![THE UNIVERSITY of EDINBURGH]

Edinburgh Research Explorer

Genomic epidemiology of SARS-CoV-2 variants during the first two years of the pandemic in Colombia

OPEN ACCESS

Genomic epidemiology of SARS-CoV-2 variants during the first two years of the pandemic in Colombia

Cinthy Jimenez-Silva [1,2,15✉], Ricardo Rivero [3,4,15✉], Jordan Douglas [1,5], Remco Bouckaert[1,6], Ch. Julian Villabona-Arenas [7], Katherine E. Atkins [7,8], Bertha Gastelbondo[3,9,10], Alfonso Calderon [3], Camilo Guzman [3,11], Daniel Echeverri-De la Hoz [3], Marina Muñoz[12], Nathalia Ballesteros[12], Sergio Castañeda[12], Luz H. Patiño[12], Angie Ramirez[12], Nicolas Luna[12], Alberto Paniz-Mondolfi [13], Hector Serrano-Coll[14], Juan David Ramirez[12,13✉], Salim Mattar [3✉] & Alexei J. Drummond [1,2,6]

Abstract

Background The emergence of highly transmissible SARS-CoV-2 variants has led to surges in cases and the need for global genomic surveillance. While some variants rapidly spread worldwide, other variants only persist nationally. There is a need for more fine-scale analysis to understand transmission dynamics at a country scale. For instance, the Mu variant of interest, also known as lineage B.1.621, was first detected in Colombia and was responsible for a large local wave but only a few sporadic cases elsewhere.

Methods To better understand the epidemiology of SARS-Cov-2 variants in Colombia, we used 14,049 complete SARS-CoV-2 genomes from the 32 states of Colombia. We performed Bayesian phylodynamic analyses to estimate the time of variants' introduction, their respective effective reproductive number, and effective population size, and the impact of disease control measures.

Results Here, we detect a total of 188 SARS-CoV-2 Pango lineages circulating in Colombia since the pandemic's start. We show that the effective reproduction number oscillated drastically throughout the first two years of the pandemic, with Mu showing the highest transmissibility (Re and growth rate estimation).

Conclusions Our results reinforce that genomic surveillance programs are essential for countries to make evidence-driven interventions toward the emergence and circulation of novel SARS-CoV-2 variants.

Plain Language Summary

Colombia reported its first COVID-19 case on 6th March 2020. By April 2022, the country had reported over 6 million infections and over 135,000 deaths. Here, we aim to understand how SARS-CoV-2, the virus that causes COVID-19, spread through Colombia over this time and how the predominant version of the virus (variant) changed over time. We found that there were multiple introductions of different variants from other countries into Colombia during the first two years of the pandemic. The Gamma variant was dominant earlier in 2021 but was replaced by the Delta variant. The Mu variant had the highest potential to be transmitted. Our findings provide valuable insights into the pandemic in Colombia and highlight the importance of continued surveillance of the virus to guide the public health response.

[1] Centre for Computational Evolution, University of Auckland, Auckland, New Zealand. [2] School of Biological Sciences, University of Auckland, Auckland, New Zealand. [3] Instituto de Investigaciones Biológicas del Trópico (IIBT), Facultad de Medicina Veterinaria y Zootecnia, Universidad de Córdoba, Montería, Colombia. [4] Paul G. Allen School for Global Health, Washington State University, Pullman, Washington, USA. [5] Department of Physics, University of Auckland, Auckland, New Zealand. [6] School of Computer Science, University of Auckland, Auckland, New Zealand. [7] Centre for Mathematical Modelling of Infectious Diseases and Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK. [8] Centre for Global Health, Usher Institute, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK. [9] Grupo de Investigaciones Microbiológicas y Biomédicas de Córdoba-GIMBIC, Universidad de Córdoba, Monteria, Colombia. [10] Grupo de Salud Pública y Auditoría en Salud, Corporación Universitaria del Caribe- CECAR, Sincelejo, Colombia. [11] Grupo de Investigación, Evaluación y Desarrollo de Farmacos y Afines - IDEFARMA, Universidad de Córdoba, Montería, Colombia. [12] Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia. [13] Molecular Microbiology Laboratory, Department of Pathology, Molecular and Cell-based Medicine, Icahn School of Medicine at Mount Sinai, New York, USA. [14] Instituto Colombiano de Medicina Tropical-Universidad CES, Medellín, Colombia. [15] These authors contributed equally: Cinthy Jimenez-Silva, Ricardo Rivero. ✉email: cjim882@aucklanduni.ac.nz; ricardoriveroh@correo.unicordoba.edu.co; juan.ramirezgonzalez@mountsinai.org; smattar@correo.unicordoba.edu.co

Colombia reported its first confirmed SARS-CoV-2 infection on 06 March 2020 in a traveler returning from Milan, Italy[1]. By April 2022, the country had reported more than 6 million SARS-CoV-2 infections and over 135,000 deaths[2] (Fig. 1, Table 1). According to epidemiological data, the SARS-CoV-2 epidemic in Colombia has been characterized by four pandemic waves with exponential growth in cases[3]. A wide range of strategies has been implemented to mitigate these surges of cases. That includes restrictions on mobility (such as school and airport closures), advice on mask use and physical distancing in public places, and vaccination[4–7]. Nevertheless, the current number of cases shows that the transmission of the virus is far from being under control, and those mitigation strategies may be insufficient[8]. However, the difficulties in achieving control were unlikely caused by strategy choice but rather by changes in the virus's transmissibility[9].

SARS-CoV-2 genetic diversity has been described using lineages, and a multiple nomenclature system has been established[10]. Notably, large-scale sequencing has led to the identification of genetic variations with enhanced transmissibility, virulence, or evasion of host immune response around the world[11,12]. The Technical Advisory Group on Virus Evolution from World Health Organization (WHO) labels the variants that pose an increased risk to global public health[13] as "Variants of Concern" (VOC). These include Alpha (WHO nomenclature) or B.1.1.7 (Pango nomenclature), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2 + AY.x), and Omicron (BA.1 & BA.2). Genomic surveillance has also led to identifying variants that carry mutations in the spike protein that may confer higher transmissibility and immune escape (such as mutations D614G, E484K/Q, K417T, N501Y, and P681H)[14–17]. These variants are termed "Variants of Interest" (VOI) and include Mu (B.1.621) and Lambda (C.37)[14,18–21]. Moreover, genomic surveillance has enabled phylodynamic investigation that has been vital to understanding global and local dynamics and tracing the zoonotic and time of origins[15,22,23].

After its first report in September 2020, the alpha variant soon spread around the world, and became the dominant variant in

many countries. But this was not the case in Colombia, where other important variants might have been circulating instead. For instance, the VOI Mu was first detected in Colombia and was responsible for large local outbreaks (in the presence of other SARS-CoV-2 variant of concern (VOCIs), including Alpha) but caused a few sporadic outbreaks elsewhere. Phylodynamics analysis of SARS-CoV-2 has led to many relevant findings, but more insights are needed from different epidemiological settings to understand better its spread and more effective approaches to control. To help achieve this, we aim to describe epidemiological trends and characterize the spatio-temporal dynamics of the most prevalent SARS-CoV-2 variants in Colombia using Bayesian Coalescent Skyline and Birth-death Skyline phylodynamic models. This study describes the epidemiological trends and circulating SARS-CoV-2 variants during the first two years of the pandemic in Colombia from a phylodynamic perspective. Our analysis reveals significant fluctuations in the effective reproductive number over the first two years of the pandemic, with the Mu variant demonstrating the highest transmissibility (as determined by Re and growth rate estimation). These findings underscore the crucial role of genomic surveillance programs in enabling countries to implement evidence-based interventions to address the emergence and spread of new SARS-CoV-2 variants.

Methods

Novel Genome sequence data. For this study, We collected Nasopharyngeal swabs from 10,674 residents from: Bogotá (the Capital District), Cali (the Capital of Valle del Cauca state), and Córdoba state (The Capital city and small towns) for testing by RT-qPCR using an in-house protocol based on the amplification of SARS-CoV-2 E gene according to WHO guidelines[24]. The samples were obtained from three different localities. In all cases, the samples came from active cases, which means patients with symptoms described as COVID. All of them were collected in health institutes such as hospitals, clinics, or minor health centers in these three localities. We sequenced positive samples with the following data availabe: travel history(the latest country of travel), patient status (Asymptomatic, mild, severe, critic, and fatal), sample collection date, and vaccination status. Our selection criteria resulted in 610 samples: 86 samples from Córdoba, 122 from Cali, and 402 from Bogotá. We purified the ARN of SARS-CoV-2 from the selected samples using the GeneJet RNA Extraction Kit (ThermoFisher Scientific, cat no. K0732) and prepared the sequencing library following the ARTIC Network

---

Table 1 SARS-CoV-2 sequences available in GISAID (the Global Initiative on Sharing All Influenza Data) from each country of South America until 18th February 2022 vs.

Country	Total of sequences	Number of cases
Brazil	114,181 (0.40%)	27,940,119
Chile	21,539 (7.83%)	2,747,55
Argentina	16,501 (0.18%)	8,799,858
Peru	15,736 (0.45%)	3,474,965
Colombia	14,653 (0.24%)	6,035,143
Ecuador	4299 (0.53%)	808,925
Paraguay	1215 (0.19%)	632,444
Suriname	1027 (1.32%)	77,549
Uruguay	743 (0.09%)	800,833
Venezuela	297 (0.05%)	508,968
Bolivia	236 (0.02%)	887,089
Guyana	63 (0.10%)	62,537

number of cases (https://www.statista.com/statistics/1101643/latin-america-caribbean-coronavirus-cases/).
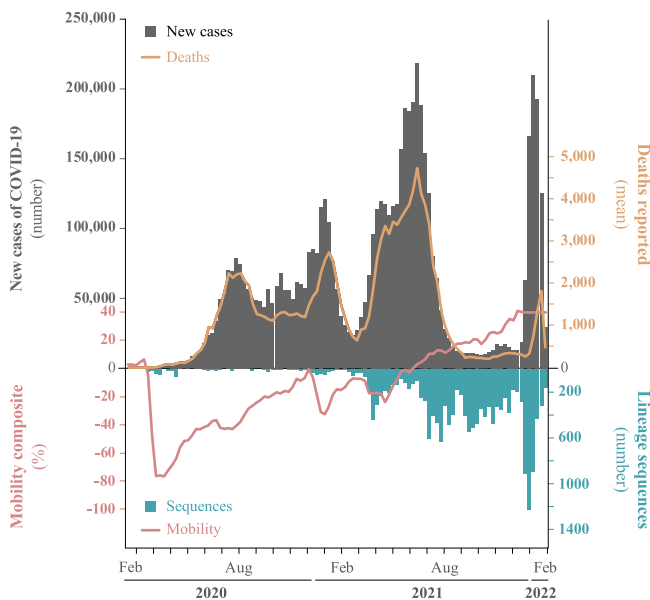

Fig. 1 Overview of the COVID-19 confirmed cases were sampled in Colombia during the two first years of the pandemic. Top: Number of daily reported cases (gray bars) and deaths (yellow line), up until February 2022. Bottom: total number of sequences (blue bars) and mobility data from 32 states in Colombia (red line) taken from covid19.healthdata.org.

Table 2 Summary of the SARS-CoV-2 sequences' alignment from Colombia per variant of interest assuming down-sampling criteria of at least two sequences per day.

Variant	Total-seq	Min date	Max date	Days	Weeks	First report Colombia (World)
B.1.1	145/504	2020-03-12	2022-01-03	510	72	2020-03-12 (2020-01-08)
B.1.111	83/251	2020-03-13	2021-06-12	454	64	2020-03-13 (2020-03-07)
B.1.1.348	74/189	2020-04-30	2021-12-31	384	54	2020-04-30 (2020-04-30)
B.1.420	79/118	2020-03-11	2021-06-03	118	16	2020-03-11 (2020-07-13)
B.1.1.7 (Alpha)	79/168	2021-02-15	2021-10-03	213	30	2021-02-15 (2020-09-03)
P.1 (Gamma)	337/922	2021-01-04	2021-12-01	332	47	2021-01-04 (2020-10-01)
B.1.617+AY (Delta)	192/4545	2020-12-07	2022-01-17	252	58	2020-12-07 (2020-10-05)
B.1.621 (Mu)	416/5225	2020-10-14	2021-12-15	330	47	2020-10-14 (2020-12-15)
C.37 (Lambda)	106/202	2021-03-30	2021-09-16	156	22	2021-03-30 (2020-07-21)
Omicron	159/766	2021-12-04	2022-01-21	42	06	2021-12-04 (2021-09-11)

The name of each variant is defined as Pangolin lineage, and WHO nomenclature is indicated in parentheses. Height: values given in days and years in parentheses. First report: Earliest date of each lineage is reported at https://cov-lineages.org/. Omicron: B.1.1.529 + BA.1/BA.1.1 lineages

protocol[25] and sequenced the libraries using the Oxford Nanopore MinION sequencer. Then, we processed (base-calling and demultiplexing) the raw data using Guppy v3.4.6[26] and filtered reads by quality and length to remove short, and low-quality reads (threshold lower than 20X was assumed as N). Finally, we assembled consensus genomes following the ARTIC bioinformatics pipeline[27]. Sample collection was led by the Instituto de Investigaciones Biologicas del Tropico (IIBT) at Universidad de Córdoba and the Centro de Investigaciones en Microbiologia y Biotecnologia (CIUMBIUR) at Universidad del Rosario, which are part of the official laboratories authorized by Colombia's Ministry of Health for testing and genomic surveillance or SARS-CoV-2. Sample collection in Córdoba was approved by the Ethics committee of Universidad de Córdoba/IIBT (Acta N° 0410-2020) in compliance with CDC's guidelines for safe work practices in human diagnostic[28]. Sample collection in Bogotá and Cali was approved by Universidad del Rosario's Research Ethics committee (DVO005 1550-CV1400) in compliance with Helsinki's declaration[29]. Informed consent was obtained from all patients.

Data curation. In Colombia, the National Genomic Surveillance Network (made up of 23 laboratories distributed across the country) of the National Institute of Health (Red de Vigilancia Genómica del Instituto Nacional de Salud) sequenced and curated most of the genomic data reported to GISAID. Therefore, we retrieved all SARS-COV-2 genome sequences from Colombia shared via GISAID (N=14,049, last accessed on 2022-02-02) and combined them with the novel genome sequences. We identified the variant of each genome sequence using the Pango nomenclature[30]. We excluded sequences with bad quality based on six different control metrics implemented in Nextclade[31]: no more than 10% ambiguous characters, no more than ten mixed sites, no more than 10% of missing data (Ns > 3000), no more than two mutation clusters, number of insertions or deletions that are not a multiple of three and number of stop codons that occur in unexpected places (2 stop codons are bad), and any outlier sequence as reported by Nextstrain[32]. We also removed sequences with incongruent lineage classification between Pangolin and Nextclade. Additional information for all sequences submitted and downloaded from GISAID is available in (Supplementary Data 1 and 2). We down-sampled the alignments by variant and homogeneously through the time (to have at least one sequence per day); any variant with ≥100 samples was considered a major variant. This down-sampling resulted in 1670 sequences distributed in 10 different alignments (Table 2, Supplementary Fig. 7).

Measuring the variant's growth rate and the effective reproduction number R_e. To estimate the growth advantage of each variant, we used the frequencies (weekly) of the SARS-CoV2 variants to fit multinomial logistic regression models that include a natural cubic spline to allow for slight variation in the growth rate of a given variant as a function of the sampling date. These multinomial spline models consider the frequencies of the major SARS-CoV2 variants as separate outcome levels (the remaining variants are aggregated in the category of "other variants") to simultaneously model the competition among all variants. Four models were fitted for each of the four wave periods in Colombia using the nnet package v.7.3-17 in R. v.3.5.0[33], one without splines and three with splines and Degrees of Freedom values ranging from 1 to 3. Best-fit model for each wave was selected based on Bayesian Information Criterion (BIC). The models were used to produce Muller plots to display the change in the relative frequencies of the major SARS-CoV-2 variants. Furthermore, estimates of the expected multiplicative effect on R_e based on the relative abundance of each variant were calculated assuming a gamma distributed generation time (Mean = 4.7 days, standard deviation = 2.9) using weighted effects contrasts and the package emtrends v.1.7.3[34,35] in R v.3.5.0[33].

Bayesian phylodynamic analysis. We aligned the sequence data of each major variant using MAFFT v7[36] and all the alignments were split by codon position. We tested the temporal signal of each alignment using a Neighbor-Joining tree that was inferred using the ape package v.5.6-2[37] in R v.3.5.0[33] and a regression of root-to-tip genetic distance against sampling time using TempEST v1.5.338[38]. The levels of temporal signal were assessed by visual inspection and by the correlation coefficient. All alignments showed a positive correlation (the correlation coefficient ranged between 0.0042 and 0.8) and appear to be suitable for phylogenetic molecular clock analysis in BEAST (Supplementary Fig. 5). We assumed a strict molecular clock as a prior for the clock rate in all cases as a log-normal distribution with a mean of 0.001 subs/site/year and standard deviation of 0.03 (parameterized using the shape and rate of that distribution), assuming from previous analysis[39]. We used this clock model with an informative prior reflecting recent estimates for the substitution rate of SARS-CoV-2 because the data did not evidence a robust temporal signal. All clock rates estimated were congruent between methods (Supplementary Fig. 6). We attribute it to the down-sampling strategy, which chose an alignment per variant.

We performed bayesian inference of phylogeny and estimated TMRCA of each node and the demographic dynamics (in terms of effective population sizes, N_e) over time of ten different alignments

(group of sequences from ten variants that were documented in Colombia) using two different tree priors: The Bayesian Coalescent Skyline (BCS)[40] and the recently implemented Bayesian Integrated Coalescent Epoch PlotS (BICEPS) model[41]. We evaluated the congruence between both models and encourage to use the last one (BICEPS) because it is computationally more efficient than BCS and allows extensive data sets analysis. We estimated the effective reproduction number (R_e) through time using a Bayesian birth-death skyline model[42] with ten and fourteen dimensions. Estimates of R_e using two different dimensions were compared to evaluate changes in the inferred R_e in some periods. All the posteriors values for the parameters of interest were reported as mean and credible intervals (CI), which is referred to as the 95% of high posterior density (HPD). We used the R-package bdskytools (https://github.com/laduplessis/bdskytools) to plot the smooth skyline, marginalizing our R_e estimates on a regular time grid (defined as the number of weeks that each variant has circulated) and calculating the HPD at each gridpoint. The models are available as packages in the platform BEAST v2.6.7[43]. In order to confirm the origin of Mu variant, we performed a phylogeography analysis using the Bayesian discrete phylogeography model (DPG)[44]. We considered migration between seven demes (Global regions and Colombia Country) assuming that the transition rates between locations were reversible.

We determined the Hasegawa-Kishino-Yano model (HKY) to be the best-fit nucleotide substitution models[45] without site heterogeneity for all alignments using BModelTest v1.2.1[46]. We used three independent Markov Chain Monte Carlo with 400 million iterations using the CoupledMCMC package (MC3) v1.0.2[47]. We diagnosed the MCMC samples using Tracer v1.7.2 (http://tree.bio.ed.ac.uk/software/tracer) until they reached effective sample sizes over 200 for all parameters. We summarized Maximum clade credibility trees (MCC) using TreeAnotator package. To visualize trees and outputs, we used Figtree v1.4.4 and R v.3.5.0[33] with packages: ape v5.6-2[48] and ggtree v3.4.0[49].

Metadata and statistical analysis. We accessed socio-demographic and COVID policy intervention variables to determine the association between each variant's R_e and N_e using generalized linear models. The variables were change in human mobility given in % units (as measured by cell phone mobility data); vaccine coverage (shows the percentage of people who receive at least one dose of a vaccine, and those who are fully vaccinated against COVID-19); estimated infections (the number of people we estimate are infected with COVID-19 each day, including those not tested); mask use (represents the percentage of the population who say they always wear a mask in public)[50], and lockdown policies[51], which is given by the stringency index (composite measure based on nine response indicators including school closures, workplace closures, and travel bans with value from 0 to 100 = strictest). We calculated Pearson correlation coefficients to avoid excessive co-linearity among explanatory predictors removing variables that exceeded 0.7. We transformed Predictors into log space and standardized to eliminate the effect of the magnitude of different co-variants. We perform the uni-variable linear regression model between dependent parameters (R_e and N_e) and possible explanatory variables (socio-demographic and COVID policy intervention variables). These statistical analyses were performed using package stats v4.3.0 in R v.3.5.0[33]. We reported the adjusted R square and p-value of <0.05 were considered statistically significant.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Initially, molecular testing and genome sequencing of SARS-CoV-2 were only performed by the National Health institute (INS), but capacity was increased with an additional 21 sequencing laboratories serving the 32 Colombian states[52].

As of February 2022, genomics surveillance in Colombia has generated 14,049 SARS-CoV-2 complete genomes, which represent 0.2% of the 5.9 million confirmed cases during this period. Compared with other South American countries, Colombia generated the fifth highest volume of SARS-CoV-2 genomic data during the first two years of the COVID-19 pandemic (Table 1). We generated 610 novel sequences from three different Colombian states for this study, and 13,444 were downloaded from GISAID.

We reconstructed the dynamics of the 10 predominant SARS-CoV-2 variants that circulated in Colombia using Bayesian phylodynamic modeling. These methods allowed us to estimate each variant's transmissibility with, effective reproductive number (R_e) and effective population size (N_e) (further details about this analysis are in the methods section).

Variant classification and distribution. The 14,049 SARS-CoV-2 complete genomes were grouped in 188 SARS-CoV-2 Pango lineages, which have circulated in Colombia. Despite the vast genetic diversity documented, only ten SARS-CoV-2 variants were predominant during the two first years of the pandemic (Fig. 2). Herein, we use the pango nomenclature name for those variants that are not labeled as VOIC (variants of interest or concern). These ten lineages include four pango lineages (B.1, B.1.1.1, B.1.420, and B.1.1.348), four variants of concern (Alpha: B.1.1.7; Gamma: P.1; Delta: B.1.617, AY.x; and Omicron: B.1.1.529, BA.1, BA.2.x), and two variants of interest (Lambda: C.37; and Mu: B.1.621, BB.1 and BB.2)[13]. The most populated states of Colombia (Antioquia, Cundinamarca, and Valle del Cauca) generated the highest number of sequences (<2000 sequences each). Mu, Delta, and Gamma variants were documented in 31, 28 and 29 states (out of 32). Alpha and Lambda were documented in 12 states, and Omicron was documented in 17 states. The most widespread variant, Mu, showed the highest prevalence in the capital district (Bogotá) (19.43%), and Antioquia state (19%) (Fig. 2b, c).

In the following sections, we describe the dynamics of COVID-19 in terms of the four COVID-19 waves reported in Colombia and considering the dates on which specific Colombian measures and strategies were raised and implemented, as was used by[4] which defined the first two pandemic periods in Colombia:

First period of the pandemic: from 06 March to 10 August 2020. The first wave of the COVID-19 pandemic in Colombia ran from when the first case was reported until just before the subsequent relaxation of the stringent non-pharmaceutical interventions (NPIs) that were implemented (10 August 2020). These NPIs included the declaration of a national emergency, closing of schools and universities, restriction of international flights, the closing of the international borders, and the first lockdown (Between 25 March 2020 and 18 June 2020). 552,523 cases were reported during this wave, 408 (0.07%) were sequenced, and 22 variants were identified. B.1 was the predominant (46.3%) variant, followed by B.1.111 (21.8%), B.1.420 (9.8%), and B.1.1.348 (1.71%), which co-circulated after their emergence. We dated the most recent common ancestors to exist around the 23 February 2020 (95% credible interval (CI) of 28 November 2019 to 28 February 2020), 27 February 2020 (CI: 10 February 2020–14 March 2020), 10 April 2020 (CI: 26 February 2020–24 March 2020), and 28 March 2020 (CI: 02 March 2020–20 April 2020) for
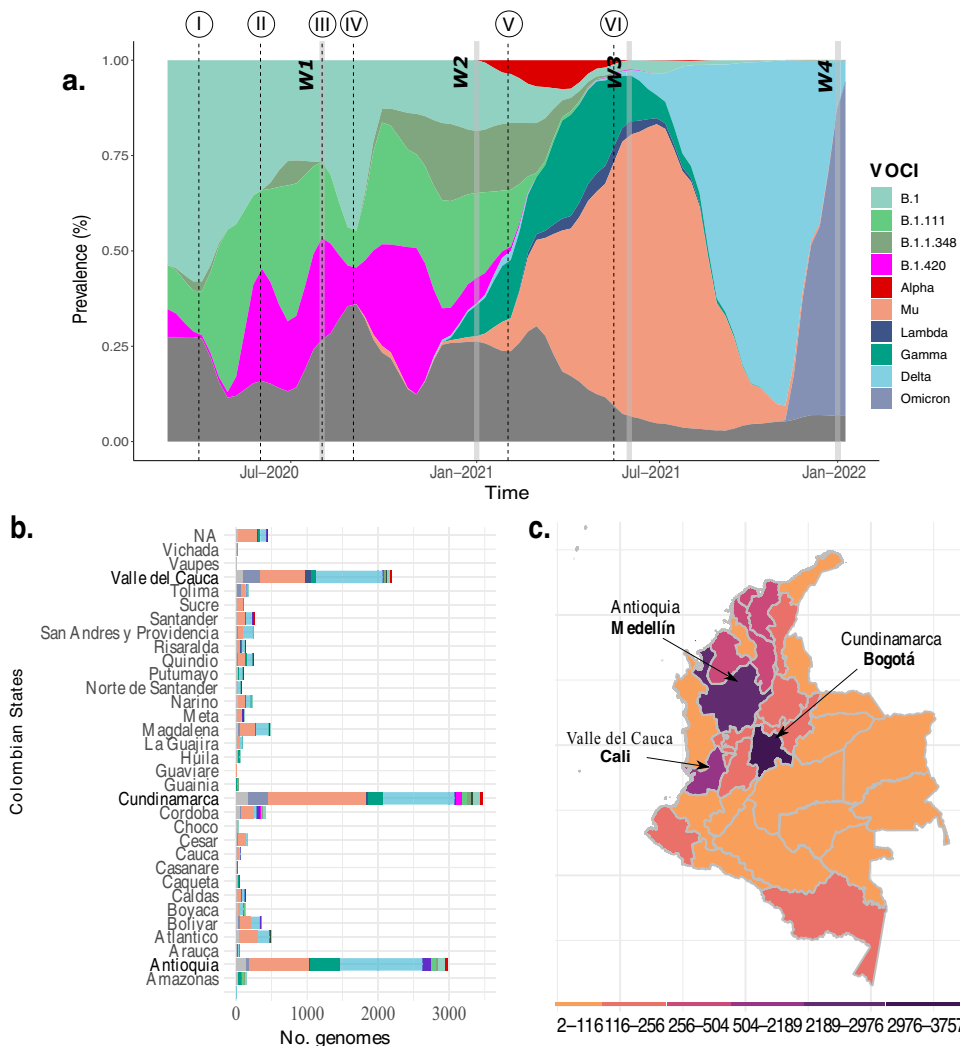
Fig. 2 SARS-CoV-2 genetic diversity. a The most prevalent Pangolin COVID-19 global lineages, SARS-CoV-2 Variants of Interest and SARS-CoV-2 variant of concern (VOCIs) until February 2022 in Colombia. Horizontal gray bars represent the four waves (W1, 2, 3, 4). The black dotted line indicated I) the first National Lockdown in Colombia on March 25, 2020, II) end of the first National lockdown on 18 June 2020, decreasing the restriction measures for controlling COVID-19, III) Policies implementation for economic reactivation 10 August 2020, IV) Reopening of domestic and international flights during September 2020, V) Vaccination phase 1 implemented on 17 February 2021, and VI) Vaccination phase 2 implemented on 17 June 2021. **b** The number of COVID-19 genomes isolated from 32 states of Colombia, and the capital city (Bogotá) available in GISAID (the Global Initiative on Sharing All Influenza Data) and the sequences obtained in this study. **c** Colombian map indicating the total genome collected per state.

B.1, B.1.111, B.1.420, and B.1.1.348, respectively (Fig. 3). Values of Re ranged between 0.37 and 1.88 for B.1, between 0.26 and 3.49 for B.1.111, between 0.47 and 2.08 for B.1.420, and between 0.47 and 2.54 for B.1.1.348 (Supplementary Fig. 1). The values of R_e remained relatively constant, with an average value of around 0.92 and 1.33 for variants B.1 and B.1.1.348, respectively. In contrast, there were two peaks with values <1.5 for variants B.1.420, B.1.1.348, and B.1.111. After their emergence, the population size (N_e) rapidly increased for all these variants. N_e reached a plateau for all variants and remained constant for variants B.1, B.111, and B.1.420. Still, it decreased for variant B.1.1.348 after experiencing some oscillations (Supplementary Fig. 2). During this period, the R_e trend was congruent with the multinomial fit analysis (Supplementary Fig. 3e). We observed a significant ($p < 0.05$) positive correlation between R_e changes and the stringency index for the variants B.1.111 (Table 3) and a positive correlation between differences in the effective population sizes (N_e) and mobility changes for the variants B.1, B.1.111, B.1.1.348, and B.1.420 (Table 4).

Second period of the pandemic: from 10 August 2020 to 6 March 2021. By 10 August, there was promoted a COVID-19 testing program, contact tracing, and sustainable selective isolation[53]. At the same time, another COVID-19 measures were relaxed across Colombia (which included lifting mobility restrictions and opening domestic travel), the country experienced a new surge of cases, 2,499,104 cases were reported during this wave, 668 (0.034%) were sequenced, and 51 variants were identified. The variants B.1.111, B.1, B.1.1.348, and B.1.420 continue to predominate during this period and represented respectively 21.4%, 17.5%, 14.8% and 11.3% of the cases. Values of R_e ranged between 0.21 and 1.37 for B.1.111, between 0.39 and 2.04 for B.1, between 0.31 and 2.01 for B.1.1.348, and between 0.47 and 2.54 for B.1.420 (Supplementary Fig. 1). The values of R_e remained relatively constant with an average value of around 0.91, 1.08, and 1.08 for variants B.1.111, B.1, and B.1.1.348, respectively. The multinomial fit analysis showed that the highest R_e values for this period were for B.1 and B.1.1.348 (Supplementary Fig. 3f). During this period, the Technical Advisory
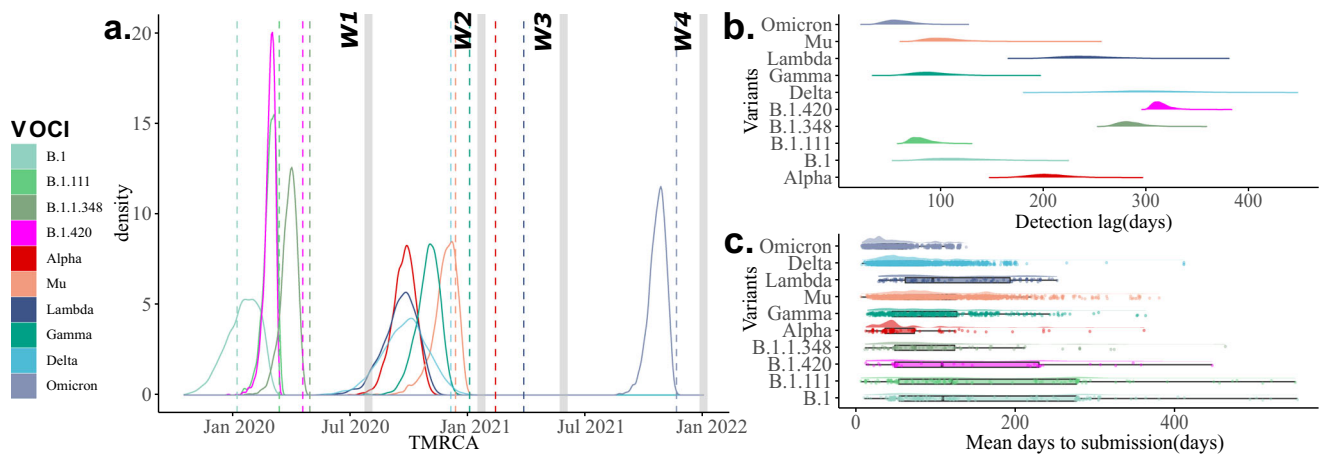
Fig. 3 Temporal dynamics of the most prevalent variants. a Time of the most common ancestor (TMRCA) of the most prevalent Pangolin COVID-19 global lineages, SARS-CoV-2 variants of interest and SARS-CoV-2 variant of concern (VOCIs) that have circulated in Colombia. Horizontal gray bars represent the four waves (W1, 2, 3, 4). The dotted line indicated the first Colombian report for each variant. **b** Detection lag over time as a function of TMRCA for each variant given by days. **c** Mean days to submission per variant and variant as a function of submission date versus collection date.

Table 3 Relationships between possible predictors of the effective reproductive number of the most predominant variants of COVID-19.

Lineage	Mobility	Stringency index	Vaccinated	Mask use	Cases
B.1.1	0.03 (0.05)	−0.003 (0.39)	NA	−0.01 (0.64)	−0.01 (0.92)
B.1.111	−0.01 (0.63)	0.4 (1.1e−08)*	NA	0.53 (3.22e−12)*	−0.008 (0.49)
B.1.1.348	0.01 (0.15)	−0.01 (0.70)	NA	0.2 (8.88e−05)*	0.28 (1.98e−05)*
B.1.420	0.20 (0.04)*	−0.02 (0.44)	NA	−0.04 (0.56)	0.53 (0.0007)*
B.1.617 (Delta)	0.54 (<2.2e−11)*	0.40 (<3.82e−08)*	0.36 (<2.62e−05)*	0.47 (<1.67e−09)*	−0.01 (0.91)
P.1 (Gamma)	0.51 (8.17e−09)*	0.50 (1.33e−08)*	0.57 (2.44e−09)*	0.50 (1.5e−08)*	0.09 (0.02)*
B.1.1.7 (Alpha)	−0.001 (0.33)	−0.01 (0.51)	−0.02 (0.54)	0.25 (0.002)*	0.15 (0.01)*
B.1.621 (Mu)	0.49 (2.21e−08)*	0.25 (0.0001)*	0.49 (1.80e−08)*	0.46 (8.92e−08)*	0.01 (0.21)
C.37 (Lambda)	−0.04 (0.84)	0.03 (0.21)	0.62 (0.0002)*	0.15 (0.03)*	0.75 (1.002e−07)*
Omicron	0.1 (0.14)	0.48 (0.006)*	0.39 (0.017)*	0.16 (0.101)*	0.79 (5.89e−05)*

We tested five predictors using a linear regression between a response variable (Re of each variant) and one variable or predictor (explanatory variables). The values in the table show the adjusted R-squared and (p-value). *: statistically significant. NA means that the variable was not recorded for all period of time that a particular variant circulated.

Table 4 Relationships between possible predictors of the effective population sizes of the most predominant variants of COVID-19.

Lineage	Mobility	Stringency index	Vaccinated	Mask use	Cases
B.1.1	0.06 (0.009)*	0.35 (6.64e−11)*	NA	0.82 (<2.2e−16)*	0.21 (1.07e−06)*
B.1.111	0.75 (<2.2e−16)*	0.17 (1.02e−05)*	NA	0.21 (9.996e−07)*	0.61 (<2.2e−16)*
B.1.1.348	0.37 (9.93e−12)*	0.20 (1.96e−06)*	NA	0.02 (0.08)	0.11(0.0003)*
B.1.420	0.72 (<2.2e−16)*	0.33 (1.65e−10)*	NA	0.27 (1.006e−08)*	0.75 (<2.2e−16)*
B.1.617+AY (Delta)	0.81 (<2.2e−16)*	0.60 (<2.2e−16)*	0.67 (<2.2e−16)*	0.76 (<2.2e−16)*	0.10 (0.0004)*
P.1 (Gamma)	0.51 (<2.2e−16)*	0.54 (<2.2e−16)*	0.60 (<2.2e−16)*	0.49 (<2.2e−16)*	0.03 (0.03)*
B.1.621 (Mu)	−0.0006 (0.33)	0.01 (0.15)	−0.009 (0.84)	−0.008 (0.65)	0.28 (7.907e−09)*
B.1.1.7 (Alpha)	0.10 (0.0007)*	0.09 (0.001)*	0.08 (0.007)*	−0.01 (0.92)	0.002 (0.25)
C.37 (Lambda)	−0.007 (0.60)	0.21 (6.51e−07)*	0.48 (1.58e−13)*	0.41 (2.26e−13)*	0.05 (0.01)*
Omicron	0.11 (0.0004)*	0.14 (4.39e−05)*	0.87 (<2.2e−16)*	0.61 (<2.2e−16)*	0.11 (0.0004)*

We tested five predictors using a linear regression between a response variable (N_e of each variant) and one variable or predictor (explanatory variables). The values in the table show the adjusted R-squared and (p-value). *: statistically significant. NA means that the variable was not recorded for all period of time that a particular variant circulated.

Group on Virus Evolution from World Health Organization (WHO) introduced the notion of "Variants of Concern" (VOC) and "Variants of Interest" (VOI)[13]. During this period in Colombia, three variants of concern were reported, Gamma (5% of the cases), Delta (0.5%), and Alpha (0.59%), and one variant of interest, Mu (2.9%). All these variants co-circulated after their

emergence, and the most recent common ancestors of the variants documented in Colombia were dated to 23 August 2020 (CI: 18 July 2020–02 October 2020), 01 December 2020 (CI: 25 September 2020–06 December 2020), 09 December 2020 (CI: 27 October 2020–09 January 2021), and 30 September 2020 (CI: 27 August 2020–11 October 2020) for Delta, Gamma, Mu, and
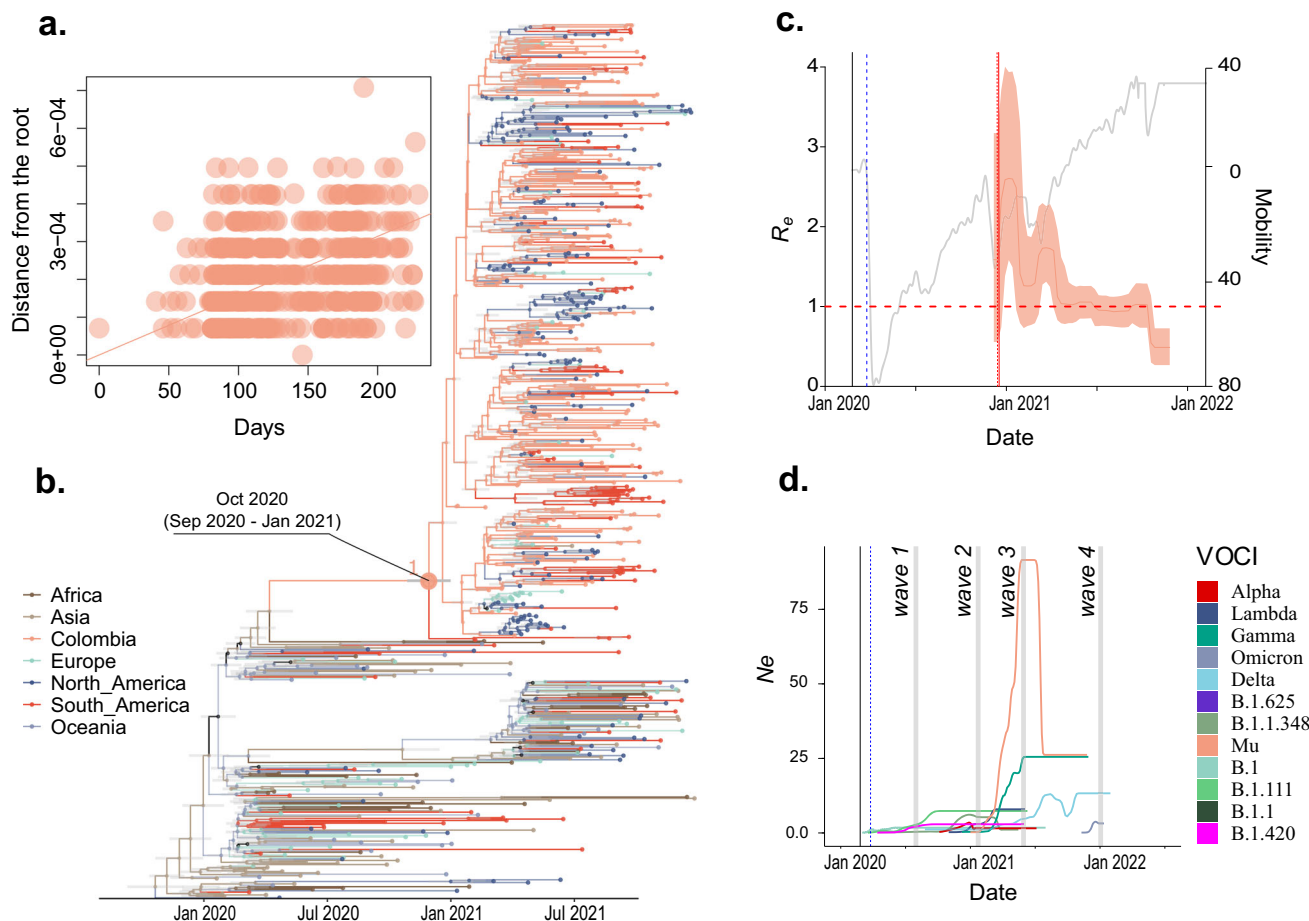
Fig. 4 Phylodynamic of Mu variant population circulated in Colombia. a Clock signal in Colombian sequences of the Mu variant. **b** MCC tree pointing out the Mu variant origin. **c** Effective reproductive number (R_e) of Mu and **d** Effective population size (N_e) of all variants in Colombia over time. The phylogenetic tree was built with all the major Global Outbreak variants with the most representation on B.1.621 lineage (Mu variant). Branches' colors represent Global regions and Colombia (country).

Alpha, respectively (Fig. 3). Based on the global available sequences, Colombia is the more probable geographical origin for Mu variant (Fig. 4b).

Although the Lambda variant was not reported during this period, the time to the most recent common ancestor (TMRCA) was estimated around 15 December 2020 (CI: 22 October 2020–01 February 2021) (Fig. 3). The values of R_e ranged between 0.45 and 2.68 for Delta, between 0.40 and 2.78 for Gamma, between 0.73 and 4.0 for Mu, and between 0.28 and 2.11 for Alpha (Supplementary Fig. 1). The values of R_e remained relatively constant with an average value of around 0.91, 1.08, and 1.08 for variants B.1.111, B.1, and B.1.1.348, respectively. Concerning the VOCIs circulated during this period, values of R_e ranged between 0.45 and 2.68 for Delta, between 0.40 and 2.78 for Gamma, between 0.73 and 4.0 for Mu, and between 0.28 and 2.11 for Alpha (Supplementary Fig. 1). The values of R_e remained relatively constant with an average value of around 1.29, 1.24, 1.8, and 1.20 for variants Delta, Gamma, Mu, and Alpha, respectively. The population size (N_e) showed a clear upward trend before the highest peak during this period for Lambda and Alpha variants after their emergence. This was followed by a plateau and remained constant for both variants. Although Gamma, Mu, and Delta were reported during this period, N_e reached low values and increased rapidly before waves 3 and 4, respectively (Supplementary Fig. 2). We observed a significant (<0.05) positive correlation between R_e changes and three or four predictors evaluated (mobility, vaccinated people, and stringency

index for Delta, Gamma and Mu; and mask use for Alpha as well). However, the highest R-squared values were for mobility ($p > 0.5$) for Delta, Gamma, and Alpha variants (Table 3). There is a positive correlation between differences in the effective population sizes (N_e) and mobility changes for the variants Delta, Gamma, and Alpha (Table 4).

Third period of the pandemic: from 7 March 2021 to July 2021. During the third wave, Colombia reported 1,797,454 cases, higher than previous waves. 2984 samples (0.16%) were sequenced, and 44 variants were identified. The most prevalent variant during this time was Mu, with 48.42% of the cases. Gamma, Lambda, and Alpha were predominated representing (21.41%), (4.59%), and (4.79% of the cases), respectively. Delta was documented in low proportion (0.33%). Values of R_e ranged between 0.64 and 2.76 for Gamma, between 0.59 and 2.86 for Lambda, between 0.25 and 1.60 for Alpha, between 0.75 and 2.28 for Mu, and between 0.56 and 2.32 for Delta (Supplementary Fig. 1). The values of R_e remained relatively constant with an average value of around 1.52, 1.23, 0.82, 1.17, and 1.37 for all variants previously mentioned. The highest peaks were for Mu and Delta, and this trend was congruent between Phylodynamics and multinomial fit analyses (Supplementary Fig. 1; 3g). We observed the highest peak of the population size (N_e) for Mu (Fig. 4c) and Gamma variants after their emergence. This was followed by a relevant decrease and remained constant for both variants. Concerning to Delta variant,

N_e increased rapidly after waves 3 and 4 (Supplementary Fig. 2). The growth advantage dynamic values evidenced that Mu variant had an advantage over Lambda, Gamma, Alpha, and Delta during wave 3 (Supplementary Fig. 3c). We observed a significant ($p < 0.05$) positive correlation between R_e changes and four predictors evaluated (mobility, stringency index, vaccinated people, and mask use) for the Mu variant and vaccinated people for the Lambda variant, being this predictor the highest R-squared values with 0.49 and 0.62, respectively. There is a positive correlation between differences in the effective population sizes (N_e) and the number of case changes for the Mu variant. Still, there was no significant association between R_e and mobility changes for the Mu variant during this wave (Tables 3 and 4).

Fourth period of the pandemic: from August 2021 to February 2022. This period was characterized by a relevant decrease at the end of 2021 and an exponential increase in the number of cases called wave 4. During this period was, reported 1,769,695 cases, 9946 samples (0.56%) were sequenced, and 41 variants were identified, including 83 Delta AY.x sub-variants and two Omicron, Mu, and Lambda sub-variants. As of February 2022, Omicron is the last variant of concern globally reported. In Colombia, Omicron displaced the previous variants described with a predominance of 87.6% in two first months in 2022 and has been detected in all the 32 states of Colombia. Since its identification, its prevalence has been 12% of the total variant samples identified in Colombia. The estimation of the TMRCA suggested that Omicron was introduced on 24 October 2021 (CI: 25 September 2021–15 November 2021) (Fig. 3), which is congruent with the first epidemiological report. Delta prevalence was 5.3%, and other variants were 7% during the same period. Values of R_e ranged between 0.68 and 1.79 for Delta, between 0.26 and 1.29 for Mu, and between 0.45 and 3.46 for Omicron (Supplementary Fig. 1). The highest peak of R_e was for Omicron, which was congruent with both performed analyses (Supplementary Fig. 1; 3h). N_e values increased rapidly before the highest peak of wave four, followed by a plateau, while N_e values maintained constant for Delta and Mu variants for this last period (Supplementary Fig. 2). The highest average value of R_e was 1.47 for Omicron compared to 1.09 and 0.82 for Delta and Mu, respectively, during this period. The growth advantage dynamic values showed the same trend as the R_e values, which omicron variant evidenced an advantage compared to Delta and Mu variant during wave 4 (Supplementary Fig. 3d, h). A significant ($p < 0.05$) positive correlation between R_e changes and the number of cases was significant, while between N_e changes and all five predictors evaluated were significant, with the highest R-squared for vaccinated people with 0.61 (Tables 3 and 4).

Discussion

The present study provides a comprehensive description of the emergence and dynamic of SARS-CoV-2 variants in Colombia based on genomic surveillance and a phylodynamic approach. Variant diversity in Colombia was characterized by multiple SARS-CoV-2 variants and multiple introductions derived from ancestral B.1 lineage, which was imported mainly from European countries (Spain and Italy)[54,55]. Despite documenting at least 188 lineages in the community, only ten dominated, suggesting high transmissibility of those variants of interest and concern compared with other emerged variants. Thus, we were interested in comparing transmission differences between Colombia's main circulated SARS-CoV-2 variants.

We employed Bayesian phylodynamic methods to recover Colombia's most prevalent genetic variants following the first reported case in March 2020[1], which led to COVID-19 being declared a health emergency one week later. Different control strategies have been implemented since then, such as mandatory isolation, epidemiological follow-up for air passengers who arrived in Colombia with COVID-19 symptoms, the closing of borders, and international air travel being banned on 20 March 2020[56]. Contention measures were taken on 25 March when lockdown and domestic air-travel ban was decreed, except for essential workers (such as bank tellers, post officers, and healthcare professionals)[57].

Despite the contention measures and an observed 80% reduction in mobility, cases continued to surge throughout the country (Fig. 1). The effective reproductive number (R_e) was shown with particular fluctuation through time per each variant circulating in the country. All analyzed variants recovered peaks >1, suggesting that a variant's spread was greater than another during a specific period (Supplementary Fig. 4). These values are congruent with previous studies for the two first periods, with medians from 1.07 to 2.13 for the first period and from 0.99 to 1.09 for the second one, respectively[4]. Comparing variants, Mu showed the highest R_e values indicating higher transmissibility than Alpha, Gamma, Lambda, and even Delta, which had been reported as dominant variants in the countries they have circulated[58]. This suggests that once Mu emerged in Colombia, it out-competed the other variants and became the dominant one. In late 2021, Mu was eventually out-competed by Delta.

Differences in transmissibility between variants could be explained by partial immune evasion. It has been reported that VOCIs that carry K417N, E484K, and N501Y have a higher affinity towards the hACE2 receptor and enhanced immune escape abilities as observed with Gamma and Alpha in Brazil and the United Kingdom, respectively[59,60]. Late 2020 and early 2021 were characterized by the emergence of variants exhibiting advantage-conferring mutations, and despite Alpha's increased transmissibility and innate immune escape ability (represented in mutations N501Y and Δ69-70)[17,60] it did not manage to establish as the dominant variant after its introduction around 26 November 2020 and was shortly displaced by Gamma and Mu. A different scenario occurred with highly evasive variants such as Gamma, Lambda, and Mu which dominated the transmission dynamics during Colombia's third pandemic wave.

Gamma was first detected in Manaus, a city in the Brazilian Amazon state, and has been determined to have emerged around November of 2020 as a result of an accelerated evolutionary rate of locally circulating clades. Due to its increased viral load, it rapidly spread throughout Brazil[59,61]. This variant was first detected in Colombia on 4 January 2021. Based on TMRCA estimation, we suggest that it could have been circulating in the Colombian Amazonian region by December of 2020 (Fig. 3) before its introduction into 29 states.

However, the emergence of the Mu variant in Colombia caused a displacement of other variants whose circulation had previously been characterized by geographic heterogeneity, with the Pacific region (Valle del Cauca, Cauca, Nariño, and Choco states) being dominated by Lambda. In contrast, Andean states (Huila, Risaralda, Quindio, Tolima, Cundinamarca, Boyaca, Santander, Norte de Santander and Antioquia) and Amazonian states (Putumayo, Caqueta, Guaviare, Guania, Vaupes and Amazonas) had a high circulation of Gamma (Fig. 2). Our findings can be explained by Mu's high R_e and its partial immune escape. Mu variant is 10.6 and 9.1 times more resistant to convalescent, and BNT162b2-immunized patient sera[62,63]. Previous studies on the impact of enhanced transmissibility and partial variant immune escape have demonstrated that epidemic sizes become larger after the introduction of a highly transmissible and immune-evasive variant. It happened commonly in a scenario comprised of slow vaccine rollout and depletion of NPIs. Furthermore, the partial

Table 5 Posterior summary of the time to the most recent common ancestor (TMRCA) showing as Tree Height parameter per variant estimated using the Birth-Death Skyline model (BDSKY), Coalescent Skyline model (SKY), and Bayesian Integrated Coalescent Epoch PlotS (BICEPS).

Variant	BDSKY	95% HPD	BICEPS	95% HPD	SKY	95% HPD
B.1.1	1.55	1.47-1.66	1.54	1.42-1.68	1.50	1.41-1.62
B.1.111	1.28	1.26-1.31	1.28	1.24-1.33	1.26	1.23-1.30
B.1.1.348	1.11	1.07-1.16	1.13	1.07-1.20	1.11	1.06-1.18
B.1.420	1.09	1.06-1.12	1.16	1.12-1.20	1.13	1.12-1.15
B.1.1.7 (Alpha)	0.87	0.80-0.94	0.85	0.75-0.95	0.83	0.75-0.92
P.1 (Gamma)	1.07	0.9-1.2	1.07	0.93-1.39	1.03	0.9-1.13
B.1.617+AY (Delta)	1.68	1.51-1.86	1.63	1.43-1.83	1.63	1.43-1.85
B.1.621 (Mu)	0.72	0.62-0.90	0.66	0.62-0.73	0.70	0.62-0.80
B.1.1.7 (Alpha)	0.52	0.45-0.60	0.48	0.39-0.57	0.50	0.41-0.60
C.37 (Lambda)	0.70	0.60-0.80	0.64	0.52-0.77	0.68	0.56-0.82
Omicron	0.70	0.60-0.80	0.64	0.52-0.77	0.68	0.56-0.82

Each value is numerical as a year.

immune evasion of Mu could account for reinfections and breakthroughs among previously highly immune populations[64]. These data support that Mu's higher R_e as described in our study (Supplementary Fig. 1) and its ability to partially escape antibody-mediated neutralization might account for Colombia's third wave of COVID-19 cases[65].

On the other hand, our results suggest the opposite phenomenon occurred with Delta. Once it was introduced to Colombia on 3 April 2021, it remained undetected until 10 May 2021, coinciding with Mu's establishment and expansion. Delta prevalence increased after July of 2021 with a steady increment in the share of reported variants (Supplementary Fig. 3c). However, cases remained low throughout July until November 2021. We propose it might be due to Delta circulating in a population with a high level of immunity elicited both by vaccination and previous exposure to Mu, which has been found to cross-neutralize Delta[66] effectively. Despite a high Delta's R_0[67], our findings show that its circulation in Colombia did not cause an exponential surge in cases, as reflected by its R_e and N_e.

In contrast, we found the Omicron variant could be responsible for the surge in cases observed through the fourth wave after its introduction into Colombia on 24 October 2021. This is based on a marked increase of N_e and a steady R_e over 1, displacing Delta circulation in the country (Supplementary Fig. 3). Our results support the predicted scenarios of introducing a highly immune evasive and highly transmissible variant in a population with high levels of immunity, with an observed out-competing of variants with high transmissibility but mild immune escape such as Delta[64,68]. The probable causes for the steep rise in Omicron's prevalence are the control measures weakness and the 1.4-fold augment in mobility (Supplementary Fig. 2b). Although a positive correlation between mobility and Omicron's N_e and R_e was observed, suggesting Omicron's advantages (immune escape), it was not statistically significant. Compared with the Mu variant, the impact of Omicron on public health was considerably lower, which could be explained by Colombia's higher vaccine coverage by the end of 2021 (62%). Therefore, even though some studies have found that population immunity wanes through time either by previous infection and vaccination and confers mild protection against reinfection and breakthrough cases by Omicron variant. Vaccination continues to effectively reduce the risk of severe disease and death as found with previous variants[8,69].

The effective population sizes (N_e) estimations increment of each variant precede an increment in the number of cases, followed by extensive of community transmission. The oscillations in N_e and R_e could be explained by the fluctuations in mobility

and preventive and control measures applied after each reported wave. As an exploratory data analysis, a general linear regression model was evaluated to identify which actions could effectively control the variant transmission represented by R_e and viral population growth. We evaluated four different control strategy measures, and two variables can explain N_e and R_e values (Tables 2 and 3). In most cases, mobility showed higher values of R squared with considerable values, suggesting that it affects N_e and R_e. Mobility indicates the more meaningful potential for personal contact, which can contribute to the spread of the disease. When mobility is high, the risk of COVID-19 spread may also be increased[70,71]. However, as mobility increases, taking precautions such as getting vaccinated, Colombian COVID-19 responses such as the implementation of stringent government policies (school closures, workplace closures, cancellation of public events, restriction on public gatherings, closures of public transport, stay at home requirements, information campaigns, restrictions on international movements; and international controls), and wearing masks in public areas can all reduce the risk of disease transmission. This regression analysis had limitations, such as a small number of data points. It was performed assuming a Re value per week; there were more than 20 points in all cases. In the future, it is necessary to include available seroprevalence data[72,73], also, perform a GLM including all the proposed explanatory variants into the model and assuming more epochs. Due to the low sequencing intensity during 2020, the estimations of lineage proportion and multiplicative effect on R_e obtained by fitting Multinomial logistic models resulted in wide CIs and were therefore less accurate than models for waves 3 and 4. However, this analyses are important for recovering and understanding the growth rate advantages of the variants that dominated the first year of the pandemic in Colombia.

The estimated mean time to the most recent common ancestor of the viral population for Alpha, Lambda and B.1.1.348 lineages detected in Colombia was before their first detection in the country (Fig. 3). Although this could indicate that lineages were circulating long before being identified in Colombia, there is not enough evidence to fully support this claim because there have been multiple introductions of these lineages into the country. Thus this common ancestor may have existed outside of the country.

The applicability of Bayesian phylodynamic methods is limited considering large genomic datasets, such as that of SARS-CoV-2. We employed down-sampling strategies to address these complications, allowing us to use a representative sample of both time and geography. Furthermore, we used a novel Bayesian Integrated

Coalescent Epoch PlotS (BICEPS) for efficient inference of coalescent epoch models. It integrates population size parameters and introduces a set of more powerful Markov Chain Monte Carlo (MCMC) proposals for flexing and stretching trees[41]. The present work compared the traditional Bayesian skyline model with this novel model and found congruence in effective population size estimates (Supplementary Fig. 2) and TMRCA per variant estimation (Table 5) between methods. The novel implementation of tree priors and proposals allows larger genomic datasets to be analyzed for tracing an emerging virus's spread, transmission, and population dynamics for genetic surveillance reports.

In summary, the study highlights the dynamics of the most predominant genetic variants that have been reported in Colombia in terms of transmissibility and demographic dynamic. The high transmission and effective population sizes of each variant could be explained by the increase in mobility and the reduction in the government response tracker (implementation of control measures) in Colombia. Each wave was characterized by the circulation of at least one of these prevalent variants. The emergence of the highly transmissible Mu in Colombia could explain why Delta and Alpha, which were introduced previously, did not have the same impact as in other countries such as England or Brazil. Genomic surveillance has been instrumental in informing public health response against COVID-19 in many parts of the world, including New Zealand, Australia, Iceland, and Taiwan, showing how these implementations helped to successfully control the increase of COVID-19[23]. This is made accessible by pathogen surveillance platforms such as GISAID[74], NextStrain[39], and Microreact[75]. Here, we have demonstrated how these technologies can inform public health response in Colombia. We advocate for the widespread adoption of such technologies in the Colombian public health infrastructure and worldwide.

Data availability

Accession codes for all sequencing data utilized in this study, as well as any other raw datasets, are available in the supplementary material accompanying this paper. Additionally, all source data for the figures presented in the main manuscript, including the numerical results underlying the graphs and charts, are provided as supplementary files in a machine-readable format. These source data files can also be accessed online at (https://github.com/cinthylorein/Colombia-COVID-19-phylodynamics.git/). Researchers and readers interested in accessing the data are encouraged to refer to the supplementary material for detailed instructions on data retrieval and utilization. For any further inquiries or assistance, the corresponding author can be contacted via email. Please note that access to certain datasets or raw data may require approval from the appropriate ethics committee and adherence to relevant data sharing agreements.

Code availability

The code used in this study is available on GitHub at the following repository: (https://github.com/cinthylorein/Colombia-COVID-19-phylodynamics.git/). The repository has been archived and made citeable via Zenodo, with a DOI assigned (https://zenodo.org/badge/latestdoi/416098836). Researchers can access the code by visiting the GitHub repository or by citing the Zenodo DOI in the references section. We encourage researchers and readers interested in utilizing the code to visit the GitHub repository for detailed instructions on code retrieval, installation, and usage. For any inquiries or support related to the code, please contact the corresponding author via email.

References

1. Minsalud. Colombia confirma su primer caso de COVID-19 https://www.minsalud.gov.co/Paginas/Colombia-confirma-su-primer-caso-de-COVID-19.aspx (2020).
2. Worldometer. COVID-19 coronavirus pandemic https://www.worldometers.info/coronavirus/ (2020).
3. Ritchie, H. et al. Coronavirus pandemic (covid-19). Our World in Data https://ourworldindata.org/coronavirus (2020).
4. Castañeda, S. et al. Evolution and epidemic spread of sars-cov-2 in Colombia: a year into the pandemic. Vaccines https://www.mdpi.com/2076-393X/9/8/837 (2021).
5. Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on Covid-19 in Europe. Nature 584, 257–261 (2020).
6. Jaja, I. F., Anyanwu, M. U. & Jaja, C.-J. I. Social distancing: how religion, culture and burial ceremony undermine the effort to curb Covid-19 in South Africa. Emerg. Microbes Infect. 9, 1077–1079 (2020).
7. Dreher, N. et al. Policy interventions, social distancing, and sars-cov-2 transmission in the United States: a retrospective state-level analysis. Am. J. Med. Sci. 361, 575–584 (2021).
8. Haug, N. et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat. Hum. Behav. 4, 1303–1312 (2020).
9. Islam, S., Islam, T. & Islam, M. R. New coronavirus variants are creating more challenges to global healthcare system: a brief report on the current knowledge. Clin. Pathol. 15, 2632010X221075584 (2022).
10. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407 (2020).
11. Bernal, J. L. et al. Effectiveness of covid-19 vaccines against the b. 1.617. 2 (delta) variant. N. Engl. J. Med. 385, 585–594 (2021).
12. Li, L. et al. Transmission and containment of the sars-cov-2 delta variant of concern in Guangzhou, China: a population-based study. PLoS Negl. Trop. Dis. 16, e0010048 (2022).
13. Organization, W. H. Tracking SARS-CoV-2 variants. https://www.who.int/activities/tracking-SARS-CoV-2-variants.
14. Janik, E., Niemcewicz, M., Podogrocki, M., Majsterek, I. & Bijak, M. The emerging concern and interest sars-cov-2 variants https://www.mdpi.com/2076-0817/10/6/633/htm, https://www.mdpi.com/2076-0817/10/6/633 (2021).
15. Resende, P. C. et al. Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B.1.1.33 during the early pandemic phase in Brazil. Front. Microbiol. 11, 3565 (2021).
16. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell 184, 2372–2383.e9 (2021).
17. Volz, E. et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 184, 64–75.e11 (2021).
18. Konings, F. et al. SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse https://www.nature.com/articles/s41564-021-00932-w (2021).
19. Campbell, F. et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Eurosurveillance 26, 2100509 (2021).
20. Laiton-Donato, K. et al. Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. Infect. Genet. Evol. 95, 105038 (2021).
21. Viana, R. et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in Southern Africa. Nature 603, 679–686 (2022).
22. Geoghegan, J. et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. Nat. Commun. 11, 6351 (2020).
23. Douglas, J. et al. Phylodynamics reveals the role of human travel and contact tracing in controlling the first wave of covid-19 in four island nations. Virus Evol. 7, veab052 (2021).
24. Corman, V. M. et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 25, 1–8 (2020).
25. Quick, J. ncov-2019 sequencing protocol v3 (locost) v.3. https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bp2l6n26rgqe/v3.
26. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for oxford nanopore sequencing. Genome Biol. 20, 1–10 (2019).
27. Loman, N., Rowe, W., Rambaut, A. ncov-2019 sequencing protocol v3 (locost) v.3. https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html.
28. Miller, J. M. et al. Guidelines for Safe Work Practices in Human and Animal Medical Diagnostic Laboratories Recommendations of a CDC-convened, Biosafety Blue Ribbon Panel Centers for Disease Control and Prevention MMWR Editorial and Production Staff MMWR Editorial Board. Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report 61, 105 (2012).
29. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J. Am. College Dentists 81, 14–18 (2014).
30. O'Toole, Á. et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. https://academic.oup.com/ve/article/7/2/veab064/6315289 (2021).
31. Aksamentov, I., Roemer, C., Hodcroft, E. B. & Neher, R. A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J. Open Source Softw. 6, 3773 (2021).
32. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121–4123 (2018).

33. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014).
34. Davies, N. G. et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021).
35. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (Covid-19) infections. *Int. J. Infect. Dis.* **93**, 284–286 (2020).
36. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. In *Bioinformatics for DNA sequence analysis*, 39–64 (Springer, 2009).
37. Paradis, E. et al. Package 'ape'. *Analyses of phylogenetics and evolution, version 2* (2019).
38. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using tempest (formerly path-o-gen). *Virus Evol.* **2**, vew007 (2016).
39. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
40. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
41. Bouckaert, R. R. An efficient coalescent epoch model for bayesian phylogenetic inference. *Syst. Biol.* **71**, 1549–1560 (2022).
42. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in hiv and hepatitis c virus (hcv). *Proc. Natil Acad. Sci. USA* **110**, 228–233 (2013).
43. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis c virus (HCV). *Proc. Natl Acad. Sci. USA* **110**, 228–233 (2013).
44. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comp. Biol.* **5**, e1000520 (2009).
45. Hasegawa, M., Kishino, H. & aki Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
46. Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 42 (2017).
47. Müller, N. F. & Bouckaert, R. R. Adaptive metropolis-coupled mcmc for beast 2. *PeerJ* **8**, e9473 (2020).
48. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
49. Yu, G. Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinformatics* **69**, e96 (2020).
50. for Health Metrics, I. & (IHME), E. Covid-19 projections. https://covid19.healthdata.org/colombia?view=cumulative-deaths&tab=trend (2022).
51. (OxCGRT), O. C.-. G. R. T. Covid-19 government response tracker. https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker (2022).
52. Instituto nacional de salud (ins). https://www.ins.gov.co/Noticias/Paginas/coronavirus-genoma.aspx (2022).
53. Haug, N. et al. Ranking the effectiveness of worldwide covid-19 government interventions. *Nat. Hum. Behav.* **4**, 1303–1312 (2020).
54. Laiton-Donato, K. et al. Genomic epidemiology of sars-cov-2 in Colombia. *medRxiv*.06 (2020).
55. Ramírez, J. D. et al. The arrival and spread of SARS-CoV-2 in Colombia. *J. Med. Virol.* https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.26393 (2020).
56. Ministerio de Salud y Protección Social. Resolución 00000380 del Ministerio de Salud y Protección Social 5 https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/resolucion-380-de-2020.pdf (2020).
57. Ministerio del Interior. Republica de Colombia. Decreto Número 457. Decreto Número 457 1–14 https://dapre.presidencia.gov.co/normativa/normativa/DECRETO457DEL22DEMARZODE2020.pdf (2020).
58. Petrone, M. E. et al. Insights into the limited global spread of the immune evasive sars-cov-2 variant Mu. *medRxiv* 2022-03 (2022).
59. Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372** (2021).
60. du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
61. Naveca, F. G. et al. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nat. Med.* **27**, 1230–1238 (2021).
62. Uriu, K. et al. Neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine serum. *N. Engl. J. Med.* **385**, 2397–2399 (2021).
63. Tada, T. et al. High-titer neutralization of Mu and C.1.2 SARS-CoV-2 variants by vaccine-elicited antibodies of previously infected individuals. *Cell Rep.* **38**, 110237 (2022).
64. Bushman, M., Kahn, R., Taylor, B. P., Lipsitch, M. & Hanage, W. P. Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape. *Cell* **184**, 6229–6242.e18 (2021).
65. Ramirez, A. L. et al. Impact of SARS-CoV-2 Mu variant on vaccine effectiveness: a comparative genomics study at the peak of the third wave in Bogota, Colombia. *J. Med. Virol.* https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.27808, https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.27808, https://onlinelibrary.wiley.com/doi/10.1002/jmv.27808 (2022).
66. Uriu, K. et al. Characterization of the immune resistance of severe acute respiratory syndrome coronavirus 2 Mu variant and the robust immunity induced by Mu infection. *J. Infect. Dis.* https://academic.oup.com/jid/advance-article/doi/10.1093/infdis/jiac053/6530359 (2022).
67. Liu, Y. & Rocklöv, J. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *J. Travel Med.* https://academic.oup.com/jtm/article/28/7/taab124/6346388 (2021).
68. Cai, J. et al. Modeling transmission of SARS-CoV-2 Omicron in China. *Nat. Med.* https://www.nature.com/articles/s41591-022-01855-7, http://www.ncbi.nlm.nih.gov/pubmed/35537471 (2022).
69. Serrano-Coll, H. et al. Effectiveness of the CoronaVac® vaccine in a region of the Colombian Amazon, was herd immunity achieved? *Trop. Dis. Travel Med. Vaccines* **8**, 4–9 (2022).
70. Mattar, S. et al. Severe Acute Respiratory Syndrome Coronavirus 2 Seroprevalence among adults in a tropical city of the Caribbean Area, Colombia: are we closer to herd immunity than developed countries? *Open Forum Infect. Dis.* https://academic.oup.com/ofid/article/7/12/ofaa550/5977863 (2020).
71. Laajaj, R. et al. COVID-19 spread, detection, and dynamics in Bogota, Colombia. *Nat. Commun.* https://doi.org/10.1038/s41467-021-25038-z (2021).
72. Mercado-Reyes, M. et al. Seroprevalence of anti-sars-cov-2 antibodies in Colombia, 2020: a population-based study. *Lancet Regional Health - Americas* **9**, 100195 (2022).
73. Serrano-Coll, H. et al. High prevalence of sars-cov-2 in an indigenous community of the colombian amazon region. *Trop. Med. Infect. Dis.* https://www.mdpi.com/2414-6366/6/4/191 (2021).
74. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro. Surveill.* **22**, 30494 (2017).
75. Argimón, S. et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb. Genom.* **2** (2016).

Acknowledgements

Author contributions

C.J.S. and R.R. contributed to the design of the study, conceived and conducted the analysis, analyzed the results and writing of the manuscript; J.D. and R.B. contributed to analysis of the results and writing of the manuscript; R.R., J.D.R., B.G., A.C., C.G., D.E.D., M.M., S.C., and L.H.P. contributed to sample collection, viral detection, clinical characterization, genome isolation and sequencing; J.D.R., M.M., S.C., and L.H.P. contributed to genome assembly, analysis, and quality control; N.B., A.R., and N.L. contributed to sample sequencing and genome assembly; A.P.M. and H.S.C. contributed to the writing and critical review of the manuscript; C.J.S., R.R., J.D.R., C.J.V.A., K.E.A., S.M., and A.J.D. contributed to result discussion and reviewing of the manuscript. All the authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-023-00328-3.

Correspondence and requests for materials should be addressed to Cinthy Jimenez-Silva, Ricardo Rivero, Juan David Ramirez or Salim Mattar.

Peer review information *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.