VIRUS EVOLUTION

Virus Evolution, 2021, **7(2)**, 1–12

DOI: https://doi.org/10.1093/ve/veab087

Advance access publication date: 1 October 2021

Research Article

Epidemiological dynamics of SARS-CoV-2 VOC Gamma in Rio de Janeiro, Brazil

Filipe Romero Rebello Moreira,[1,†,§] Mirela D'arc,[2,†] Diana Mariani,[1] Alice Laschuk Herlinger,[1] Francine Bittencourt Schiffler,[2] Átila Duque Rossi,[1] Isabela de Carvalho Leitão,[3] Thamiris dos Santos Miranda,[2] Matheus Augusto Calvano Cosentino,[2] Marcelo Calado de Paula Tôrres,[1] Raíssa Mirella dos Santos Cunha da Costa,[1] Cássia Cristina Alves Gonçalves,[1] Débora Souza Faffe,[3] Rafael Mello Galliez,[4] Orlando da Costa Ferreira Junior,[1] Renato Santana Aguiar,[5,6,**] André Felipe Andrade dos Santos,[2] Carolina Moreira Voloch,[1,††] Terezinha Marta Pereira Pinto Castiñeiras,[4,‡] and Amilcar Tanuri[1,‡,*] on behalf of the COVID-19-UFRJ Workgroup

[1]Departamento de Genética, Laboratório de Virologia Molecular, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, Centro de Ciências da Saúde, Bloco A, lab 121, Cidade Universitária, Rio de Janeiro 21941-902, Brazil, [2]Departamento de Genética, Laboratório de Diversidade e Doenças Virais, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, Centro de Ciências da Saúde, Bloco A, lab 120, Cidade Universitária, Rio de Janeiro 21941-902, Brazil, [3]Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, Centro de Cincias da Saúde, Bloco C, Cidade Universitária, Rio de Janeiro 21941-902, Brazil, [4]Departamento de Doenças Infecciosas e Parasitárias, Universidade Federal do Rio de Janeiro, Av. Carlos Chagas Filho 373, Centro de Ciências da Saúde, Bloco K, Cidade Universitária, Rio de Janeiro 21941-902, Brazil, [5]Departamento de Genética, Ecologia e Evolução, Laboratório de Biologia Integrativa, Universidade Federal de Minas Gerais, Belo Horizonte, Av. Antônio Carlos, 6627, Instituto de Ciências Biológicas, G3-60, Pampulha, Belo Horizonte 31270901, Brazil and [6]Instituto D'Or de Pesquisa e Ensino (IDOR), Rio de Janeiro, Rua Diniz Cordeiro 30, Botafogo, Rio de Janeiro 22281-100, Brazil

[†]These authors equally contributed to this work.

[‡]These authors share the senior authorship.

[§]https://orcid.org/0000-0002-7162-5070

[**]https://orcid.org/0000-0001-5180-3717

[††]https://orcid.org/0000-0001-5182-4366

*Corresponding author: E-mail: atanuri@gmail.com

Abstract

The emergence and widespread circulation of severe acute respiratory syndrome coronavirus 2 variants of concern (VOCs) or interest impose an enhanced threat to global public health. In Brazil, one of the countries most severely impacted throughout the pandemic, a complex dynamics involving variants co-circulation and turnover events has been recorded with the emergence and spread of VOC Gamma in Manaus in late 2020. In this context, we present a genomic epidemiology investigation based on samples collected between December 2020 and May 2021 in the second major Brazilian metropolis, Rio de Janeiro. By sequencing 244 novel genomes through all epidemiological weeks in this period, we were able to document the introduction and rapid dissemination of VOC Gamma in the city, driving the rise of the third local epidemic wave. Molecular clock analysis indicates that this variant has circulated locally since the first weeks of 2021 and only 7 weeks were necessary for it to achieve a frequency above 70 per cent, consistent with rates of growth observed in Manaus and other states. Moreover, a Bayesian phylogeographic reconstruction indicates that VOC Gamma spread throughout Brazil between December 2020 and January 2021 and that it was introduced in Rio de Janeiro through at least 13 events coming from nearly all regions of the country. Comparative analysis of reverse transcription-quantitative polymerase chain reaction (RT-qPCR) cycle threshold (Ct) values provides further evidence that VOC Gamma induces higher viral loads (N1 target; mean reduction of Ct: 2.7, 95 per cent confidence interval = ± 0.7). This analysis corroborates the previously proposed mechanistic basis for this variant-enhanced transmissibility and distinguished epidemiological behavior. Our results document the evolution of VOC Gamma and provide independent assessment of scenarios previously studied in Manaus, therefore contributing to the better understanding of the epidemiological dynamics currently being surveyed in other Brazilian regions.

Key words: SARS-CoV-2; COVID-19; variant of concern Gamma; lineage P.1; Rio de Janeiro; Brazil; phylodynamics; phylogeography

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in China in late 2019 and rapidly spread around the globe, leading the World Health Organization (WHO) to declare a pandemic state on 11 March 2020 (Wu et al., 2020; WHO 2020). In Brazil, the first report of SARS-CoV-2 infection occurred in late February 2020 in the state of São Paulo (de Jesus et al., 2020). Soon after that, multiple introductions from diverse locations and lineages were documented in the country (Candido et al., 2020). These events led to transmission chains that perpetuated and

caused a heavy burden on Brazilian public health, accounting for more than 18 million cases and 500,000 deaths reported (as of 23 June 2021; available in https://covid.saude.gov.br//).

The predominance of two major lineages, B.1.1.28 and B.1.1.33, characterized the first SARS-CoV-2 epidemic wave in Brazil (Candido et al., 2020). This scenario started to change in the second semester of 2020, with the worldwide emergence of variants of concern (VOCs). VOCs and variants of interest (VOIs) bear mutations with biological significance, potentially associated with distinct epidemiological characteristics. Diverse VOCs have been characterized in different parts of the world, as Alpha (Pango lineage B.1.1.7; Nextclade 20I/V1), Beta (Pango lineage B.1.351; Nextclade 20H/V2), Gamma (Pango lineage P.1; Nextclade 20J/V3), and Delta (Pango lineage B.1.617.2; Nextclade 21A) in the UK, South Africa, Brazil, and India, respectively (Tegally et al., 2020; Cherian et al., 2021; Volz et al., 2021; Faria et al., 2021; Naveca et al., 2021). Similarly, genomic surveillance studies support that VOI Zeta (Pango lineage P.2; Nexclade 20B/S.484K) (Voloch et al., 2021) and potential VOIs N.9 (Resende et al., 2021a) and N.10 (Resende et al., 2021b) have originated in different Brazilian regions.

Interestingly, these lineages share mutations in biologically relevant sites of the receptor-binding domain (RBD) of the spike protein. For instance, the substitutions E484K and N501Y are implicated in the distinct epidemiological characteristics reported for VOCs Alpha, Beta, and Gamma (Tegally et al., 2020; Faria et al., 2021; Volz et al., 2021). Some evidence suggest that VOCs can be more transmissible (Faria et al., 2021; Naveca et al., 2021; Tegally et al., 2021; Volz et al., 2021), cause increased disease severity (Davies et al., 2021; Faria et al., 2021), and display reduced neutralization by antibodies elicited by previous infections or vaccines (Cele et al., 2021; Garcia-beltran et al., 2021).

Given these distinct epidemiological behaviors and the putative fitness advantage of VOCs, current evidence supports that they have rapidly risen in frequency, becoming predominant and quickly spreading to other regions (Tegally et al., 2020; Volz et al., 2021). This phenomenon has already been recorded in Brazil. The VOI Zeta, which emerged in Rio de Janeiro in mid-July 2020, rapidly became the primary local lineage and was exported to several regions of the country and abroad (Voloch et al., 2021). Likewise, the emergence of VOC Gamma in mid-November in Manaus is marked by its fast rise in frequency and a massive increase in the number of cases (Faria et al., 2021; Naveca et al., 2021). Currently, diverse surveillance efforts indicate that this lineage has become predominant in several country regions, replacing previously circulating variants (Barbosa et al., 2021; Faria et al., 2021; Franceschi et al., 2021; Naveca et al., 2021; Moreira et al., 2021a). Moreover, VOCs that originated elsewhere, like Alpha, Beta, and Delta, have already been detected in Brazil (Claro et al., 2021; Slavov et al., 2021; SES/MA 2021), adding up to a complex epidemiological scenario (Souza et al., 2020; Santos et al., 2021).

As the circulation of these lineages has broad epidemiological implications for public health, including ongoing vaccination efforts in Brazil, we sought to determine the genetic background of the SARS-CoV-2 epidemic in the city of Rio de Janeiro between early December 2020 and early May 2021. Rio de Janeiro is the second major Brazilian metropolis, an essential hub for business, and a majorly connected city by air travel (IATA 2020). By combining novel genomic and epidemiological data, we were able to characterize a comprehensive shift in the composition of the local SARS-CoV-2 epidemic population induced by VOC Gamma. In addition, we provide further evidence that this variant causes higher viral loads than previously circulating variants, reinforcing its enhanced transmissibility. Altogether, this study documents the complex epidemiological dynamics of SARS-CoV-2 in a major Brazilian metropolis and highlights a possible mechanism by which VOC Gamma dominates distinct epidemiological scenarios.

2. Methods

2.1 Study population and COVID-19 diagnostics

The study population was composed by convenience sampling among the reverse transcription-quantitative polymerase chain reaction (RT-qPCR) positive cases amidst individuals evaluated at the coronavirus disease 2019 (COVID-19) Diagnostic Center of the Federal University of Rio de Janeiro, between 1 December 2020 and 12 May 2021 (total number of tests: 14,080). Nasopharyngeal swab samples were collected from both nostrils, placed in viral transport medium (2 ml), and kept at 4°C until transportation to the laboratory. Total viral RNA from swab samples were extracted in a KingFisher Flex System® (Thermofisher, USA), using the MagMax Viral/Pathogen Kit (Thermofisher, USA), according to manufacturer's instructions. Viral RNA was detected using the SARS-CoV-2 (2019-nCoV) multiplex CDC qPCR Probe Assay (Integrated DNA Technologies, USA), targeting the SARS-CoV-2 *N1* and *N2* genes and the human ribonuclease P (*RNaseP*) gene (endogenous control). The GoTaq® Probe 1-Step RT-qPCR System (Promega, USA) was used, according to the manufacturer's instructions. All reactions were performed in a 7500 Thermal Cycler (Applied Biosystems, USA).

The RT-qPCR result interpretation was as follows: positive for SARS-CoV-2 when both targets (N1 and N2) amplified with cycle threshold (Ct) ≤ 37; undetermined when only one target amplified with $Ct \leq 37$, or both targets amplified with Ct between 37 and 40; and negative when one or both targets amplified with $Ct > 40$, or absence of amplification.

The present study was approved by the local ethics review committee from Clementino Fraga Filho University Hospital (Certificado de Apresentação de Apreciação Ética (CAAE): 30161620.0.0000.5257) and by the national ethical review board (CAAE: 30127020.0.0000.0068). All enrolled participants were over 18 years old and gave written informed consent.

2.2 Genome sequencing

In total, 278 RT-qPCR-positive samples with $Ct < 30$, collected between 11 December 2020 and 5 May 2021, were selected for genome sequencing. The temporal distribution of samples comprehended two epidemic waves of SARS-CoV-2 in Rio de Janeiro city (COE/RJ 2021). Sequencing was carried out using a widely employed protocol (Quick et al., 2017). Briefly, viral RNA was converted to cDNA using SuperScript III or IV (ThermoFisher, USA), followed by a multiplex amplification reaction with the ARTIC SARS-CoV-2 v3 Panel and the Q5 hotstart polymerase (NewEngland Biolabs, USA). After purification, amplicons for each sample were normalized and converted into Illumina sequencing libraries using either the Nextera XT library kit (Illumina, USA) or the QIAseq FX library kit (QIAGEN, Germany), following the manufacturers' protocols. Library fragments were quantified using the Qubit dsDNA High Sensitivity assay (ThermoFisher, USA) and/or the QIAseq Library Quantification kit (QIAGEN, Germany), and their lengths were estimated using the Bioanalyzer High Sensitivity DNA Analysis kit (Agilent, USA). Finally, libraries were diluted into equimolar pools and sequenced in five distinct Illumina MiSeq runs with two V2 Nano (300 cycles) and three V3 (600 cycles) cartridges.

2.3 Viral genome assembly

Sequencing reads were filtered with fastp v0.20.1 (Chen et al., 2018), which removed adapters, short reads (<50 nucleotides) and trimmed low-quality bases (phread < 30). Reads from each sample were mapped against the SARS-CoV-2 reference genome (NCBI accession: NC_045512.2) with Bowtie2 v2.4.2 (Langmead and Salzberg 2012), and mapping files were indexed and sorted with SAMtools v1.12 (Li et al., 2009). BCFtools v1.12 was used for variant calling and consensus genome inference, while BEDtools v2.30.0 (Quinlan and Hall 2010) was used to mask low coverage sites (<100-fold). Sequences with less than 70 per cent genome coverage were removed from downstream analysis.

2.4 Lineage identification and phylogenetic analysis

Lineage identification was performed with pangolin tool v3.0.6 (pango v1.2.12; pangoLEARN model from 5 June 2021; constellations v0.0.4 and scorpio v0.3) (Rambaut et al., 2020; O'Toole et al., 2021). To further confirm lineage assignments and contextualize the novel sequences, a representative global dataset of SARS-CoV-2 genome sequences was assembled ($n = 3{,}609$). This dataset is drawn from a random sample of a larger dataset ($n = 10{,}838$; high-quality data available on GISAID on 13 May 2021) that comprehends all Brazilian sequences, plus one international sequence per country per epidemiological week since the first reported SARS-CoV-2 genome. The reference dataset was designed to be enriched for sequences from lineages relevant in the Brazilian epidemiological scenario while also harboring representativeness of SARS-CoV-2 lineages circulating around the world. These sequences were all aligned to the genome sequences herein described with MAFFT v7.475 (Katoh and Standley 2013), and a maximum likelihood tree was inferred with IQ-tree v2.0.3 (Minh et al., 2020) under the GTR+F+I+G4 model (Tavaré 1986; Yang 1994).

2.5 Phylodynamics

To further access the temporal dynamics of introduction of VOC Gamma in Rio de Janeiro city, we performed molecular clock analyses on a fully Bayesian framework using BEAST v1.10.4 (Suchard et al., 2018). To assemble the reference dataset, all Brazilian VOC Gamma sequences available on (https://www.gisaid.org/) Global Initiative on Sharing Avian Influenza Data (GISAID) were downloaded ($n = 3{,}398$, as of 18 May 2021) and categorized into the following five discrete locations, matching Brazilian geopolitical regions: Southeast ($n = 2{,}090$, except Rio de Janeiro), South ($n = 107$), Northeast ($n = 106$), North ($n = 138$), and Central West ($n = 177$). The 113 novel Gamma genome sequences from Rio de Janeiro characterized in this study were added to a subset of this dataset. All sequences from locations with less than 113 representatives were included, while the same number of sequences from the remaining locations were randomly sampled, composing a geographically balanced dataset ($n = 665$). Preliminary maximum likelihood phylogenetic analysis indicated that 38 reference sequences of this dataset belonged to the recently identified P.1-like clades (Gräf et al., 2021) and were removed from downstream analysis. The final dataset ($n = 627$) was composed of 113, 113, 70, 105, 113, and 113 sequences from Rio de Janeiro, Southeast, South, Northeast, North, and Central West, respectively. The temporal distribution of sequences per location may be visualized in Supplementary Figure S1. Root-to-tip regression was used to assess the temporal signal available in the dataset with TempEst v1.5.3 (Rambaut et al., 2016).

The time-scaled phylogeographic reconstructions used are as follows: *i*—the strict molecular clock model; *ii*—a uniform prior distribution (range: 8×10^{-4}–10^{-3}) on evolutionary rate; *iii*—the coalescent exponential growth tree prior (Laplace prior with scale 100) (Griffiths and Tavare 1994); *iv*—the HKY + I + G4 nucleotide substitution model (Hasegawa, Yano, and Kishino 1984; Yang 1994); and *v*—a discrete phylogeographic model (Lemey et al., 2009). As Brazilian regions are widely connected by roads and air travel, with no clear asymmetry pattern over short time scales (months), a symmetric phylogeographic model was selected. Eight independent chains of 50 million generations sampling every 10,000 states were performed, and convergence (effective sample size > 200 for all parameters) was verified on Tracer v1.7.1 (Rambaut et al., 2018) after 10 per cent burn-in removal. Logcombiner was used to combine posterior distributions and a maximum clade credibility tree was inferred with TreeAnnotator (Drummond and Rambaut 2007).

An additional set of analyses using the nonparametric coalescent skygrid tree prior (Gill et al., 2013), with 25 grid points between the date of the most recently sampled sequence (4 May 2021) and the previously estimated date of VOC Gamma emergence (15 November 2020) (Faria et al., 2021), was also performed. These grids approximately match the number of epidemiological weeks comprehended in this period. For this analysis, six independent runs of 100 million generations, sampling every 10,000 steps, were executed. Maximum clade credibility trees and log files are available in Supplementary File S1.

2.6 Analysis of epidemiological data

To evaluate the hypothesis that the transmissibility enhancement of VOC Gamma is due to increments in viral loads (Faria et al., 2021; Naveca et al., 2021), we performed a series of analyses based on epidemiological data (age, sex, days of symptoms at diagnosis time, and Ct values) of patients infected by viral lineages identified by genome sequencing ($n = 244$ genomes). Throughout this study, the confidence level considered for hypothesis testing was $\alpha = 0.05$. First, a linear regression was used to measure the association between viral lineages (VOC Gamma or non-Gamma) and Ct values (N1 target). This model was also adjusted to account for the effects of age, sex, and number of symptomatic days at diagnosis time.

We also evaluated the time series of all Ct values (N1 and N2 targets, RNaseP control) characterized in the studied period ($n = 1{,}224$), aiming to identify variations of viral loads related to the predominating lineages. To perform a direct comparison, we selected epidemiological weeks for which sequenced samples have shown frequency of VOC Gamma below 20 per cent or above 80 per cent, eliminating intermediate frequency periods. Lineages were then imputed to samples according to the epidemiological week they have been collected, and a linear model was estimated to measure the effect of lineage on Ct values. These analyses have been performed with R software (R Core Team 2020). Code and data are available in Supplementary File S2.

3. Results

3.1 Diagnostics and viral genome sequencing

A total of 1,224 individuals were positive for SARS-CoV-2 detection by RT-qPCR between 1 December 2020 and 12 May 2021 at the COVID-19 Diagnostic Center of the Federal University of Rio de Janeiro. From these, 278 samples collected between 11 December 2020 and 5 May 2021 were selected for complete genome sequencing. In this effort, 244 novel genome sequences with coverage

greater than 70 per cent (median coverage: 97.2 per cent, range: 71.1–99.8 per cent; median depth: 891-fold, depth range: 168–14,595-fold) were characterized, which corresponds to 19.93 per cent of all positive individuals in the studied period. The generated data accounts for 52.5 per cent (244/465) and 14.3 per cent (244/1,711) of all genome sequences available on GISAID EpiCoV database for the city and state of Rio de Janeiro, respectively (as of 30 May 2021). Notwithstanding, these sequenced samples correspond to only approximately 0.04 per cent (244/565,359) of the cases in the city in the studied period. The time between sample collection to sequencing and analysis ranged from 1 to 6 months. Metadata associated with samples are available in Supplementary File S3, including complete sequencing statistics.

3.2 Analysis of viral genomes reveals a VOC-Gamma-driven lineage replacement event

The following diverse lineages were identified in the novel genome sequences with the pangolin tool: P.2 (43.44 per cent; 106/244), P.1 (36.47 per cent; 89/244), B.1.1 (5.73 per cent; 14/244), B.1 (3.69 per cent; 9/244), B.1.1.28 (3.28 per cent; 8/244), P.1.2 (3.28 per cent; 8/244), B.1.1.33 (1.23 per cent; 3/244), B.1.1.7 (1.23 per cent; 3/244), P.4 (0.82 per cent; 2/244), B.1.407 (0.41 per cent; 1/244), and N.9 (0.41 per cent; 1/244; Supplementary File S4). To confirm these classifications and further contextualize these data, we performed a maximum likelihood phylogenetic inference with a globally representative dataset, which showed a different proportion of identified lineages: P.2 (46.72 per cent; 114/244), P.1 (46.31 per cent; 113/244), B.1.1.28 (3.68 per cent; 9/244), B.1.1.33 (1.23 per cent; 3/244), B.1.1.7 (1.23 per cent; 3/244), and N.9 (0.82 per cent; 2/244; Fig. 1A).

This analysis indicates divergent classifications between the pangolin tool and the phylogenetic reconstruction, emphasizing that the latter analysis is important to confirm lineage classifications, especially for incomplete genome sequences. Sequences identified phylogenetically as P.1, P.2, B.1.1.28, and N.9 were previously misclassified as belonging to other lineages ($n = 26/244 = 10.65$ per cent; Supplementary Files S3 and S4).

Noticeably, the genetic composition of the SARS-CoV-2 epidemic population in Rio de Janeiro varied in the studied period (Fig. 1B and C). Lineage P.2 (VOI Zeta) was predominant in December 2020 and January 2021, a pattern that changed with the introduction of Lineage P.1 (VOC Gamma), first appearing in our dataset in the sixth epidemiological week of 2021. Since then, this lineage increased in frequency and rapidly became the major circulating lineage in the city, reaching 100 per cent of the sampled sequences by the twelfth epidemiological week of 2021, with little variation after that. These analyses reveal a complex scenario of co-circulation of multiple VOCs or VOIs in Rio de Janeiro and highlight a fast lineage displacement event induced by VOC Gamma.

3.3 Phylogeographic reconstructions suggest multiple introductions of VOC Gamma into Rio de Janeiro

To further explore the temporal dynamics of introduction and spread of VOC Gamma in Rio de Janeiro city, we performed two Bayesian time-scaled phylogeographic reconstructions using different demographic priors. Root-to-tip regression suggests that a modest temporal signal is available in the dataset ($R^2 = 0.27$; Supplementary Figure S2). The model based on the coalescent exponential growth prior is initially presented, and the one based on the nonparametric coalescent skygrid is described bellow by. Results from both models are summarized in Supplementary Table S1.

The first model places the time for the most recent common ancestor (tMRCA) of VOC Gamma on 5 November 2020 (95 per cent highest posterior density interval (HPD): 11 October 2020–23 November 2020) in the North region of the country (Fig. 2A). After emergence, the phylogeographic model suggests the variant spread to all Brazilian regions through multiple independent events occurring since late 2020. In Rio de Janeiro, 13 separate introductions that led to the emergence of local clades have been reconstructed, coming from the Northeast ($n = 6$), Central West ($n = 2$), South ($n = 1$), and Southeast ($n = 4$) regions (Fig. 2B). A total of seven introductions represented by single sequences have also been identified (South: 2; Southeast: 5).

The model also suggests that the first introduction came from the Northeast region in early January (95 per cent HPD: 7–9 January 2021), after an initial dissemination of the VOC Gamma from the northern region. Posterior independent introductions from other regions of the country have been estimated over the following months (earliest event: 22 January 2021, 95 per cent HPD: 22 January 2021–6 February 2021; oldest event: 7 March 2021, 95 per cent HPD: 17 February 2021–18 March 2021). Likewise, exportation events from Rio de Janeiro have also been estimated, to the Central West ($n = 2$), Southeast ($n = 9$), and South ($n = 1$) regions. This molecular clock analysis provides context for the last upsurge in the number of SARS-CoV-2 infections reported in the city, mainly driven by VOC Gamma (Fig. 2C).

The model based on the nonparametric coalescent skygrid tree prior yielded similar results, dating the emergence of VOC Gamma to 2 December 2020 (95 per cent HPD: 23 November 2020–4 December 2020) in the North region. The model supports that this variant spread from the North to different regions of the country in the following months, being first introduced in Rio de Janeiro in early 2021 (Supplementary Figure S3). After this initial event, 13 other introductions from diverse locations have been estimated (Central West: 1; North: 4; Northeast: 2; South: 1; Southeast: 5) between January and March 2021 (earliest event: 19 January 2021, 95 per cent HPD: 11 January 2021–6 February 2021; oldest event: 7 March 2021, 95 per cent HPD: 17 February 2021–7 March 2021). In addition, 11 introductions supported by single sequences have also been estimated (South: 4; Southeast: 7).

3.4 Analysis of RT-qPCR Ct values suggests VOC Gamma is associated with higher viral loads

To explicitly address the hypothesis that the transmissibility enhancement characterized for VOC Gamma is associated with the induction of higher viral loads, we compared the distribution of Ct values (N1 target) measured for patients infected by VOC Gamma or non-Gamma viruses. First, we conducted an exploratory data analysis to evaluate the impact of sex, age, and symptomatic days at diagnosis time on Ct values using generalized linear models. While the analysis of age and days of symptoms yielded models with statistical significance, despite moderate effects, no association between patient sex and Ct values could be observed (age: $P < 0.01$, $\beta = -0.05$, 95 per cent confidence interval (CI) $= \pm 0.03$; days of symptoms: $P < 0.01$, $\beta = 0.48$, 95 per cent CI $= \pm 0.16$; sex: $P > 0.05$). In this sense, multivariate linear models were performed to estimate the effect of viral lineage on Ct values, also adjusting for the individual and combined effects of age and days of symptoms All models have been summarized in Table 1. No statistical association between viral lineages and

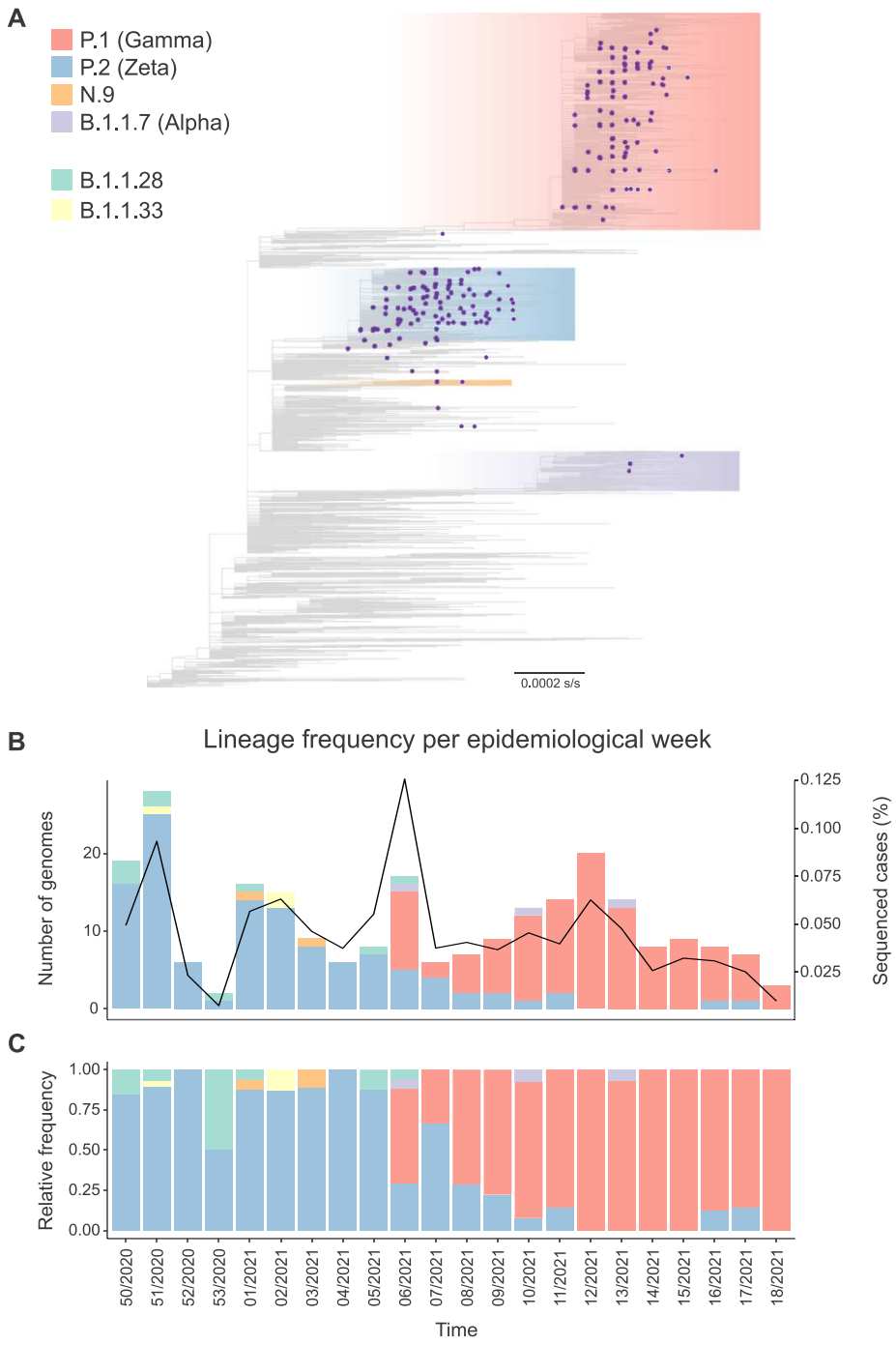

Figure 1. Genetic composition of SARS-CoV-2 lineages in Rio de Janeiro city in the sampled period. (A) Maximum likelihood phylogenetic tree estimated for lineage identification with a globally representative dataset. Genome sequences from this study are highlighted with purple circle tips. Lineages P.1 (Gamma), P.2 (Zeta), N.9, and B.1.1.7 (Alpha) are highlighted in red, blue, orange, and light purple, respectively. Branch length scale shows 0.0002 substitutions per site (s/s). (B) Bar plot exhibiting in absolute numbers the variation in lineages' frequencies across epidemiological weeks from late 2020 to May 2021, as estimated from the novel SARS-CoV-2 genome sequences (Primary Y-axis). The secondary Y-axis indicates the percentage of sequenced cases among the total registered in Rio de Janeiro. (C) Bar plot showing variation in lineages frequencies in relative terms. Lineages are color labeled as in the phylogenetic tree, plus: B.1.1.28 (green) and B.1.1.33 (yellow).

Ct values could be established with statistical significance in any of the evaluated models ($P > 0.05$ for lineages, all models; Fig. 3A).

To further examine the effect of lineages on viral loads, we evaluated the time series of Ct values (N1 target) for all positive individuals in our cohort ($n = 1{,}224$; Fig. 3B). We hypothesized that if VOC Gamma in fact induces higher viral loads, a continuous decrease in Ct values should be observed as its frequency increases. This pattern is consistent with the observed global Ct values between December 2020 (median Ct value: 21.61) and March 2021 (median Ct value: 18.37), in agreement with the increase in the frequency of VOC Gamma. Nevertheless, this tendency moved slightly upward in April 2020 (median Ct value: 18.88), suggesting that other factors could also be influencing

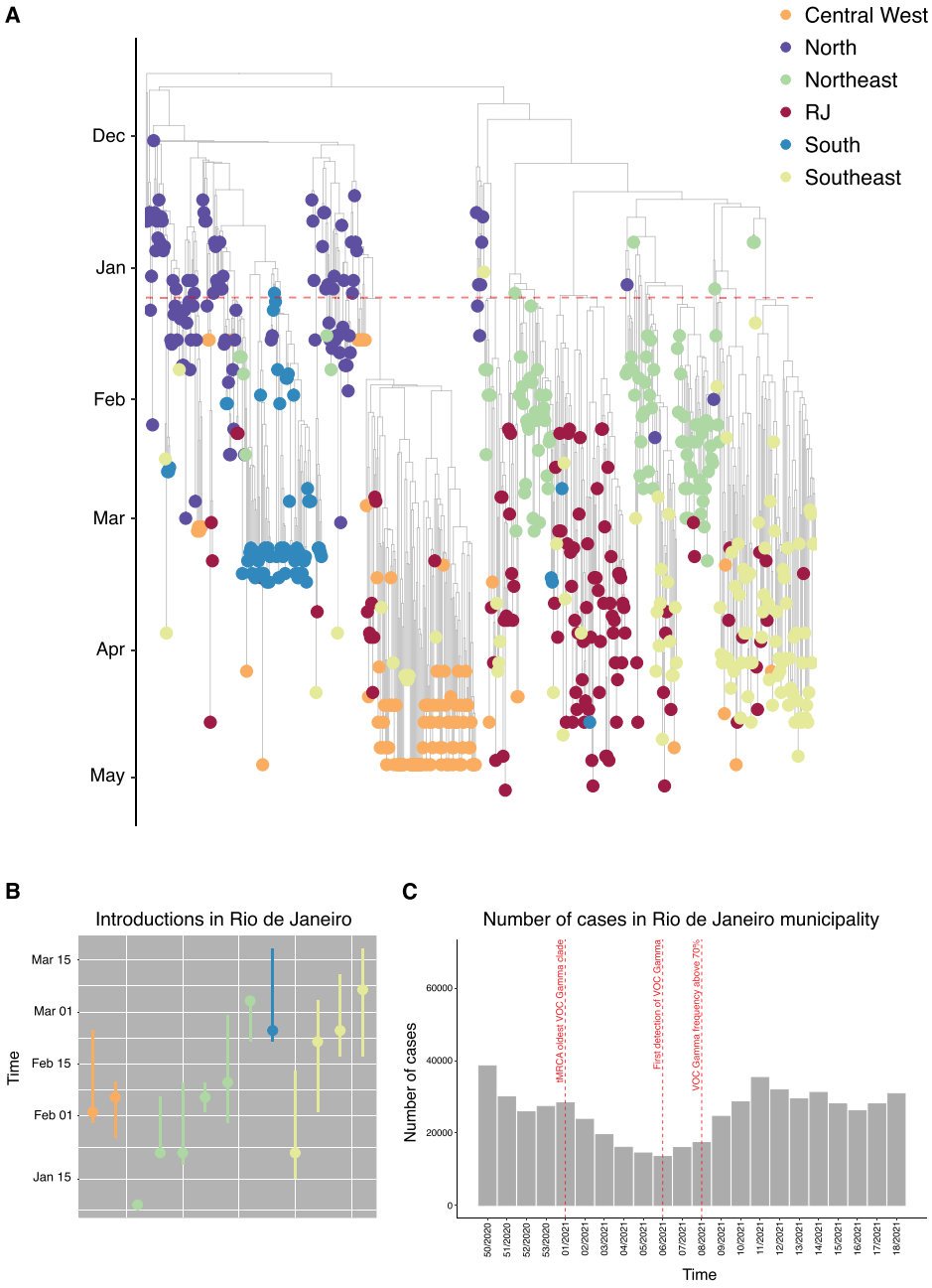

Figure 2. Temporal dynamics of introduction and spread of VOC Gamma in Rio de Janeiro municipality. (A) Phylogeographic reconstructions using a model based on coalescent exponential growth tree prior inferred from a dataset comprehending 514 publicly available VOC Gamma sequences and the 113 new genomes characterized in this study. (B) Dot plot showing the time of each independent VOC Gamma introduction that led to the emergence of local clades. (C) Time series of the general number of cases highlighting the VOC-Gamma-related events. The number of cases was obtained from the Rio COVID-19 panel (available on https://experience.arcgis.com/experience/38efc69787a346959c931568bd9e2cc4, last accessed 22 June 2021). Sampling locations are color indicated: Central West (orange), North (purple), Northeast (green), Rio de Janeiro (red), South (blue), and Southeast (yellow). The dashed red lines on phylogeographic reconstructions and time series highlight (1) the time of tMRCA for the oldest VOC Gamma clade (showed on both panels); (2) first detection of VOC Gamma in Rio de Janeiro (only in time series); and (3) VOC Gamma frequency above 70 per cent (only in time series).

Ct values distribution. Time-series analysis of Ct values for the N2 target yielded similar results, while for the RNaseP control no tendencies could be identified, validating previous results (Supplementary Figure S4).

To perform a direct comparison, lineages have been imputed to all positive samples based on their frequency per epidemiological week, estimated with sequencing data. Samples collected before the sixth epidemiological week of 2021 ($n = 749$; median N1 target Ct value: 21.43) were imputed as non-Gamma, while samples collected from tenth epidemiological week of 2021 forward ($n = 420$; median N1 target Ct value: 18.54) were imputed as VOC Gamma. This comparison suggests the existence of a significant statistical association between lineages and Ct values ($P < 0.01$, $\beta = -2.69$, 95 per cent CI $= \pm 0.70$; Fig. 3C). This result

Table 1. Linear models estimated to evaluate the impact of sex, age, viral lineage, and symptomatic days at diagnosis time on N1 target Ct values of sequenced samples.

Model	Parameters	P-value	β	95% CI
m1	Days of symptoms	<0.01	0.48	±0.16
m2	Age	<0.01	−0.05	±0.04
m3	Sex	0.14	–	–
M1	Lineage	0.09	–	–
M2	Lineage	0.33	–	–
	Days of symptoms	<0.01	0.47	±0.16
M3	Lineage	0.07	–	–
	Age	<0.01	−0.05	±0.03
M4	Lineage	0.29	–	–
	Days of symptoms	<0.01	0.46	±0.16
	Age	<0.01	−0.05	±0.03

Statistically significant values are in italics.

was further corroborated for the N2 target ($P < 0.01$, $\beta = -2.43$, 95 per cent CI $= \pm 0.75$) and no significant statistical association was identified for RNaseP ($P = 0.83$).

4. Discussion

Brazil has been severely impacted by the SARS-CoV-2 pandemic, accounting for 9.5 per cent of the cases and 10.4 per cent of the deaths reported worldwide, even though the country harbors approximately 2.7 per cent of the global population (Castro et al., 2021). While the epidemiological dynamics on a national scale has been shown to be affected by a range of complex factors (Souza et al., 2020; Santos et al., 2021), the scenario is further aggravated by the emergence and widespread circulation of multiple VOCs and VOIs in the country (Claro et al., 2021; Faria et al., 2021; Naveca et al., 2021; Slavov et al., 2021; Voloch et al., 2021; SES/MA 2021; Resende et al., 2021a,b). As these lineages harbor a myriad of mutations of biological significance and epidemiological implications (Cele et al., 2021; Davies et al., 2021; Faria et al., 2021; Garcia-beltran et al., 2021; Tegally et al., 2021; Volz et al., 2021), we sought to investigate their circulation dynamics in the city of Rio de Janeiro, the second largest Brazilian metropolis. By combining novel genetic and epidemiological data, we documented a fast lineage replacement event induced by VOC Gamma and provide additional evidence of its altered epidemiological characteristics.

Our study was conducted between early December 2020 and early May 2021, capturing the second and third epidemic waves in the city of Rio de Janeiro (COE/RJ 2021). Analysis of the 244 novel viral genomes led to the identification of six circulating lineages by phylogenetic inference: P.1, P.2, B.1.1.28, B.1.1.7, B.1.1.33, and N.9 (Fig. 1A). Temporal analysis of lineage frequencies revealed a shift in the genetic composition of SARS-CoV-2 epidemic population in Rio de Janeiro (Fig. 1B and C), with lineage P.2 (VOI Zeta), most frequent between December 2020 and January 2021, being replaced by lineage P.1 (VOC Gamma), first identified in the sixth epidemiological week of 2021. In fact, VOC Gamma has rapidly risen in frequency, being responsible for over 70 per cent of cases only 2 weeks later, that is 7 weeks after its initial detection, coinciding with the third epidemic wave in Rio de Janeiro city (COE/RJ 2021). This result reveals a complex epidemiological scenario, marked by co-circulation of multiple VOCs (Alpha and Gamma) and VOIs (Zeta and potential VOI N.9) and a fast lineage turnover event induced by VOC Gamma, consistent with previous reports on other Brazilian states (Barbosa et al., 2021; Faria et al., 2021; Franceschi et al., 2021; Naveca et al., 2021; Moreira et al., 2021a). Moreover, these results corroborate previous data showing an early detection of VOC Gamma in Rio de Janeiro and provide further context for preliminary genomic analysis performed for the whole state (Almeida et al., 2021; Lamarca et al., 2021).

Beyond the reported lineage displacement event, the dissemination of VOC Gamma in Rio de Janeiro is also marked by a small interval between autochthonous circulation and an associated epidemic wave. While VOI Zeta emerged in mid-July 2020 (Voloch et al., 2021) and rose in frequency through the second semester, causing an epidemic peak over the last months of that year, our molecular clock analysis indicates that only 10 weeks separate the tMRCA of the earliest Gamma clade from Rio de Janeiro and the epidemic peak of the third wave, mostly driven by this lineage (Fig. 2A and C). This observation implies that VOC Gamma not only displayed higher transmissibility than VOI Zeta but also has been associated with an upsurge in the number of cases, in close agreement with reports from Manaus (Faria et al., 2021; Naveca et al., 2021).

Except for B.1.1.28 and B.1.1.33, the predominant lineages in the first epidemic wave in Brazil (Candido et al., 2020), all the remaining lineages are VOCs, VOIs, or potential VOIs and share the mutation E484K on the RBD of the spike protein, shown to weaken neutralizing antibodies' response (Greaney et al., 2021). This observation highlights the previously reported evolutionary convergence on SARS-CoV-2 genomes (Martin et al., 2021). Likewise, both VOCs Alpha and Gamma share the mutation N501Y, also implicated in immune escape and enhancement of the affinity between the RBD and human angiotensin converting enzyme 2, the cellular receptor used by SARS-CoV-2 for cell entry (Lan et al., 2020; Ramanathan et al., 2021). These functional effects, combined with the fast observed increase in frequency and number of cases across diverse epidemiological settings (Barbosa et al., 2021; Faria et al., 2021; Franceschi et al., 2021; Naveca et al., 2021; Moreira et al., 2021a)—including the one herein reported—provide evidence that the VOC Gamma has a fitness advantage over previously circulating lineages.

These results are consistent with the singular epidemiological characteristics reported for other VOCs in distinct epidemiological scenarios, as Alpha (Volz et al., 2021) and Beta (Tegally et al., 2021). Interestingly, although both lineages have been shown to circulate in Brazil (Claro et al., 2021; Slavov et al., 2021), with evidence that VOC Alpha circulates in nearly all regions of the country (Moreira et al., 2021b), no rise in frequency for the later variant could be observed in our dataset. This result suggests that these lineages do not have a fitness advantage in a scenario dominated by VOC Gamma, a hypothesis that must be evaluated across other settings.

Molecular clock analysis performed with the coalescent exponential growth and skygrid tree priors yielded only marginally different results, so the first model will be considered for the purpose of this discussion. The time-scaled phylogenetic reconstruction suggests that VOC Gamma emerged in the North region between middle October and late November 2020, consistent with previous reports from Manaus (Faria et al., 2021; Naveca et al., 2021). The phylogeographic model suggests that after initial dissemination to other regions of the country in early 2021, this lineage was introduced in Rio de Janeiro through multiple events from diverse locations (Fig. 2). While the first estimated introduction came from the Northeast region in early January (95 per cent HPD: 7–9 January 2021), later events from nearly all other regions in the following months have been estimated. This phylogeographic reconstruction emphasizes the mixture of viruses from diverse regions of the

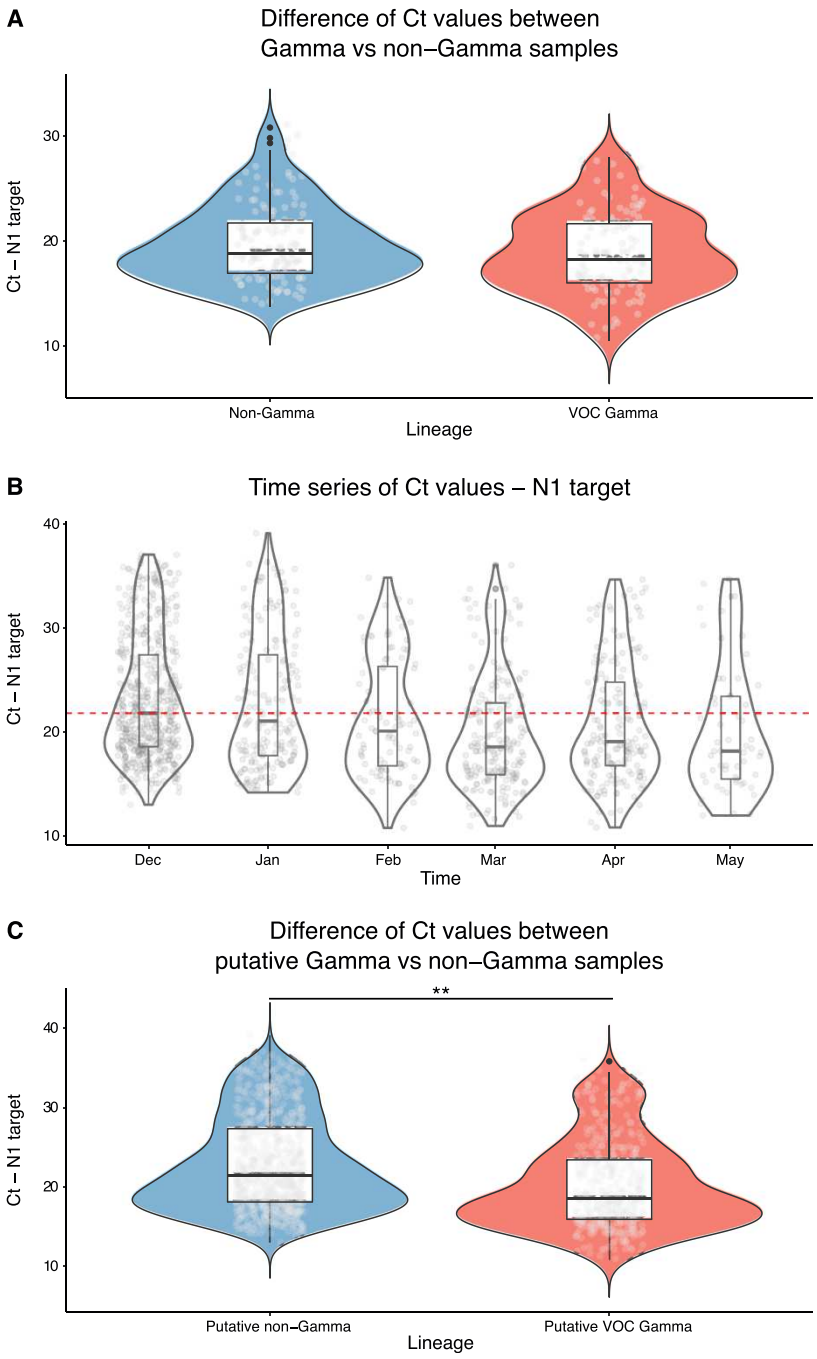

Figure 3. Comparison of the distribution of Ct values (N1 target) measured for patients infected by VOC Gamma or non-Gamma viruses. (A) Violin plots displaying the distribution of Ct values for different viral lineages, as characterized by genome sequencing. (B) Time series of Ct values for all positive samples in our cohort. The dashed red line represents the median Ct value observed in December 2020. (C) Violin plot displaying the distribution of Ct values for groups of samples imputed with distinct viral lineages based on collection dates. Samples collected prior to the sixth epidemiological week of 2021 were imputed as non-Gamma, while samples collected from the tenth week forward were imputed as VOC Gamma. Asterisks indicate significant statistical association between imputed viral lineages and Ct values (linear model: $P < 0.01$, $\beta = -2.69$, 95 per cent CI $= \pm 0.70$).

country, evidencing widespread transmission. This pattern is the outcome of the lack of mobility control measures and reinforces the necessity of coordinated responses and surveillance efforts on a national scale. Although the estimated number of introduction events is certainly underestimated, given the limited number of genomes analyzed, this phylogeographic reconstruction likely reveals general patterns of viral spread across the country and, ultimately, the dynamics of VOC Gamma dissemination into Rio de Janeiro.

The molecular clock analysis also implies that it took approximately 5 weeks for VOC Gamma to be detected by genome sequencing in our randomly selected set of samples and 7 weeks to achieve 70 per cent frequency, consistent with previous rates of growth observed in Manaus (Faria et al., 2021). The scope of these estimates is limited by the employed sampling effort (0.04 per cent of registered COVID-19 cases in Rio de Janeiro) and is conceivable that increasing sample sizes would have led to an earlier detection of VOC Gamma circulating at initially

lower frequencies. Therefore, increasing genomic sampling and reducing the time from sample collection to sequencing and analysis is needed to enhance the performance of surveillance initiatives. Notwithstanding, our results are consistent with the reported epidemic growth of VOC Gamma in all Brazilian regions.

Although current evidence supports that VOC Gamma in fact has a fitness advantage over other lineages circulating in Brazil, the mechanistic basis of this phenomenon has not been unequivocally established. The main hypothesis in this regard implies that this variant is more transmissible due to the induction of infections characterized by higher viral loads (Faria et al., 2021; Naveca et al., 2021), as already observed for other VOCs (Tegally et al., 2021; Volz et al., 2021). Our results agree with this conjecture. No effect of viral lineages on N1 target Ct values distributions could be determined when considering only sequenced samples (Fig. 3A). However, the total number of samples and possible biases associated with sample selection criteria for genome sequencing ($Ct < 30$) limit the scope of this analysis. Notwithstanding, analysis of N1 target Ct values time series was suggestive of a consistent decrement coincident with the increase in the frequency of VOC Gamma (Fig. 3B). This result is further corroborated by the analysis of Ct values for the N2 target and the RNaseP control (Supplementary Figure S4).

These observations were further supported by analysis based on imputed viral lineages through sample collection dates, which shows that samples from individuals likely infected with VOC Gamma viruses have lower Ct values, i.e. higher viral loads (Fig. 3C). These results are in line with previous estimates (Faria et al., 2021; Naveca et al., 2021) and suggest that alternative methods of lineage identification (e.g. RT-qPCR; Naveca et al., 2021; Vogels et al., 2021) may be especially useful in detecting the effect of lineages on Ct values in an unbiased fashion.

It has been suggested that Ct values also vary as a function of epidemiological trajectories (Hay et al., 2021). To ascertain that this factor has not imposed a significant bias on our analysis, we subsetted epidemiological weeks from periods of decay of the second (40/2020 to 01/2021) and third (11/2021 to 16/2021) epidemic waves in Rio de Janeiro. Samples from the weeks of decay of the second and third waves were imputed as non-Gamma and VOC Gamma, respectively. This analysis yielded the same qualitative results (N1 target; $P < 0.01$; $\beta = -2.96$; 95 per cent CI $= \pm 0.81$). The presented findings would benefit from further studies over broader time scales and across different epidemiological settings.

Altogether, this study describes a complex epidemiological dynamics for SARS-CoV-2 in a major metropolis of Brazil, one of the countries most severely hit by the COVID-19 pandemic (Castro et al., 2021). We document the joint circulation of multiple VOCs and VOIs and report a fast lineage displacement event induced by VOC Gamma, providing further evidence of its altered epidemiological characteristics. In addition, through evolutionary analyses, we describe the timing and origins of VOC Gamma introductions in the city. Finally, we bring forth data and analyses that support the hypothesis that the transmissibility enhancement associated with this lineage is at least partially explained by the induction of higher viral loads. Overall, our results document the evolution of this variant and provide independent assessment of scenarios previously studied in Manaus, therefore contributing to the better understanding of the epidemiological dynamics currently being observed in other Brazilian regions.

Data availability

All consensus genome sequences characterized in this study have been deposited on GISAID (Epicodes: EPI_ISL_2629675 to 2629820). Associated metadata and data used in the Ct time-series analysis are available in Supplementary Files S3 and S2, respectively.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgements

The authors would like to thank all authors who submitted their data to GISAID and the EpiCoV curation team for their work. A complete list of acknowledgements is available on Supplementary File S5. The authors would also like to thank the entire technical and administrative staff from Laboratório de Virologia Molecular for their invaluable support, making this and other studies by our group feasible. The authors also thank the support of the sequencing infrastructure of the Laboratório de Uso Comum e Apoio Técnico of the Department of Genetics from the Federal University of Rio de Janeiro. Furthermore, the authors would like to acknowledge Mr Rafael Estima and Druid for their help on handling computational resources that made our analyses possible. Finally, the authors would also like to thank the Instituto Nacional do Cancer staff for the technical support on part of their sequencing runs.

Funding

This study was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (A.T.: E-26/210.178/2020; A.F.A.S.: E-26/202.738/2018, E-26/010.101094/2018, E-26/210.179/2020; R.S.A: 202.922/2018; C.M.V.: 26/010.002278/2019), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (A.F.A.S.: 313005/2020-6; R.S.A.: 312688/2017-2, 439119/2018-9), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (R.S.A.: 14/2020—23072.211119/2020-10), and Financiadora de estudos e projetos (FINEP) (R.S.A: 0494/20 01.20.0026.00). The authors also acknowledge support from the Rede Corona-ômica Ministério de Ciência e Tecnologia, Brasil (BR MCTI)/FINEP affiliated to RedeVírus/MCTI (FINEP 01.20.0029.000462/20, CNPq 404096/2020-4).

Conflict of interest: None declared.

Members of the COVID-19-UFRJ Workgroup

Anna Carla Pinto Castiñeiras, Bianca Ortiz da Silva, Cintia Policarpo, Cynthia Chester Cardoso, Érica Ramos dos Santos Nascimento, Fernanda Leitão dos Santos, Gleidson da Silva de Oliveira, Guilherme Sant'Anna de Lira, Helena D'Anunciação de Oliveira, Helena Toledo Scheid, Isabela Labarba Carvalho de Almeida, Laura Zalcberg Renault, Lidia Theodoro Boullosa, Marcelo Amaral de Souza, Mariana Campos Freire, Natacha Cunha de Araujo Faria, Raquel Fernandes Coelho, Ricardo José Barbosa Salviano, Romina Carvalho Ferreira, Thais Felix, Victor Akira Ota, and Victoria Cortes Bastos.

References

Almeida, L. G. P. et al. 2021) 'Genomic Surveillance of SARS-CoV-2 in the State of Rio de Janeiro, Brazil: A Technical Briefing'. Preprint

from Virological. <https://virological.org/t/genomic-surveillance-of-sars-cov-2-in-the-state-of-rio-de-janeiro-brazil-technical-briefing/683> accessed 14 May 2021.

Barbosa, G. R. et al. (2021) 'Rapid Spread and High Impact of the Variant of Concern P.1 In the Largest City of Brazil', *The Journal of Infection*, 83: S0163–4453(21)00195-X.

Candido, D. S. et al. (2020) 'Evolution and Epidemic Spread of SARS-CoV-2 in Brazil', *Science*, 369: 1255–60.

Castro, M. C. et al. (2021) 'Spatiotemporal Pattern of COVID-19 Spread in Brazil', *Science*, 372: 821–6.

Cele, S. et al. (2021) 'Escape of SARS-CoV-2 501Y.V2 From Neutralization by Convalescent Plasma', *Nature*, 593: 142–6.

Chen, S. et al. (2018) 'Fastp: An Ultra-fast All-in-one FASTQ Preprocessor', *Bioinformatics*, 34: i884–90.

Cherian, S. et al. (2021) 'Convergent Evolution of SARS-CoV-2 Spike Mutations, L452R, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India', *BioRxiv*.

Claro, I. M. et al. (2021) 'Local Transmission of SARS-CoV-2 Lineage B.1.1.7, Brazil, December 2020', *Emerging Infectious Diseases*, 27: 970–2.

COE/RJ - Centro de Operações em Emergência do Rio de Janeiro (2021) COVID-19 - Boletim Epidemiológico - Semana 19. Prefeitura do Rio. <https://coronavirus.rio/wp-content/uploads/2021/05/Covid_BoletimEpid_Sem19-_v1.pdf> accessed 14 May 2021.

Davies, N. G. et al. (2021) 'Increased Mortality in Community-tested', *Nature*, 593: 270–4.

de Jesus, J. G. et al. (2020) 'Importation and Early Local Transmission of Covid-19 in Brazil, 2020', *Revista do Instituto de Medicina Tropical de São Paulo*, 62: e30.

Drummond, A. J., and Rambaut, A. (2007) 'BEAST: Bayesian Evolutionary Analysis by Sampling Trees', *BMC Evolutionary Biology*, 7: 214.

Faria, N. R. et al. (2021) 'Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil', *Science*, 372: 815–21.

Franceschi, V. B. et al. (2021) 'Predominance of the SARS-CoV-2 Lineage P.1 And Its Sublineage P.1.2 In Patients from the Metropolitan Region of Porto Alegre, Southern Brazil in March. *Pathogens*, 10: 988.

Garcia-beltran, W. F. et al. (2021) 'Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-induced Humoral Immunity Ll Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-induced Humoral Immunity', *Cell*, 184: 2372–83.e9.

Gill, M. S. et al. (2013) 'Improving Bayesian Population Dynamics Inference: A Coalescent-based Model for Multiple Loci', *Molecular Biology and Evolution*, 30: 713–24.

Gräf, T. et al. (2021) 'Identification of SARS-CoV-2 P.1-related Lineages in Brazil Provides New Insights about the Mechanisms of Emergence of Variants of Concern', *Preprint from Research Square*.

Greaney, A. J. et al. (2021) 'Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition', *Cell Host & Microbe*, 29: 44–57.e9.

Griffiths, R. C., and Tavare, S. (1994) 'Sampling Theory for Neutral Alleles in a Varying Environment', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 344: 403–10.

Hasegawa, M., Yano, T., and Kishino, H. (1984) 'A New Molecular Clock of Mitochondrial DNA and the Evolution of Hominoids', *Proceedings of the Japan Academy, Series B*, 60: 95–8.

Hay, J. A. et al. (2021) 'Estimating Epidemiologic Dynamics from Cross-sectional Viral Load Distributions', *Science*, 373: eabh0635.

IATA (2020) Air Connectivity: Measuring the connections that drive economic growth. <https://www.iata.org/en/iata-repository/publications/economic-reports/air-connectivity-measuring-the-connections-that-drive-economic-growth/> accessed 15 May 2021.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Lamarca, A. P. et al. (2021) 'Genomic Surveillance of SARS-CoV-2 Tracks Early Interstate Transmission of P.1 Lineage and Diversification within P.2 Clade in Brazil', *PLOS Neglected Tropical Diseases*, 15: e0009835.

Lan, J. et al. (2020) 'Structure of the SARS-CoV-2 Spike Receptor-binding Domain Bound to the ACE2 Receptor', *Nature*, 581: 215–20.

Langmead, B., and Salzberg, S. L. (2012) 'Fast Gapped-read Alignment with Bowtie 2', *Nature Methods*, 9: 357–9.

Lemey, P. et al. (2009) 'Bayesian Phylogeography Finds Its Roots 5', *PLoS Computational Biology*, 5: e1000520.

Li, H. et al. (2009) 'The Sequence Alignment/Map Format and SAMtools', *Bioinformatics*, 25: 2078–9.

Martin, D. P. et al. (2021) 'The Emergence and Ongoing Convergent Evolution of the N501Y Lineages Coincides with a Major Global Shift in the SARS-CoV-2 Selective Landscape'. Preprint from medRxiv.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Moreira, F. et al. (2021a) 'Increasing Frequency of SARS-CoV-2 Lineages B.1.1.7, P.1 And P.2 And Identification of a Novel Lineage Harboring E484Q and N501T Spike Mutations in Minas Gerais, Southeast Brazil'. Preprint from Virological. <https://virological.org/t/increasing-frequency-of-sars-cov-2-lineages-b-1-1-7-p-1-and-p-2-and-identification-of-a-novel-lineage-harboring-e484q-and-n501t-spike-mutations-in-minas-gerais-southeast-brazil/676> accessed 14 May 2021.

Moreira, F. R. R. et al. (2021b) 'Epidemic Spread of SARS-CoV-2 Lineage B.1.1.7 In Brazil', *Viruses*, 13: 984.

Naveca, F. G. et al. (2021) 'COVID-19 in Amazonas, Brazil, Was Driven by the Persistence of Endemic Lineages and P.1 Emergence', *Nature Medicine*, 27: 1230–8.

O'Toole, Á. et al. (2021) 'Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool', *Virus Evolution*, 7: veab064.

Quick, J. et al. (2017) 'Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples', *Nature Protocols*, 12: 1261–76.

Quinlan, A. R., and Hall, I. M. (2010) 'BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features', *Bioinformatics*, 26: 841–2.

R Core Team (2020) *R: A language and environment for statistical computing. R Foundation for Statistical Computing* (Vienna, Austria). <https://www.R-project.org/> accessed 14 May 2021.

Ramanathan, M. et al. (2021) 'SARS-CoV-2 B.1.1.7 And B.1.351 Spike Variants Bind Human ACE2 with Increased Affinity', *The Lancet Infectious Diseases*, 21: S1473–3099(21)00262-0.

Rambaut, A. et al. (2018) 'Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.

——— et al. (2020) 'A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology', *Nature Microbiology*, 5: 1403–7.

——— et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst', *Virus Evolution*, 2: vew007.

Resende, P. C. et al. (2021a) 'A Potential Sars-cov-2 Variant of Interest (Voi) Harboring Mutation E484k in the Spike Protein Was Identified within Lineage B.1.1.33 Circulating in Brazil', *Viruses*, 13: 724.

——— et al. (2021b) 'Identification of a New B.1.1.33 SARS-CoV-2 Variant of Interest (VOI) Circulating in Brazil with Mutation E484K and Multiple Deletions in the Amino (N)-terminal Domain of

the Spike Protein'. Preprint from Virological. <https://virological.org/t/identification-of-a-new-b-1-1-33-sars-cov-2-variant-of-interest-voi-circulating-in-brazil-with-mutation-e484k-and-multiple-deletions-in-the-amino-n-terminal-domain-of-the-spike-protein/675> accessed 14 May 2021.

Santos, A. A. et al. (2021) 'Dataset on SARS-CoV-2 Non-pharmaceutical Interventions in Brazilian Municipalities', *Scientific Data*, 8: 73.

SES/MA (2021) Secretaria Estadual de Saúde do estado do Maranhão, Brasil. <https://www.saude.ma.gov.br/destaques/nota-variante-indiana/> accessed 20 May 2021.

Slavov, S. N. et al. (2021) 'Genomic Monitoring Unveil the Early Detection of the SARS-CoV-2 B.1.351 Lineage (20H/501Y.V2) in Brazil'. Preprint from medRxiv.

Souza, W. M. et al. (2020) 'Epidemiological and Clinical Characteristics of the COVID-19 Epidemic in Brazil', *Nature Human Behaviour*, 4: 856–65.

Suchard, M. A. et al. (2018) 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10', *Systematic Biology*, 4: vey016.

Tavaré, S. (1986) 'Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences', *American Mathematical Society Lectures on Mathematics in the Life Sciences*, 17: 57–86. doi: citeulike-article-id: 4801403.

Tegally, H. et al. (2020) 'Emergence and Rapid Spread of a New Severe Acute Respiratory Syndrome-related Coronavirus 2 (Sars-cov-2) Lineage with Multiple Spike Mutations in South Africa'. Preprint from medRxiv.

——— et al. (2021) 'Detection of a SARS-CoV-2 Variant of Concern in South Africa', *Nature*, 592: 438–43.

Vogels, C. B. F. et al. (2021) 'Multiplex qPCR Discriminates Variants of Concern to Enhance Global Surveillance of SARS-CoV-2', *PLoS Biology*, 19: e3001236.

Voloch, C. M. et al. (2021) 'Genomic Characterization of a Novel SARS-CoV-2 Lineage from Rio de Janeiro, Brazil', *Journal of Virology*, 95: e00119–21.

Volz, E. et al. (2021) 'Assessing Transmissibility of SARS-CoV-2 Lineage B.1.1.7 In England', *Nature*, 593: 266–9.

WHO (2020) WHO Director-General's opening remarks at the media briefing on COVID19. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020> accessed 14 May 2021.

Wu, F. et al. (2020) 'A New Coronavirus Associated with Human Respiratory Disease in China', *Nature*, 579: 265–9.

Yang, Z. (1994) 'Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods', *Journal of Molecular Evolution*, 39: 306–14.