructed yielded one major group (658 samples). The remaining 15 samples were clustered into seven minor branches, more closely related to the original Wuhan strain (Figure 1).

In total, 50 PANGO lineages were identified. Lineage B.1 was most prevalent (45.0%), followed by lineages B.1.111 (11.4%) and B.1.1.348 (9.7%). Interestingly, lineage B.1.621, the so-called 'Colombian variant', was found in seven cases (1.0%) (Table 3). Concerning GISAID clades, clade GH was predominant (48.1%); clades G, GR and 'other' clades were found in 24.8%, 19.8% and 7.1% of cases, respectively. Clades S and GRY were each detected in a single sample.

Demographic and clinical factors associated with mortality

A higher association with death was found in male patients aged >60 years with several comorbidities. Multi-organic complications were associated with higher risk of death. A distinct relationship was identified between mortality and educational level: the mortality rate was higher in patients with a low educational level (Table 4).

Hospitalization risk increased progressively as a function of age. Male sex, lower educational level and most comorbidities were associated with greater need for hospitalization. Interestingly, a history of current or previous smoking was inversely related to hospitalization rate (Table 5).

Sociodemographic and clinical factors associated with GISAID genetic clades

Clades G and GR predominated in residents of Bogotá, and clade GH predominated in residents of Cali. Associations between clades and comorbidities were identified: clade GH and 'other' clades were associated with type 2 diabetes mellitus, malignancy and obesity, and clade G was associated with nephropathy. Symptoms such as cough and fatigue were seen more frequently in patients infected with clade G, while nasal congestion was more common in patients with clade GR0, and respiratory and renal complications were more common in patients with 'other' clades.

The need for hospitalization and ICU care were associated with clade G and 'other' clades, while clades GH and GR predominated in outpatients. Sequences classified as 'other' were more common in patients aged >59 years (Table 6).

GISAID clade composition changed throughout the study time window (Figure 2). Clade G and 'other' clades were more abundant during the first months. The prevalence of clade GR grew from 4%

Table 4
Baseline demographics and medical conditions of patients, stratified by mortality.

	Category	Deceased (*n*=69)	Living (*n*=926)	*P*-value[a]
Age (years), *n* (%)	<40	3 (4.3%)	346 (37.3%)	3.46E-18
	40–60	9 (13%)	303 (32.7%)	
	>60	57 (82.6%)	279 (30%)	
Sex, *n* (%)	Female	23 (33 3%)	462 (49.7%)	0.011972952
	Male	46 (66.6%)	466 (50.2%)	
Highest educational level, *n* (%)	Secondary school	44(63.7)	391 (42.1%)	0.000750963
	Professional	25(36.3%)	537 (58%)	
Comorbidities, *n* (%)	Hypertension	38 (55%)	226 (24.3%)	5.38366E-08
	Type 2 diabetes mellitus	19 (27.5%)	105 (11.2%)	0.984705638
	Asthma	1 (1.4%)	21 (2.2%)	0.984705638
	COPD	12 (17.3%)	27 (2.9%)	1.47512E-08
	Cardiovascular disease	12 (17.3%)	62 (66.8%)	2.39E-03
	Nephropathy	13 (18.8%)	29 (3.1%)	2.53539E-09
	Malignancy	8 (11.5%)	33 (3.5%)	0.003391579
	Autoimmune disease	3 (4.3%)	24 (2.6%)	0.627410658
	Smoking (former)	12 (17.3%)	150 (16.1%)	0.1949283
	Smoking (current)	0	42 (45.%)	
	Obesity	14 (20.2%)	137 (14.7%)	0.288462289
	HIV	1 (1.4%)	7 (0.7%)	1
	Thyroid disease	11 (15.9%)	82 (8.8%)	0.08123423
Symptoms, *n* (%)	Fever	34 (49.2%)	435 (46.9%)	0.794550071
	Cough	47 (68.1%)	497 (53.5%)	0.026544714
	Fatigue	38 (55%)	485 (52.2%)	0.744480128
	Dyspnoea	45 (65.2%)	370 (39.8%)	6.49034E-05
	Diarrhoea	10 (14.4%)	163 (17.6%)	0.622079854
	Sore throat	10 (14.4%)	225 (24.2%)	0.090170707
	Anosmia	5 (7.2%)	266 (28.6%)	0.000201008
	Dysgeusia	4 (5.7%)	228 (24.6%)	0.000643501
	Chest pain	2 (2.8%)	108 (11.6%)	0.041717244
	Nasal congestion	8 (11.5%)	92 (9.9%)	0.809861204
	Abdominal pain	7 (10.1%)	55 (5.9%)	0.253675689
	Cyanosis	1 (1.4%)	6 (0.6%)	0.981464039
	Headache	8 (11.5%)	316 (34%)	0.000207739
	Conjunctivitis	0	25 (2.6%)	0.326209045
	Chills	13 (18.8%)	241 (26%)	0.242789718
Complications, *n* (%)	Cardiac	19 (27.5%)	29 (3.1%)	9.45464E-19
	Respiratory	59 (85.5%)	234 (25.2%)	1.30E-25
	Renal	30 (43.4%)	49 (5.2%)	1.32638E-28
	Neurological	15 (21.7%)	21 (2.2%)	1.04509E-15
	Thromboembolic	7 (10.1%)	21 (2.2%)	0.000575997
Body mass index (kg/m^2) (mean, range)		26.127 (16.2–35)	26.437 (16.3–52)	0.588035743

COPD, chronic obstructive pulmonary disease; HIV, human immunodeficiency virus.
[a] Chi-squared test.

initially to 26% at the end of the study. Similarly, the prevalence of clade GH increased from 21% to 64%.

Regarding the results of the decision tree, Figure 3 shows the most important variables determining patient mortality. Of all variables considered, only age (>88 years), and infection with viral lineages B.1.1, B.1.1.388, B.1.523 or B.1.621 in patients aged 71–88 years were associated with death [odds ratio (OR) 6.048036, 95% confidence interval (CI) 1.346567–32.92521; *P*=0.01718674].

Similarly, for hospitalization, the tree showed that patients aged ≥59 years, and patients aged 51–59 years with comorbidities had higher hospitalization rates. It is noteworthy that in patients aged 38–51 years, infection with certain viral lineages was associated with higher hospitalization rates, namely lineages A, B, B.1.1.388, B.1.1.434, B.1.153, B.1.36.10, B.1.411, B.1.471, B.1.558 and B.1.621 (OR 8.368427, 95% CI 2.573145–39.10672; *P*=0.00012) (Figure 4).

Discussion

Despite the high number of genetic variants and lineages identified among the obtained sequences, and consistent with previously published evidence (Peacock et al., 2021), only two lineages were associated with both increased hospitalization and mortality rates: lineages B.1.621 and B.1.1.388.

Lineage B.1.621, recently labelled as the 'Colombian variant', was detected in Colombia on 11 January 2021 by the National Institute of Health (Laiton-Donato et al., 2021). Interestingly, the present study reports the detection of this lineage in a sample from September 2020 collected in Bogotá. Lineage B.1.621 has disseminated nationally with significant acceleration since March 2021, reaching accumulated prevalence of 26%, and peaking with prevalence of 71% (7-day rolling average) by the end of July 2021, according to data uploaded to GISAID by the Colombian National Genomic Surveillance Program, as per the online data aggregating and lineage/mutation tracker outbreak.info (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017). Lineage B.1.621 was estimated to represent 52.7% of circulating lineages between April and June 2021 (http://www.ins.gov.co/BibliotecaDigital/Estrategia-de-caracterizacion-genomica-SARS-CoV2_Colombia.pdf). Lineage B.1.621 has been detected in 28 countries, prompting active monitoring by WHO since 26 May 2021 (World Health Organization, 2021); Public Health England upgraded it from active monitoring to 'variant under investigation' (VUI-21JUL-01) on 21 July 2021 (Public Health England, 2021); the European Centre for Disease Prevention and Control (ECDC) classified it as a VOI on 29 July 2021 (European Centre for Disease Prevention and Control, 2021). At the date of submission of this article, the US

Table 5
Baseline demographics and medical conditions of hospitalized and non-hospitalized patients.

	Category	Non-hospitalized (*n*=508)	Hospitalized (*n*=489)	*P*-value[a]
Age (years), *n* (%)	<40	300 (59%)	49 (10%)	2.90E-75
	40–60	157 (30.9%)	155 (31.7%)	
	>60	51 (10.1%)	285 (58.3%)	
Sex, *n* (%)	Female	283 (55.8%)	202 (41.3%)	7.31642E-06
	Male	225 (44.2%)	287 (58.7%)	
Highest educational level, *n* (%)	Secondary school	175 (34.4%)	260 (53.1%)	3.75178E-09
	Professional	333 (65.5%)	229 (46.9%)	
Comorbidities, *n* (%)	Hypertension	55 (10.8%)	209 (42.7%)	7.82792E-30
	Type 2 diabetes mellitus	27 (5.3%)	97 (19.8%)	7.39187E-12
	Asthma	12 (2.3%)	10 (2%)	0.900344245
	COPD	4 (0.7%)	35 (7.1%)	5.08776E-07
	Cardiovascular disease	10 (1.9%)	64 (13%)	4.87E-11
	Nephropathy	2 (0.3%)	40 (8.1%)	2.51175E-09
	Malignancy	9 (1.7%)	32 (6.5%)	0.000279061
	Autoimmune disease	2 (0.3%)	25 (5.1%)	1.11482E-05
	Smoking (former)	86 (16.9%)	76 (15.5%)	6.17715E-05
	Smoking (current)	35 (6.8%)	7 (1.4%)	
	Obesity	42 (8.3%)	109 (22.2%)	1.15776E-09
	HIV	2 (0.3%)	6 (1.2%)	0.263027525
	Thyroid disease	30 (5.9%)	63(12.8%)	0.000234681
Symptoms, *n* (%)	Fever	196 (38.5%)	273 (55.8%)	7.02857E-08
	Cough	206 (40.5%)	338 (69.1%)	2.39592E-19
	Fatigue	236 (46.4%)	287 (58.6%)	0.000142592
	Dyspnoea	98 (19.2%)	317 (64.8%)	9.46952E-48
	Diarrhoea	87 (17.1%)	86 (17.5%)	0.913616235
	Sore throat	151 (29.8%)	84 (17.1%)	4.40343E-06
	Anosmia	222 (43.7%)	49(10%)	1.53071E-32
	Disgeusia	188 (37%)	44 (8.9%)	2.80033E-25
	Chest pain	59 (11.7%)	51(10.4%)	0.62004627
	Nasal congestion	142 (28%)	35(7.1%)	1.77962E-17
	Abdominal pain	29 (5.7%)	33 (6.7%)	0.583361711
	Cyanosis	1 (0.1%)	6 (1.2%)	0.116863239
	Headache	220 (43.5%)	103 (21%)	6.6173E-14
	Conjunctivitis	25 (4.9%)	0	1.882E-06
	Chills	157 (30.9%)	97 (19.8%)	8.24311E-05
Complications, *n* (%)	Cardiac	1 (0.19%)	47 (9.6%)	1.14455E-11
	Respiratory	6 (1.1%)	287 (58.6%)	1.29436E-87
	Renal	0	79 (16.1%)	1.2216E-20
	Neurological	2 (0.3%)	34 (6.9%)	7.43065E-08
	Thromboembolic	0	28 (5.7%)	7.43065E-08
Body mass index (kg/m^2) (mean, range)		25.6 (16.3–43)	27.1 (16.2–52.7)	1.88697E-07
Age (years) (mean, range)		39.8 (18.6–92.5)	61.8 (20.7–99.3)	3.14239E-97

COPD, chronic obstructive pulmonary disease; HIV, human immunodeficiency virus.
[a] Chi-squared test.

Centers for Disease Control and Prevention had not yet designated lineage B.1.621, although it had been detected in 28 US states (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017).

The authors support the growing concern about lineage B.1.621 as reflected by its designation as a VOI by ECDC. Previous to the findings of the present study, there was insufficient real-world, experimental or model-based evidence regarding the impact of lineage B.1.621. This lack of evidential substantiation may be because the worldwide prevalence of lineage B.1.621 is very low, reportedly <0.5% (Elbe and Buckland-Merrett, 2017; Shu and McCauley, 2017). There may be under-representation of lineage B.1.621 as a result of insufficient detection and reports by genomic surveillance, although Colombian sequencing efforts have provided >23% of the global GISAID variant sequences for lineage B.1.621. Lineage B.1.621 seems to have displayed heightened ease of transmission by playing a central role in variant prevalence, considering the current and most serious pandemic wave in Colombia. The latter is coupled with the restriction of cases to well-delimited regions in a specific geography, allowing the criteria for VOI consideration to be met (Janik et al., 2021).

As pertains to genomic characterization, lineage B.1.621 shows the accumulation of substitutions including I95I, Y144T and Y145S in the N terminal domain; R346K, E484K and N501Y in the receptor-binding domain; P681H in the S1/S2 cleavage site; insertion 146N in the spike protein (Laiton-Donato et al., 2021), and K417N S gene mutation (Public Health England, 2021).

Potential designation as a VOC requires an evident increase in disease severity, in addition to the attributes shared with VOIs. Through the decision tree analysis, the findings of this study shed light on the impact of lineage B.1.621, support its consideration as a possible VOC, and could prompt action such that it may become a target of strengthened public health measures and focused genomic surveillance. Although the Delta variant (lineage B.1.617.2) was first identified at the end of July 2021 in Colombia, this lineage has not yet shown the dominant prevalence seen in many parts of the world.

Lineage B.1.1.388 was also found to be associated with higher hospitalization and mortality rates. Lineage B.1.1.388 was initially and almost exclusively reported in Colombia (until recently in Ecuador and Spain), triggering PANGO to label it as another 'Colombian lineage' (Rambaut et al., 2020), and displays several distinctive substitutions that have not been designated as VOI or VOC at the time of submission of this article.

A high percentage of patients had symptoms (90.7%), mainly with influenza-like illness. At clinical presentation, disease severity allowed for outpatient management in 50.8% of cases,

Table 6
Demographic factors and baseline clinical characteristics of the study population stratified by viral clade (*n*=654).

	Total (*n*=652)	G (*n*=167)	GH (*n*=321)	GR (*n*=120)	Other (*n*=46)	*P*-value[a]
Age (years), *n* (%)						
<40	264 (40.3%)	62 (37.1%)	135 (42%)	60 (50%)	7 (15.2%)	2.38E-09
40–60	206 (31.5%)	48 (28.7 %)	118 (36.7%)	32 (26.6%)	8 (17.3%)	
>60	184 (28.2%)	57 (34.1 %)	68 (21.2%)	28 (23.3%)	31 (67.3%)	
Sex, *n* (%)						
Female	321 (49.1%)	74 (44.3%)	165 (51.4%)	60 (50%)	22 (47.8%)	0.516063827
Male	333 (50.9%)	93 (55.6%)	156 (48.5%)	60 (50%)	24 (52.1%)	
Residence, *n* (%)						
Bogotá	413 (63.2%)	130 (77.8%)	180 (56%)	78 (65%)	25 (54.3%)	1.36E-06
Cundinamarca	40 (6.1%)	15 (8.9%)	13 (4%)	6 (5%)	6 (13%)	
Cali	187 (28.6%)	21 (12.5%)	120 (37.3%)	33 (27.5%)	13 (28.2%)	
Other	14 (2.1%)	1 (0.5%)	8 (2.5%)	3 (2.5%)	2 (4.3%)	
Comorbidities, *n* (%)						
Hypertension	137 (20.9%)	40 (23.9%)	61 (19%)	21 (17.6%)	15 (32.6%)	0.098673612
Type 2 diabetes mellitus	71 (10.8%)	19 (11.3%)	28 (8.7%)	13 (10.9%)	11 (23.9%)	0.021689056
Asthma	18 (2.7%)	5 (2.9%)	9 (2.8%)	3 (2.5%)	1 (2.1%)	0.988589637
Cardiovascular disease	45 (6.8%)	15 (8.9%)	18 (5.6%)	8 (6.6%)	4 (8.6%)	0.530351676
Nephropathy	24 (3.6%)	8 (4.7%)	6 (1.8%)	5 (4.1%)	5 (10%)	0.015694572
COPD	24 (3.6%)	11 (6.5%)	9 (2.8%)	4 (3.3%)	0	0.089988724
Malignancy	31 (4.7%)	10 (5.9%)	11 (3.4%)	3 (2.5%)	7 (15.2 %)	0.002501822
Autoimmune disease	14 (2.1%)	4 (2.3%)	9 (2.8%)	0	1 (2.1%)	0.340663261
HIV	5 (0.7%)	1 (0.6%)	3 (0.9%)	1 (0.8%)	0	0.908969613
Smoking (former)	118 (18%)	33 (20%)	55 (17.1%)	20 (16.6%)	10 (21.7%)	0.26392099
Smoking (current)	25 (3.8%)	8 (4.8%)	11 (3.4%)	3 (9.1%)	3 (6.5%)	0.26392099
Obesity	93 (14.2%)	24 (14.3%)	41 (12.7%)	14 (11.6%)	14 (30.4%)	0.011141853
Symptomatology, *n* (%)						
Symptomatic, *n* (%)	592 (90.5%)	159 (95.2%)	283 (88.1%)	105 (87%)	45 (97.8%)	0.014785209
Asymptomatic, *n* (%)	62 (9.5%)	8 (4.7%)	38 (11.8%)	15 (12.5%)	1 (0.2%)	
Symptoms, *n* (%)						
Fever	289 (44.1%)	80 (47.9%)	144 (44.8%)	46 (38.3%)	20 (43.4%)	0.446571927
Cough	329 (50.3%)	98 (58.6%)	152 (47.3%)	54 (45%)	26 (56.5%)	0.048897461
Fatigue	345 (52.7%)	98 (59.2%)	159 (49.5%)	64 (53.3%)	24 (52.1%)	0.239754642
Dyspnoea	224 (34.2%)	71 (42.5%)	91 (28.3%)	37 (30.8%)	25 (54.3%)	0.000286422
Diarrhoea	101 (15.4%)	26 (15.5%)	46 (14.3%)	26 (21.6%)	3 (6.5%)	0.083212644
Sore throat	163 (24.9%)	41 (24.5%)	78 (24.2%)	32 (26.6%)	12 (26%)	0.95862538
Anosmia	207 (31.6%)	47 (28.1%)	107 (33.3%)	46 (38.3%)	7 (15.2%)	0.022403231
Disgeusia	185 (28.2%)	46 (27.5%)	94 (29.2%)	37 (30.8%)	8 (17.3%)	0.350710859
Chest pain	77 (11.7%)	16 (9.5%)	39 (12.1%)	19 (15.8%)	3 (6.5%)	0.267781866
Nasal congestion	132 (20.1%)	38 (22.7%)	60 (18.6%)	32 (26.6%)	2 (4.3%)	0.009658253
Abdominal pain	32 (4.8%)	6 (3.5%)	16 (4.9%)	6 (5%)	4 (8.6%)	0.563171113
Cyanosis	3 (0.4%)	0	1 (0.3%)	0	2 (4.3%)	0.000809846
Headache	223 (34%)	58 (34.7%)	113 (35.2%)	47 (39.1%)	11 (23.9%)	0.332417704
Complications, *n* (%)						
Cardiac	25 (3.8%)	10 (6%)	9 (2.8%)	4 (3.3%)	2 (4.3%)	0.361438087
Respiratory	141 (21.5%)	51 (30%)	50 (15.5%)	22 (18.3%)	18 (39.1%)	2.3284E-05
Renal	41 (6.2%)	17 (10.2%)	11 (3.4%)	7 (5.8%)	6 (13%)	0.005898232
Neurological	16 (2.4%)	4 (2.4%)	6 (1.8%)	3 (2.5%)	3 (6.5%)	0.302293399
Thromboembolic	16 (2.4%)	5 (3%)	4 (1.2%)	4 (3.3%)	3 (6.5%)	0.124023067
Clinical outcomes, *n* (%)						
Ambulatory	409 (62.7%)	83 (49.7%)	231 (71.9%)	88 (73.3%)	7 (15.2%)	1.99E-13
Hospitalized	159 (24.3%)	59 (35.3%)	56 (17.4%)	19 (15.8%)	25 (54.3%)	
Intensive care unit	45 (6.8%)	10(6%)	19 (5.9%)	7 (5.8%)	9 (1.9%)	
Deceased	41(6.2%)	15 (9%)	15 (4.6%)	6 (5%)	5 (10.8%)	

COPD, chronic obstructive pulmonary disease; HIV, human immunodeficiency virus.
[a] Chi-squared test.

hospitalization in 33% of cases, and ICU admission in 9.2% of cases. The mortality rate was 6.9%. The ICU admission rate in this study is similar to rates reported previously (5–32%) (Guan et al., 2020; Huang et al., 2020).

In agreement with others, pulmonary complications were seen most frequently in this study (29.4%). Other systems had minor involvement, mainly associated with the multi-systemic impact of COVID-19. As in most series, the conditions most frequently associated with hospitalization or ICU admission were age >60 years, male sex, hypertension, cardiovascular disease, nephropathy, obesity and thyroid disease (Chen et al., 2020; Guan et al., 2020; Huang et al., 2020; Richardson et al., 2020; Wu et al., 2020).

In agreement with the literature, age was the most important variable associated with the need for hospitalization (21% in patients aged <38 years vs 77% in patients aged >59 years). Remarkably, lineages A, B, B.1.1.388, B.1.1.434, B.1.153, B.1.36.10, B.1.411, B.1.471, B.1.558 and B.1.621 increased the likelihood of hospitalization (82% vs 35%) in relatively young patients.

Similar results were seen in terms of mortality (3% in patients aged <71 years vs 26% in patients aged >88 years). Lineages B.1.1, B.1.1.388, B.1.523 and B.1.621 were associated with increased risk of death (62% vs 21%).

The bivariate analysis showed a significant association between mortality and clade G and 'other' clades, possibly because patients aged >60 years and with comorbidities were over-represented in these two clades, leading to possible bias. Furthermore, in the first phase of the study, a higher proportion of clades G and GH were identified, as opposed to clade GH and 'other' clades in the second

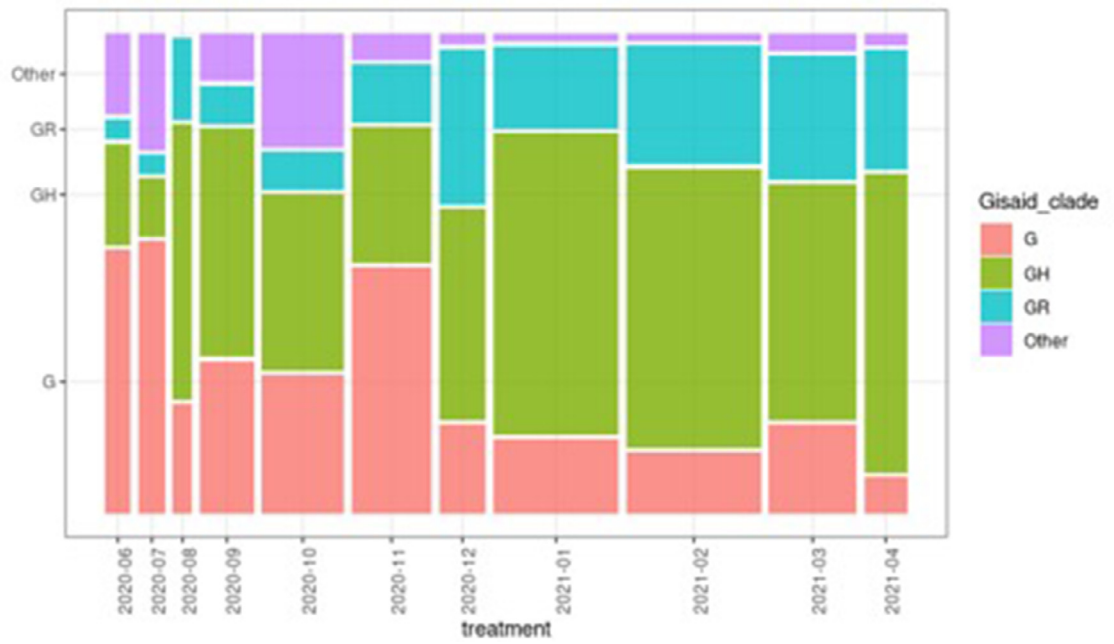

Figure 2. Mosaic plot showing the monthly distribution of Global Initiative on Sharing All Influenza Data (GISAID) clades during the study time window.

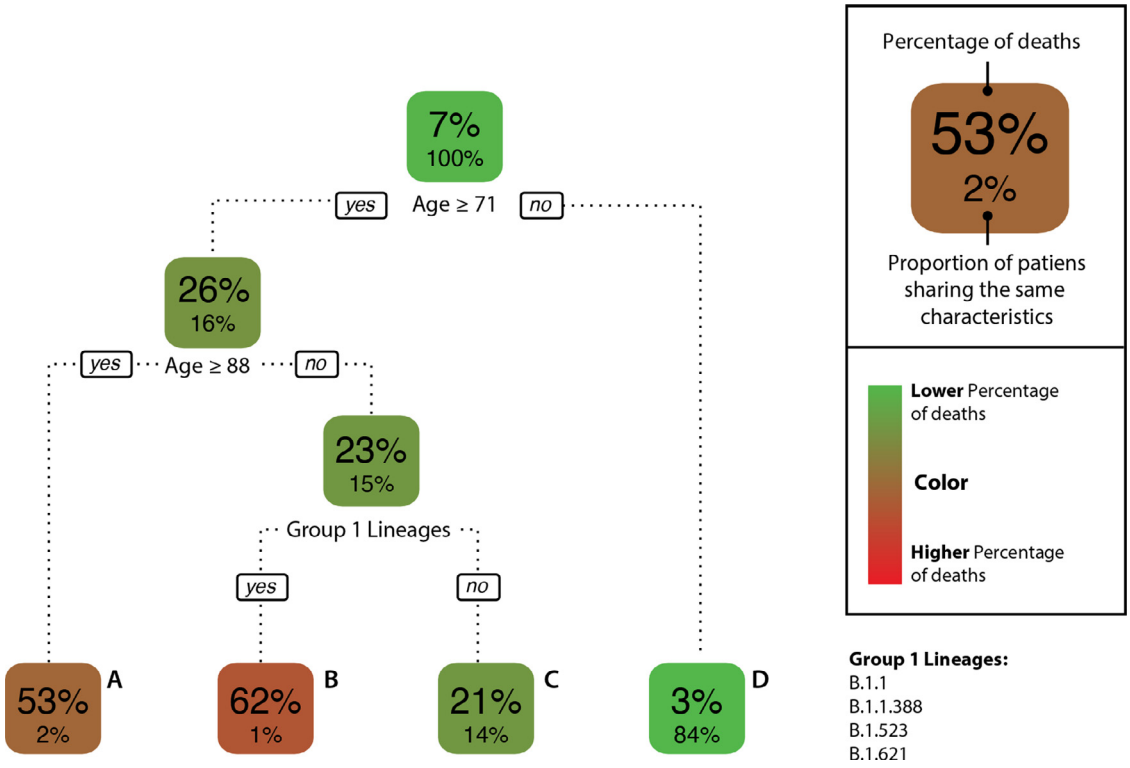

Figure 3. Classification and regression tree for mortality. Two variables (age and four lineages) were associated with higher mortality. This tree identifies patients with commonalities, classified into four subgroups. (A) Patients aged >88 years, with mortality rate of 53% (2% of the total study sample). (B) Patients aged 71–88 years who presented with any of the following four lineages: B.1.1, B.1.1.388, B.1.523 or B.1.621. In this group, 62% of patients died (1% of the sample). (C) Patients aged 71–88 years who presented with lineages different from the four described above, and had 21% mortality rate (14% of the sample). (D) Patients aged <71 years, who had 3% mortality rate (84% of the total sample).

phase. The association between mortality and clade G and 'other' clades was no longer seen on the multi-variate analysis.

Several studies have explored associations between clinical outcomes and SARS-CoV-2 clades, and findings have been broadly divergent. Hamed et al. (2021) found that clades GH and GR were associated with cases of severe disease/death, while clades S, G and GV were associated with mild/asymptomatic cases. Clades L and V showed no significant statistical association.

Young et al. (2021) showed that clades L and V were significantly associated with disease severity and a more intense systemic inflammatory response, while clade G was not associated with higher disease severity or transmissibility.

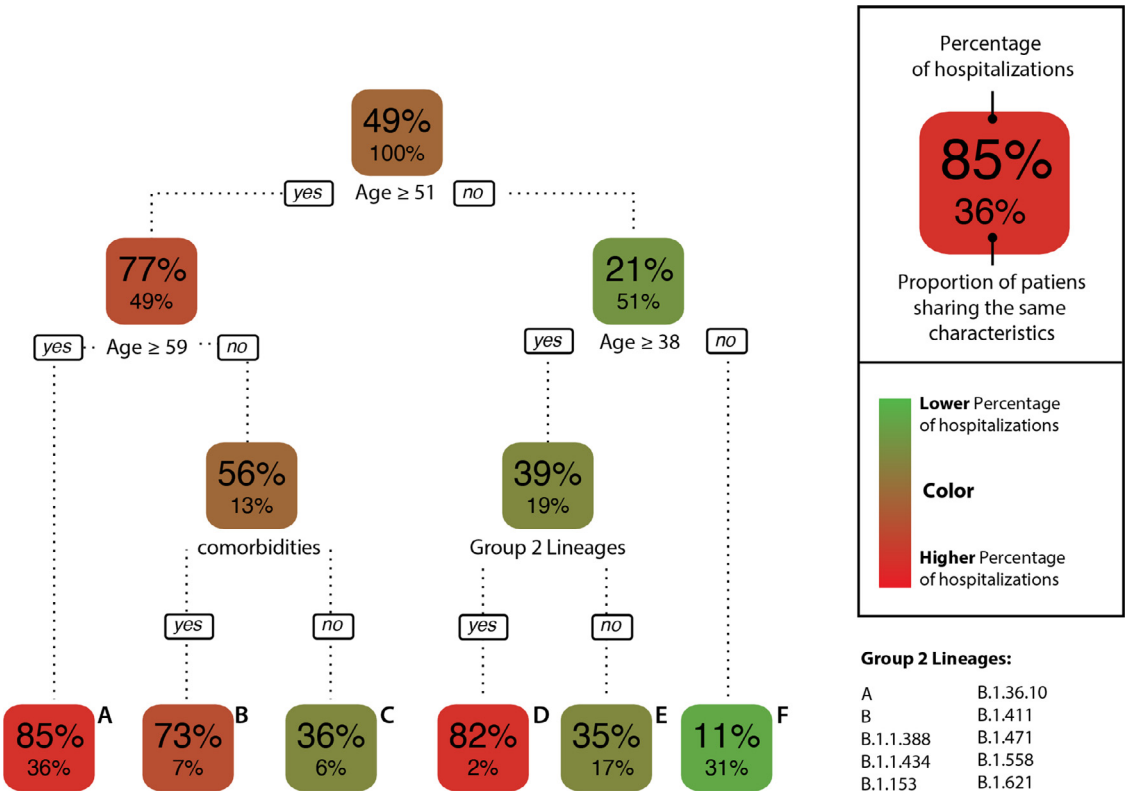

Figure 4. Classification and regression tree for hospitalization. The tree seeks to identify the variables determining the clinical severity of coronavirus disease 2019 in terms of the need for hospitalization. Six groups were identified. (A) Patients aged >59 years, with 85% hospitalization rate (36% of the sample). (B) Patients aged 51–59 years with one or more comorbidities, with 73% hospitalization rate (7% of the sample). (C) Patients aged 51–59 years with no comorbidities, with 36% hospitalization rate (6% of the sample). (D) Patients aged 38–51 years with any of the following viral lineages: A, B, B.1.1.388, B.1.1.434, B.1.153, B.1.36.10, B.1.411, B.1.471, B.1.558 or B.1.621, with 82% hospitalization rate (2% of the sample). (E) Patients aged 38–51 years but not presenting any group D lineage, with 35% hospitalization rate (17% of the sample). (F) Patients aged <38 years, with 11% hospitalization rate (31% of the sample).

Nakamichi et al. (2021) explored associations between genetic variants and hospitalization and mortality due to SARS-CoV-2 infection, designating two clear clades from hierarchical clustering of the sequence variants. Clade 2, predominantly composed of clade S, showed a trend toward poorer clinical outcomes compared with clade 1, predominantly constituted by clade GH.

Taxonomic classification into clades provides for a relatively coarse characterization, possibly lacking sufficient granularity for clinical correlation as clades are constituted by lineages designated or not as VOCs or VOIs. Additionally, clade composition could be modified over time, depending on the identification of new lineages and the understanding of the clinical impact of previously designated lineages on disease severity. In this sense, the allocation of a VOC or VOI in a specific clade could give a false perception that the clade in itself is the variable associated with higher disease severity or transmissibility, in place of lineage which may be what relates to higher disease severity. For instance, the Alpha variant belongs to clade GRY (previously clade GR); the Beta, Epsilon and Iota variants belong to clade GH; the Gamma, Zeta, Theta and Lambda variants belong to clade GR; and the Delta, Eta and Kappa variants belong to clade G. Such a wide distribution of VOIs and VOCs hinders the exploration of possible associations (World Health Organization, 2021). Finally, these are dynamic lineage groupings prone to reclassification and reallocation.

Conclusions

This study investigated associations between SARS-CoV-2 lineages and hospitalization and mortality rates. The findings, in agreement with those of others, make plausible the consideration of lineage B.1.621 as a VOI. In the authors' view, the designation of lineage B.1.621 as a VOI merits consideration due to the fixation and significant increase in detection frequency over a relatively short interval, and because of the high detection rate within the protracted third wave of SARS-CoV-2 infection in Colombia, which was the third-largest COVID-19 caseload in Latin America, and 12th largest globally. As disease severity for this lineage will be better characterized in further studies, possible designation as a VOC may be considered.

This was a cohort study, viewed in contrast to GISAID data and ecological studies. It is suggested that in public databases, such as GISAID, it would be beneficial to make available clinical information associated with the sequence data in order to foment timely association of genomic data with clinical variables. This may expedite consideration of the classification of variants as VOCs or VOIs, in turn triggering strict surveillance in terms of public health and other policies related to management of the pandemic.

This study included patients who were followed-up until the clinical outcome definitions were met, ensuring the fidelity of the information collected. In addition, the 12-month temporal coverage sheds light on the evolution of SARS-CoV-2 and the dynamics related to introduction of the pathogen from other countries. Pre-analytical specimen management included a unique platform for obtaining sequences and automated library preparation, thus controlling for cross-contamination and operator-dependent error.

This study has limitations. First, it used a convenience sample, which limits its generalizability. While participants resided in the largest and third-largest cities in Colombia, which were the main national outbreak epicentres, particularly in the first half of 2020, Caribbean coastal populations were excluded, in whom most

B.1.621 cases have been detected. Second, the rate of successful sequencing was 67%; failure was mainly due to lower viral sample contents and/or RNA degradation. Third, the study design was ambispective; recall bias may have affected the accuracy of symptom information provided in retrospective cases. Nonetheless, good agreement with the literature leads the authors to infer that recall bias was probably minor. Finally, patient recruitment concluded shortly before the third, and most serious to date, pandemic wave in Colombia, in which the detection of lineage B.1.621 soared, thus explaining its relatively low frequency in this cohort study.

Conflict of interest statement

None declared.

Acknowledgements

The authors wish to thank Lina Marcela Méndez Castillo, Andrés Felipe Torres Gómez and Danyela Faisury Valero Rubio for supporting the laboratory procedures; and Andrea del Pilar Hernández Rodríguez, Andrés Felipe Patiño Aldana, Liseth Belinda Cifuentes Saenz, Natalia Andrea Pedraza López, Mateo Andrés Díaz Quiroz, Kevin Daniel Rivera Mendoza, Isabella Caicedo, María Alejandra Urbano, Alejandra Hidalgo and all other members of the Clinical Investigation Group that supported the collection of clinical data and development of this research.

Funding

This work was supported by governmental funding from the Colombian Ministry of Science, Technology and Innovation (Grant No. 366-2020).

Ethical approval

This study was approved by the institutional review boards of Universidad del Rosario and participating hospital research centres. All international and national bioethical principles and regulations for clinical investigation in human subjects were followed.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2021.12.326.

References

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Boca Raton: Routledge; 2017.
Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020;395:507–13.
CoVsurver. 2021. Available at: https://mendel3.bii.a-star.edu.sg/METHODS/corona/beta/indexAnno2.html (accessed 11 June 2021).
European Centre for Disease Prevention and Control. SARS-CoV-2 variants of concern as of 5 August 2021. Stockholm: ECDC; 2021 Available at https://www.ecdc.europa.eu/en/covid-19/variants-concern (accessed 17 August 2021).
Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Challeng (Hoboken, NJ) 2017;1(1):33–46 doi:10.1002/gch2.1018.
Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, et al. Emergence of SARS-CoV-2 B.1.1.7 lineage – United States, December 29, 2020—January 12, 2021. MMWR 2021;70:95–9.
Guan W-J, Ni Z-Y, Hu Y, Liang W-H, Ou C-Q, He J-X, et al. China Medical Treatment Expert Group for COVID-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020;382:1708–20.
Hamed SM, Elkhatib WF, Khairalla AS, Noreddin AM. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. Sci Rep 2021;11:8435.
Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497–506.
Huang S-W, Wang S-F. SARS-CoV-2 entry related viral and host genetic variations: implications on COVID-19 severity, immune escape, and infectivity. Int J Mol Sci 2021;22:3060.
Instituto Nacional de Salud. Coronavirus Colombia. Bogotá: INS; 2021 Available at https://www.ins.gov.co/Noticias/paginas/coronavirus.aspx (accessed 11 June 2021).
Janik E, Niemcewicz M, Podogrocki M, Majsterek I, Bijak M. The emerging concern and interest SARS-CoV-2 variants. Pathogens 2021;10:633.
Johns Hopkins Coronavirus Resource Center. COVID-19 map. 2021. Available at: https://coronavirus.jhu.edu/map.html (accessed 11 June 2021).
Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucl Acid Res 2002;30:3059–66.
Laiton-Donato K, Franco-Muñoz C, Álvarez-Díaz DA, Ruiz-Moreno HA, Usme-Ciro JA, Prada DA, et al. Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. Infect Genet Evol 2021;95.
Lamptey J, Oyelami FO, Owusu M, Nkrumah B, Idowu PO, Adu-Gyamfi EA, et al. Genomic and epidemiological characteristics of SARS-CoV-2 in Africa. PLoS Negl Trop Dis 2021;15.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 2020;37:1530–4.
Nakamichi K, Shen JZ, Lee CS, Lee A, Roberts EA, Simonson PD, et al. Hospitalization and mortality associated with SARS-CoV-2 viral clades in COVID-19. Sci Rep 2021;11:4802.
Peacock TP, Penrice-Randal R, Hiscox JA, Barclay WS. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. J Gen Virol 2021;102(4).
Posada D. jModelTest: Phylogenetic model averaging. Mol Biol Evol 2008;25:1253–6.
Public Health England. SARS-CoV-2 variants of concern and variants under investigation. London: PHE; 2021.
Rambaut A, Holmes EC, Á O'Toole, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 2020;5:1403–7.
Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 2020;323:2052–9.
Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. Euro Surveill 2017;22 pii=30494.
World Health Organization. Tracking SARS-CoV-2 variants. Geneva: WHO; 2021 Available at https://www.who.int/emergencies/emergency-health-kits/trauma-emergency-surgery-kit-who-tesk-2019/tracking-SARS-CoV-2-variants (accessed 17 August 2021).
Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med 2020;180:934–43.
Young BE, Wei WE, Fong S-W, Mak T-M, DE Anderson, Chan Y-H, et al. Association of SARS-CoV-2 clades with clinical, inflammatory and virologic outcomes: an observational study. EBioMed 2021;66.