

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

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

1 - Laboratório de Virologia Molecular, Departamento de Genética, Universidade Federal do Rio de Janeiro

2 - Laboratório de Diversidade e Doenças Virais, Departamento de Genética, Universidade Federal do Rio de Janeiro

3 - Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro

4 - Departamento de Doenças Infecciosas e Parasitárias, Faculdade de Medicina, Universidade Federal do Rio de Janeiro

5 - Laboratório de Biologia Integrativa, Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais

6 – Instituto D'Or de Pesquisa e Ensino (IDOR), Rio de Janeiro

\* These authors equally contributed to this work.

† These authors share the senior authorship.

Corresponding author: Amilcar Tanuri (email: [atanuri@gmail.com](mailto:atanuri@gmail.com))

## **Abstract:**

The emergence and widespread circulation of SARS-CoV-2 variants of concern (VOC) or interest (VOI) imposes 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 this variant has circulated locally since the first weeks of 2021 and only seven weeks were necessary for it to achieve a frequency above 70%, consistent with rates of growth observed in Manaus and other states. Moreover, a Bayesian phylogeographic reconstruction indicates 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 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% CI =  $\pm 0.7$ ). 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.

## **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 United Kingdom, South Africa, Brazil and India, respectively (Cherian et al, 2021; Tegally et al., 2020; 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 (Faria et al., 2021; Tegally et al., 2020; Volz et al., 2021). Some evidences 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 spread 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 (Voloach 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; Moreira et al., 2021a; Naveca et al., 2021). 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 (Santos et al., 2021; Souza et al., 2020).

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 RT-qPCR positive cases amidst individuals evaluated at the COVID-19 Diagnostic Center of the Federal University of Rio de Janeiro, between 1st December 2020 and 12 May 2021. 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<sup>®</sup> (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<sup>®</sup> 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  $Ct \leq 37$ ; Undetermined when only one target amplified with  $Ct \leq 37$ , or both targets amplified with  $Ct$  between 37 and 40; 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 (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 declared 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 Qubit dsDNA High Sensitivity assay (ThermoFisher, USA) and/or the QIAseq Library Quantification kit (QIAGEN, Germany) and their lengths were estimated using 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% genome coverage were removed from downstream analysis.

## **2.4 – Lineage identification and phylogenetic analysis**

Lineage identification was performed with the pangolin tool v2.4.2 (Rambaut et al., 2020; O'Toole et al., 2020). 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. 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 references dataset, all Brazilian VOC Gamma sequences available on GISAID were downloaded ( $n = 3,398$ , as of 18 May 2021) and categorized into 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 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 time scaled phylogeographic reconstruction used: *i* – the strict molecular clock model; *ii* – a uniform prior distribution (range:  $8 \times 10^{-4} - 10^{-3}$ ) on evolutionary rate; *iii* – the coalescent exponential growth tree prior (Laplace prior with scale 100) (Kingman, 1982); *iv* – the HKY+I+G4 nucleotide substitution model (Hasegawa et al., 1984; Yang, 1994); and *v* – a symmetric discrete phylogeographic model (Lemey et al., 2009). 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% burnin removal. Logcombiner was used to combine posterior distributions and a maximum clade credibility tree was inferred with treeannotator.

An additional set of analyses using the non-parametric 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, 6 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% or above 80%, 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 the R software (R core team, 2020).

### **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% (median coverage: 97.2%, range: 71.1 - 99.8%; median depth: 891-fold, depth range: 168 - 14,595-fold) were characterized, which corresponds to 19.93% of all positive individuals in the studied period. The generated data accounts for 52.5% (244/465) and 14.3% (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). Metadata associated with samples are available in **Supplementary File S2**, including complete sequencing statistics.

#### **3.2 – Analysis of viral genomes reveals a VOC Gamma driven lineage replacement event**

Diverse lineages were identified in the novel genome sequences with the pangolin tool: P.2 (43.44%; 106/244), P.1 (36.47%; 89/244), B.1.1 (5.73%; 14/244), B.1 (3.69%; 9/244), B.1.1.28 (3.28%; 8/244), P.1.2 (3.28%; 8/244), B.1.1.33 (1.23%; 3/244), B.1.1.7 (1.23%; 3/244), P.4 (0.82%; 2/244), B.1.407 (0.41%; 1/244) and N.9 (0.41%; 1/244) (**Supplementary File S2**). 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%; 114/244), P.1 (46.31%; 113/244), B.1.1.28 (3.68%; 9/244), B.1.1.33 (1.23%; 3/244), B.1.1.7 (1.23%; 3/244) and N.9 (0.82%; 2/244) (**Figure 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 (**Supplementary File S2**).

Noticeably, the genetic composition of the SARS-CoV-2 epidemic population in Rio de Janeiro varied in the studied period (**Figures 1B and C**). The 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% 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. The model based on the coalescent exponential growth prior is initially presented and the one based on the non-parametric coalescent skygrid is described below. 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% highest posterior density interval, HPD: 11

October 2020 - 23 November 2020) in the North region of the country (**Figure 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 (**Figure 2B**). A total of 7 introductions represented by single sequences have also been identified (South: 2; Southeast: 5).

The model also implies that the first introduction came from the Northeast region in early January (95% 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% HPD: 22 January 2021 - 6 February 2021; oldest event: 7 March 2021, 95% 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 (**Figure 2C**).

The model based on the non-parametric coalescent skygrid tree prior yielded similar results, dating the emergence of VOC Gamma to 2 December 2020 (95% 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 S1**). 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% HPD: 11 January 2021 - 6 February

2021; oldest event: 7 March 2021, 95% HPD: 17 February 2021 - 7 March 2021). In addition, 11 introductions supported by single sequences have also been estimated (South: 4; Southeast: 7).

The evolutionary rate estimated under the exponential model was  $9.27 \times 10^{-4}$  substitutions per site per year (s/s/y) (95% HPD:  $8.53 \times 10^{-4}$  -  $9.99 \times 10^{-4}$ ) and the coalescent exponential growth rate has been estimated as 14.82 (95% HPD: 12.95 - 16.76), which implicates a doubling time of 0.047 years or approximately 17 days (15 - 19 days, considering the 95% HPD). Similarly, results obtained with the coalescent skygrid tree prior indicated an evolutionary rate of  $8.93 \times 10^{-4}$  s/s/y (95% HPD:  $8.10 \times 10^{-4}$  -  $9.73 \times 10^{-4}$ ) and a steady growth in relative genetic diversity until early 2021, which stabilized over the following months (**Supplementary Figure S2**).

### **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 analyses 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% CI =  $\pm 0.03$ ; days of symptoms:  $p < 0.01$ ,  $\beta = 0.48$ , 95% 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 Ct values could be established with statistical significance in any of the evaluated models ( $p > 0.05$  for lineages, all models) (**Figure 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 ( $n = 1,224$ ; **Figure 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 rise in frequency of VOC Gamma. Nevertheless, this tendency moved slightly upwards in April 2020 (median Ct value: 18.88), suggesting other factors could also be influencing 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 S3**).

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% CI =  $\pm 0.70$ ) (**Figure 3C**). This result was further corroborated for the N2 target ( $p < 0.01$ ,  $\beta = -2.43$ , 95% 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% of the cases and 10.4% of the deaths reported worldwide, even though the country harbors approximately 2.7% 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 (Santos et al., 2021; Souza et al., 2020), 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; Resende et al., 2021a; Resende et al., 2021b; Slavov et al., 2021; Voloch et al., 2021; SES/MA, 2021). 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 (**Figure 1A**). Temporal analysis of lineage frequencies revealed a shift in the genetic composition of SARS-CoV-2 epidemic population in Rio de Janeiro (**Figures 1B and 1C**), 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% of cases only two weeks later, 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; Moreira et al., 2021a; Naveca et al., 2021). 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 (**Figures 2A and 2C**). 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 (ACE2), the cellular receptor used by SARS-CoV-2 for cell entry (Lan et al., 2020; Ramanathan et al., 2021). These functional effects, joint to the fast observed rise in frequency and number of cases across diverse epidemiological settings (Barbosa et al., 2021; Faria et al., 2021; Franceschi et al., 2021; Moreira et al., 2021a; Naveca et al., 2021) - 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 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 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 (**Figure 2**). While the first estimated

introduction came from the Northeast region in early January (95% 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 country, evidencing the necessity of coordinated 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 five weeks for VOC Gamma to be detected by genome sequencing in our randomly selected set of samples and seven weeks to achieve 70% frequency, consistent with previous rates of growth observed in Manaus (Faria et al., 2021). Moreover, demographic analysis based on the coalescent exponential growth model implies that this lineage had a doubling time of approximately 17 days since its emergence (**Supplementary Figure S2A**). This result is consistent with the reported epidemic growth of VOC Gamma in Brazil and provides independent evidence of this phenomenon based solely on genetic data. An equivalent growth was estimated under the non-parametric coalescent skygrid tree prior (**Supplementary Figure S2B**).

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 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 (**Figure 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 were suggestive of a consistent decrement coincident with the rise in frequency of VOC Gamma (**Figure 3B**). This result is further corroborated by the analysis of  $Ct$  values for the N2 target and the RNaseP control (**Supplementary Figure S3**).

These observations were further confirmed by analysis based on imputed viral lineages through sample collection dates, supporting that samples from individuals infected with VOC Gamma viruses have lower  $Ct$  values, *i.e.*, higher viral loads (**Figure 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 variate 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% CI =  $\pm 0.81$ ). The presented findings would benefit from further studies over broader time scales and across different epidemiological settings.

Additionally, it is conceivable that other mechanisms might also play a role in the enhancement of VOC Gamma transmissibility. For instance, assays measuring binding between ACE2 and the RBD of VOCs Alpha and Beta showed 1.98- and 4.62-fold

increments on affinity in comparison to the original Wuhan isolate, respectively (Ramanathan et al., 2021). As VOC Gamma harbors multiple RBD mutations in common with VOC Beta – K417N/T, E484K, N501Y – it is likely that it also exhibits differences in the affinity with the human ACE2, enhancing its transmission potential. Future functional studies should evaluate this conjecture.

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.

## 5 – References:

- Almeida, L.G.P. de, Lamarca, A.P., Jr, R. da S.F., Cavalcante, L., Gerber, A.L., Guimarães, A.P. de C., Machado, D.T., Alves, C., Mariani, D., Cruz, T.F., Ribeiro, M.S., Carvalho, S., Silva, F.D. da, Garcia, M.H. de O., Souza, L.M. de, Silva, C.G. Da, Ribeiro, C.L.P., Cavalcanti, A.C., Mello, C.M.B. de, Tanuri, A., Vasconcelos, A.T.R., 2021. Genomic Surveillance of SARS-CoV-2 in the State of Rio de Janeiro, Brazil: a technical briefing. Preprint from Virological. [Internet]. 14 May 2021. Available from: <https://virological.org/t/genomic-surveillance-of-sars-cov-2-in-the-state-of-rio-de-janeiro-brazil-technical-briefing/683>.
- Barbosa, G.R., Moreira, L.V.L., Justo, A.F.O., Perosa, A.H., Chaves, A.P.C., Bueno, M.S., Luna, L.K. de S., Conte, D.D., Carvalho, J.M.A., Prates, J., Dantas, P.S., Faico-Filho, K.S., Camargo, C., Resende, P.C., Siqueira, M.M., Bellei, N., 2021. Rapid spread and high impact of the Variant of Concern P.1 in the largest city of Brazil. *J Infect.* S0163-4453(21)00195-X. doi: 10.1016/j.jinf.2021.04.008.
- Candido, D.S., Claro, I.M., de Jesus, J.G., de Souza, W.M., Moreira, F.R.R., Dellicour, S., Mellan, T.A., du Plessis, L., Pereira, R.H.M., Sales, F.C. da S., Manuli, E.R., Thézé, J., Almeida, L., de Menezes, M.T., Voloch, C.M., Fumagalli, M.J., Coletti, T. de M., Silva, C.A.M., Ramundo, M.S., Amorim, M.R., Hoeltgebaum, H., Mishra, S., Gill, M.S., Carvalho, L.M., Buss, L.F., Prete, C.A., Ashworth, J., Nakaya, H., Peixoto, P. da S., Brady, O.J., Nicholls, S.M., Tanuri, A., Rossi, A.D., Braga, C.K.V., Gerber, A.L., Guimarães, A.P., Gaburo, N., Alencar, C.S., Ferreira, A.C. de S., Lima, C.X., Levi, J.E., Granato, C., Ferreira, G.M., Francisco, R. da S., Granja, F., Garcia, M.T., Moretti, M.L., Perroud, M.W., Castineiras, T.M.P.P., Lazari, C.D.S., Hill, S.C., de Souza Santos, A.A., Simeoni, C.L., Forato, J., Sposito, A.C., Schreiber, A.Z., Santos, M.N.N., de Sá, C.Z., Souza, R.P., Resende-Moreira, L.C.,

- Teixeira, M.M., Hubner, J., Leme, P.A.F., Moreira, R.G., Nogueira, M.L., Ferguson, N.M., Costa, S.F., Proenca-Modena, J.L., Vasconcelos, A.T.R., Bhatt, S., Lemey, P., Wu, C.H., Rambaut, A., Loman, N.J., Aguiar, R.S., Pybus, O.G., Sabino, E.C., Faria, N.R., 2020. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. 369(6508):1255-1260. doi: 10.1126/science.abd2161.
- Castro, M.C., Kim, S., Barberia, L., Ribeiro, A.F., Gurzenda, S., Ribeiro, K.B., Abbott, E., Blossom, J., Rache, B., Singer, B.H., 2021. Spatiotemporal pattern of COVID-19 spread in Brazil. *Science*. 372(6544):821-826. doi: 10.1126/science.abh1558.
- Cele, S., Gazy, I., Jackson, L., Hwa, S.-H., Tegally, H., Lustig, G., Giandhari, J., Pillay, S., Wilkinson, E., Naidoo, Y., Karim, F., Ganga, Y., Khan, K., Bernstein, M., Balazs, A.B., Gosnell, B.I., Hanekom, W., Moosa, M.-Y.S., Abrahams, S., Alcantara, L.C.J., Alisoltani-Dehkordi, A., Allam, M., Bhiman, J.N., Davies, M.-A., Doolabh, D., Engelbrecht, S., Fonseca, V., Giovanetti, M., Glass, A.J., Godzik, A., Goedhals, D., Hardie, D., Hsiao, M., Iranzadeh, A., Ismail, A., Korsman, S., Pond, S.L.K., Laguda-Akingba, O., Lourenco, J., Marais, G., Martin, D., Maslo, C., Mlisana, K., Mohale, T., Msomi, N., Mudau, I., Petruccione, F., Preiser, W., San, E.J., Sewell, B.T., Tyers, L., Van Zyl, G., von Gottberg, A., Walaza, S., Weaver, S., Wibmer, C.K., Williamson, C., York, D., Archary, M., Dullabh, K.J., Goulder, P., Harilall, S., Harling, G., Harrichandparsad, R., Herbst, K., Jeena, P., Khoza, T., Klein, N., Kløverpris, H., Leslie, A., Madansein, R., Marakalala, M., Mazibuko, M., Moshabela, M., Mthabela, N., Naidoo, K., Ndhlovu, Z., Ndung'u, T., Nyamande, K., Padayatchi, N., Patel, V., Ramjit, D., Rodel, H., Smit, T., Steyn, A., Wong, E., Lessells, R.J., de Oliveira, T., Sigal, A., Africa, N. for G.S. in S., 2021. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature*. 593(7857):142-146. doi: 10.1038/s41586-021-03471-w.

- Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34(17):i884-i890. doi: 10.1093/bioinformatics/bty560.
- Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das M., Rakshit, P., Singh, S., Abraham, P. Panda, S., NIC team. 2021. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. bioRxiv. doi: <https://doi.org/10.1101/2021.04.22.440932>.
- Claro, I.M., Cristina, F., Ramundo, M.S., Candido, D.S., Silva, C.A.M., Jesus, J.G. De, Manuli, E.R., Oliveira, C.M. De, Scarpelli, L., Campana, G., Pybus, O.G., Sabino, E.C., Faria, N.R., Levi, J.E., 2021. Local Transmission of SARS-CoV-2 Lineage B.1.1.7, Brazil, December 2020. *Emerg Infect Dis*. 27(3):970-972. doi: 10.3201/eid2703.210038.
- COE/RJ - Centro de Operações em Emergência do Rio de Janeiro, 2021. COVID-19 - Boletim Epidemiológico - Semana 19. Prefeitura do Rio. [Internet]. 14 May 2021. Available from: [https://coronavirus.rio/wp-content/uploads/2021/05/Covid\\_BoletimEpid\\_Sem19-\\_v1.pdf](https://coronavirus.rio/wp-content/uploads/2021/05/Covid_BoletimEpid_Sem19-_v1.pdf).
- Davies, N.G., Jarvis, C.I., Group, C.C.-W., Edmunds, W.J., Jewell, N.P., Diaz-ordaz, K., Keogh, R.H., 2021. Increased mortality in community-tested. *Nature*. 593(7858):270-274. doi: 10.1038/s41586-021-03426-1.
- de Jesus, J.G., Sacchi, C., Candido, D. da S., Claro, I.M., Sales, F.C.S., Manuli, E.R., da Silva, D.B.B., de Paiva, T.M., Pinho, M.A.B., Santos, K.C. de O., Hill, S.C., Aguiar, R.S., Romero, F., Dos Santos, F.C.P., Gonçalves, C.R., Timenetsky, M.D.C., Quick, J., Croda, J.H.R., de Oliveira, W., Rambaut, A., Pybus, O.G., Loman, N.J., Sabino, E.C., Faria, N.R., 2020. Importation and early local transmission of covid-19 in



Brazil, 2020. *Rev Inst Med Trop Sao Paulo*. 62:e30. doi: 10.1590/s1678-9946202062030.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D. da S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulswit, R.J.G., Franco, L.A.M., Ramundo, M.S., Jesus, J.G. de, Andrade, P.S., Coletti, T.M., Ferreira, G.M., Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Jr., N.G., Camilo, C. da C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., Souza, L.M. de, Pinho, M.C. de, Araujo, L.J.T., Malta, F.S. V., Lima, A.B. de, Silva, J. do P., Zauli, D.A.G., Ferreira, A.C. de S., Schnekenberg, R.P., Laydon, D.J., Walker, P.G.T., Schlüter, H.M., Santos, A.L.P. dos, Vidal, M.S., Caro, V.S. Del, Filho, R.M.F., Santos, H.M. dos, Aguiar, R.S., Proença-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Jr., C.A.P., Nascimento, V.H., Suchard, M.A., Bowden, T.A., Pond, S.L.K., Wu, C.-H., Ratmann, O., Ferguson, N.M., Dye, C., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M. do P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 372(6544):815-821. doi: 10.1126/science.abh2644.

Franceschi, V.B., Caldana, G.D., Perin, C., Horn, A., Cadeiani, A., Zimmerman, R.A., Thompson, C.E., 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 2021: a phylogenomic analysis. Preprint from medRxiv. doi: 10.1101/2021.05.18.21257420v1.

Garcia-beltran, W.F., Lam, E.C., Denis, K.S., Iafrate, A.J., Naranbhai, V., Balazs, A.B., Garcia-beltran, W.F., Lam, E.C., Denis, K.S., Nitido, A.D., Garcia, Z.H., Hauser,

- B.M., 2021. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity II Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*. 184(9):2372-2383.e9. doi: 10.1016/j.cell.2021.03.013.
- Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., Suchard, M.A., 2013. Improving bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol Biol Evol*. 30(3):713-24. doi: 10.1093/molbev/mss265.
- Gräf, T., Bello, G., Venas, T.M.M., Pereira, E.C., Paixão, A.C.D., Appolinario, L.R., Lopes, R.S., Mendonça, A.C. da F., Rocha, A.S.B. da, Motta, F.C., Gregianini, T.S., Salvato, R., Fernandes, S.B., Rovaris, D.B., Cavalcanti, A.C., Leite, A.B., Riediger, I., Debur, M. do C., Bernardes, A.F.L., Ribeiro-Rodrigues, R., Grinsztejn, B., Dezordi, F.Z., Wallau, G.L., Naveca, F.G., Delatorre, E., Siqueira, M.M., Resende, and P.C., Network., F.C.-G.S., 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. doi: 10.21203/rs.3.rs-580195/v1.
- Greaney, A.J., Starr, T.N., Gilchuk, P., Carnahan, R.H., Crowe, J.E., Bloom, J.D., 2021. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe*. 29(1):44-57.e9. doi: 10.1016/j.chom.2020.11.007.
- Hasegawa, M., Yano, T., Kishino, H., 1984. A new molecular clock of mitochondrial DNA and the evolution of hominoids. *Proc. Japan Acad. Ser. B Phys. Biol. Sci.* 60, 95–98. doi: 10.2183/pjab.60.95.

- Hay, J.A., Kennedy-Shaffer, L., Kanjilal, S., Lennon, N.J., Gabriel S.B., Lipsitch M., Mina M.J., 2021. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science*. eabh0635. doi: 10.1126/science.abh0635.
- IATA, 2020. Air Connectivity: Measuring the connections that drive economic growth. [Internet]. 15 May 2021. Available from: <https://www.iata.org/en/iata-repository/publications/economic-reports/air-connectivity-measuring-the-connections-that-drive-economic-growth/>
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010.
- Kingman, J.F.C., 1982. The coalescent. *Stoch. Process. their Appl.* 13, 235–248. doi: 10.1016/0304-4149(82)90011-4.
- Lamarca, A.P., Almeida, L.G.P. de, Jr, R. da S.F., Lima, L.F.A., Scortecci, K.C., Perez, V.P., Brustolini, O.J., Sousa, E.S.S., Secco, D.A., Santos, A.M.G., Albuquerque, G.R., Mariano, A.P.M., Maciel, B.M., Gerber, A.L., Guimarães, A.P. de C., Nascimento, P.R., Neto, F.P.F., Gadelha, S.R., Porto, L.C., Campana, E.H., BezerraJeronimo, S.M., Vasconcelos, A.T.R., 2021. Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil. Preprint from medRxiv. doi: 10.1101/2021.03.21.21253418.
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 581(7807):215-220. doi: 10.1038/s41586-020-2180-5.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat*

Methods. 9(4):357-9. doi: 10.1038/nmeth.1923.

Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian Phylogeography Finds Its Roots 5. PLoS Comput Biol. 5(9):e1000520. doi: 10.1371/journal.pcbi.1000520.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078-9. doi: 10.1093/bioinformatics/btp352.

Martin, D.P., Weaver, S., Tegally, H., San, E.J., Shank, S.D., Wilkinson, E., Giandhari, J., Naidoo, S., Pillay, Y., Singh, L., Lessells, R.J., (COG-UK), C.-19 G.U., Gupta, R.K., Wertheim, J.O., Nekturenko, A., Murrell, B., Harkins, G.W., Lemey, P., MacLean, O.A., Robertson, D.L., Oliveira, T. de, Pond, S.L.K., 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. 2021.02.23.21252268. doi: 10.1101/2021.02.23.21252268.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., Lanfear, R., Teeling, E., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 37(5):1530-1534. doi: 10.1093/molbev/msaa015.

Moreira, Filipe Romero Rebello, Bonfim, D.M., Geddes, V.E.V., Zauli, D.A.G., Silva, J. do P., Lima, A.B. de, Malta, F.S.V., Ferreira, A.C. de S., Pardini, V.C., Queiroz, D.C., Souza, R.M. de, Araújo, J.L.F. de, Alves, H.J., Silva, A.V.F.G., Resende, G.G., Menezes, A.L. de, Oliveira, E.S. de, Oliveira, J.S. de, Teixeira, M.M., Luiz, L.M., Gomez, R.S., Fonseca, P.L.C., Moreira, R.G., Tanuri, A., Souza, W.M. de, Faria, N.R., Voloch, C.M., Souza, R.P. de, Aguiar, R.S., 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. [Internet]. 14 May 2021. Available from: <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>.

Moreira, Filipe R R, Bonfim, D.M., Zauli, D.A.G., Silva, J.P., Lima, A.B., Ferreira, A.C.S., Pardini, V.C., Magalh, W.C.S., Queiroz, D.C., Souza, R.M., Geddes, V.E., V, Costa, W.C., Moreira, R.G., Faria, N.R., Voloch, C.M., Souza, R.P., Aguiar, R.S., 2021b. Epidemic Spread of SARS-CoV-2 Lineage B.1.1.7 in Brazil. *Viruses*. 13(6):984. doi: 10.3390/v13060984.

Naveca, F.G., Nascimento, V., de Souza, V.C., Corado, A.D.L., Nascimento, F., Silva, G., Costa, Á., Duarte, D., Pessoa, K., Mejía, M., Brandão, M.J., Jesus, M., Gonçalves, L., da Costa, C.F., Sampaio, V., Barros, D., Silva, M., Mattos, T., Pontes, G., Abdalla, L., Santos, J.H., Arantes, I., Dezordi, F.Z., Siqueira, M.M., Wallau, G.L., Resende, P.C., Delatorre, E., Gräf, T., Bello, G., 2021. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nat Med*. doi: 10.1038/s41591-021-01378-7.

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., Plessis, L., Aanensen, D., Holmes, E., Pybus, O., Rambaut, A., 2020. Pangolin: lineage assignment in an emerging pandemic as an epidemiological tool. [github.com/cov-lineages/pangolin](https://github.com/cov-lineages/pangolin).

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., Burton, D.R., Lewis-

- Ximenez, L.L., De Jesus, J.G., Giovanetti, M., Hill, S.C., Black, A., Bedford, T., Carroll, M.W., Nunes, M., Alcantara, L.C., Sabino, E.C., Baylis, S.A., Faria, N.R., Loose, M., Simpson, J.T., Pybus, O.G., Andersen, K.G., Loman, N.J., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc.* 12(6):1261-1276. doi: 10.1038/nprot.2017.066.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics.* 26(6):841-2. doi: 10.1093/bioinformatics/btq033.
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ramanathan, M., Ferguson, I.D., Miao, W., Khavari, P.A., 2021. SARS-CoV-2 B.1.1.7 and B.1.351 spike variants bind human ACE2 with increased affinity. *Lancet Infect Dis.* S1473-3099(21)00262-0. doi: 10.1016/S1473-3099(21)00262-0.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 67(5):901-904. doi: 10.1093/sysbio/syy032.
- Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 5(11):1403-1407. doi: 10.1038/s41564-020-0770-5.
- Resende, P.C., Gräf, T., Paixão, A.C.D., Appolinario, L., Lopes, R.S., Mendonça, A.C. da F., da Rocha, A.S.B., Motta, F.C., Neto, L.G.L., Khouri, R., de Oliveira, C.I., Santos-Muccillo, P., Bezerra, J.F., Teixeira, D.L.F., Riediger, I., Debur,

- M.D.C., Ribeiro-Rodrigues, R., Leite, A.B., Do Santos, C.A., Gregianini, T.S., Fernandes, S.B., Bernardes, A.F.L., Cavalcanti, A.C., Miyajima, F., Sachhi, C., Mattos, T., da Costa, C.F., Delatorre, E., Wallau, G.L., Naveca, F.G., Bello, G., Siqueira, M.M., 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(5):724. doi: 10.3390/v13050724.
- Resende, P.C., Gräf, T., Neto, L.G.L., Silva, F.V. da, Paixão, A.C.D., Appolinario, L., Lopes, R.S., Mendonça, A.C. da F., Rocha, A.S.B. da, Motta, F.C., Delatorre, E., Wallau, G.L., Naveca, F.G., Bello, G., Network, M.M.S.C.-19 G.S., 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*. [Internet]. 14 May 2021. Available from: <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>.
- Santos, A.A. de S., Candido, D. da S., Souza, W.M. de, Buss, L., Li, S.L., Pereira, R.H.M., Wu, C.-H., Sabino4, E.C., Faria, N.R., 2021. Dataset on SARS-CoV-2 non-pharmaceutical interventions in Brazilian municipalities. *Sci Data*. 8(1):73. doi: 10.1038/s41597-021-00859-1.
- SES/MA, 2021. Secretaria Estadual de Saúde do estado do Maranhão, Brasil. [Internet]. 20 May 2021. Available from: <https://www.saude.ma.gov.br/destaques/nota-variante-indiana/>
- Slavov, S.N., Patané, J.S.L., Bezerra, R. dos S., Giovanetti, M., Fonseca, V., Martins, A.J., Viala, V.L., Rodrigues, E.S., Santos, E.V. dos, Barros, C.R. dos S., Marqueze,

- E.C., Santos, B., Aburjaile, F., Neto, R.M., Moretti, D.B., Haddad, R., Calado, R.T., Kitajima, J.P., Freitas, E., Schlesinger, D., Alcantara, L.C.J. de, Elias, M.C., Sampaio, S.C., Kashima, S., Covas, D.T., 2021. Genomic monitoring unveil the early detection of the SARS-CoV-2 B.1.351 lineage (20H/501Y.V2) in Brazil. Preprint from medRxiv. doi: 10.1101/2021.03.30.21254591.
- Souza, W.M. De, Buss, L.F., Candido, S., Carrera, J., Li, S., Zarebski, A.E., Henrique, R., Pereira, M., Jr, C.A.P., Souza-santos, A.A. De, Parag, K. V, Belotti, M.C.T.D., Vincenti-gonzalez, M.F., Messina, J., Cristina, F., Andrade, S., Nascimento, V.H., Ghilardi, F., Abade, L., Gutierrez, B., Kraemer, M.U.G., Braga, C.K. V, Aguiar, R.S., Alexander, N., Mayaud, P., Brady, O.J., Marcilio, I., Gouveia, N., Li, G., Tami, A., 2020. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat Hum Behav.* 4(8):856-865. doi: 10.1038/s41562-020-0928-4.
- Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4(1):vey016. doi: 10.1093/ve/vey016.
- Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Am. Math. Soc. Lect. Math. Life Sci.* doi: citeulike-article-id:4801403.
- Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, M., Korsman, S., Davies, M.A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver,



S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. Preprint from medRxiv. doi: 10.1101/2020.12.21.20248640.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, N. yuan, Korsman, S., Davies, M.A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 592(7854):438-443. doi: 10.1038/s41586-021-03402-9.

Vogels, C.B.F., Breban, M.I., Ott, I.M., Alpert, T., Petrone, M.E., Watkins, A.E., Kalinich, C.C., Earnest, R., Rothman, J.E., Jesus, J.G. de, Claro, I.M., Ferreira, G.M., Crispim, M.A.E., Network, B.-U.C.G., Singh, L., Tegally, H., Anyaneji, U.J., Africa, N. for G.S. in S., Hodcroft, E.B., Mason, C.E., Khullar, G., Metti, J., Dudley, J.T., MacKay, M.J., Nash, M., Wang, J., Liu, C., Hui, P., Murphy, S., Neal, C., Laszlo, E., Landry, M.L., Muyombwe, A., Downing, R., Razeq, J., Oliveira, T. de, Faria, N.R., Sabino, E.C., Neher, R.A., Fauver, J.R., Grubaugh, N.D., 2021. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol.* 19(5):e3001236. doi: 10.1371/journal.pbio.3001236.

Voloch, C.M., da Silva Francisco, R., de Almeida, L.G.P., Cardoso, C.C., Brustolini, O.J., Gerber, A.L., Guimarães, A.P. de C., Mariani, D., da Costa, R.M., Ferreira, O.C.,

- Cavalcanti, A.C., Frauches, T.S., de Mello, C.M.B., Leitão, I. de C., Galliez, R.M., Faffe, D.S., Castiñeiras, T.M.P.P., Tanuri, A., de Vasconcelos, A.T.R., 2021. Genomic Characterization of a Novel SARS-CoV-2 Lineage from Rio de Janeiro, Brazil. *J Virol.* 95(10):e00119-21. doi: 10.1128/JVI.00119-21.
- Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N.J., McCrone, J.T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D.K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D.P., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2021. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature.* 593(7858):266-269. doi: 10.1038/s41586-021-03470-x.
- WHO, 2020. WHO Director-General's opening remarks at the media briefing on COVID19. [Internet]. 11 March 2020. Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. *Nature.* 579(7798):265-269. doi: 10.1038/s41586-020-2008-3.
- Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol.* 39(3):306-14. doi: 10.1007/BF00160154.



### **Data availability:**

All consensus genome sequences characterized in this study have been deposited on GISAID (Epicodes: EPI\_ISL\_2629675 to 2629820). Associated metadata are available in **Supplementary File S2**.

### **Competing interests**

The authors declare no competing interests.

### **Acknowledgements:**

We 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 S3**. We 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. Further, we would like to acknowledge Mr. Rafael Estima and Druid for their help on handling computational resources that made our analyses possible. Finally, we would also like to thank the Instituto Nacional do Cancer staff for the technical support on part of our sequencing runs.

**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 Zalcborg 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, Victoria Cortes Bastos.

## Funding:

This study was supported by 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), CNPq (A.F.A.S.: 313005/2020-6; R.S.A.: 312688/2017-2, 439119/2018-9), CAPES (R.S.A.: 14/2020 - 23072.211119/2020-10) and FINEP (R.S.A: 0494/20 01.20.0026.00). We also acknowledge support from the Rede Coronavírica BR MCTI/FINEP affiliated to RedeVírus/MCTI (FINEP 01.20.0029.000462/20, CNPq 404096/2020-4).

## Figures and Table:

**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. The 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). Bar plots exhibit in absolute (B) and relative (C) 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. Lineages are color labelled as in the phylogenetic tree, plus: B.1.1.28 (green) and B.1.1.33 (yellow).

**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), RJ (red), South (blue) and Southeast (yellow). The dashed red lines on phylogeographic reconstructions and time series highlight: 1- the time of the most recent common ancestor (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% (only in time series).

**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% CI =  $\pm 0.70$ ).

**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.

**Supplementary Material:**

**Supplementary Figure S1:** Phylogeographic reconstructions using a model based on non-parametric coalescent skygrid tree prior. Inner panel on the right showing introductions that led to the emergence of local VOC Gamma clades in Rio de Janeiro.

**Supplementary Figure S2:** Demographic analysis based on the coalescent exponential growth model and non-parametric coalescent skygrid model. Both models imply an initial increase in relative genetic diversity, matching the reported rise in frequency of lineage P.1 across the country.

**Supplementary Figure S3:** Time series of Ct values for all positive samples in our cohort. Violin plot displaying the distribution of Ct values for N1 (A) and N2 (B) viral targets and RNaseP endogenous control (C). The dashed red line represents the median Ct value observed in December 2020.

**Supplementary Table S1:** Results from BEAST models based on the coalescent exponential growth and non-parametric coalescent skygrid tree priors.

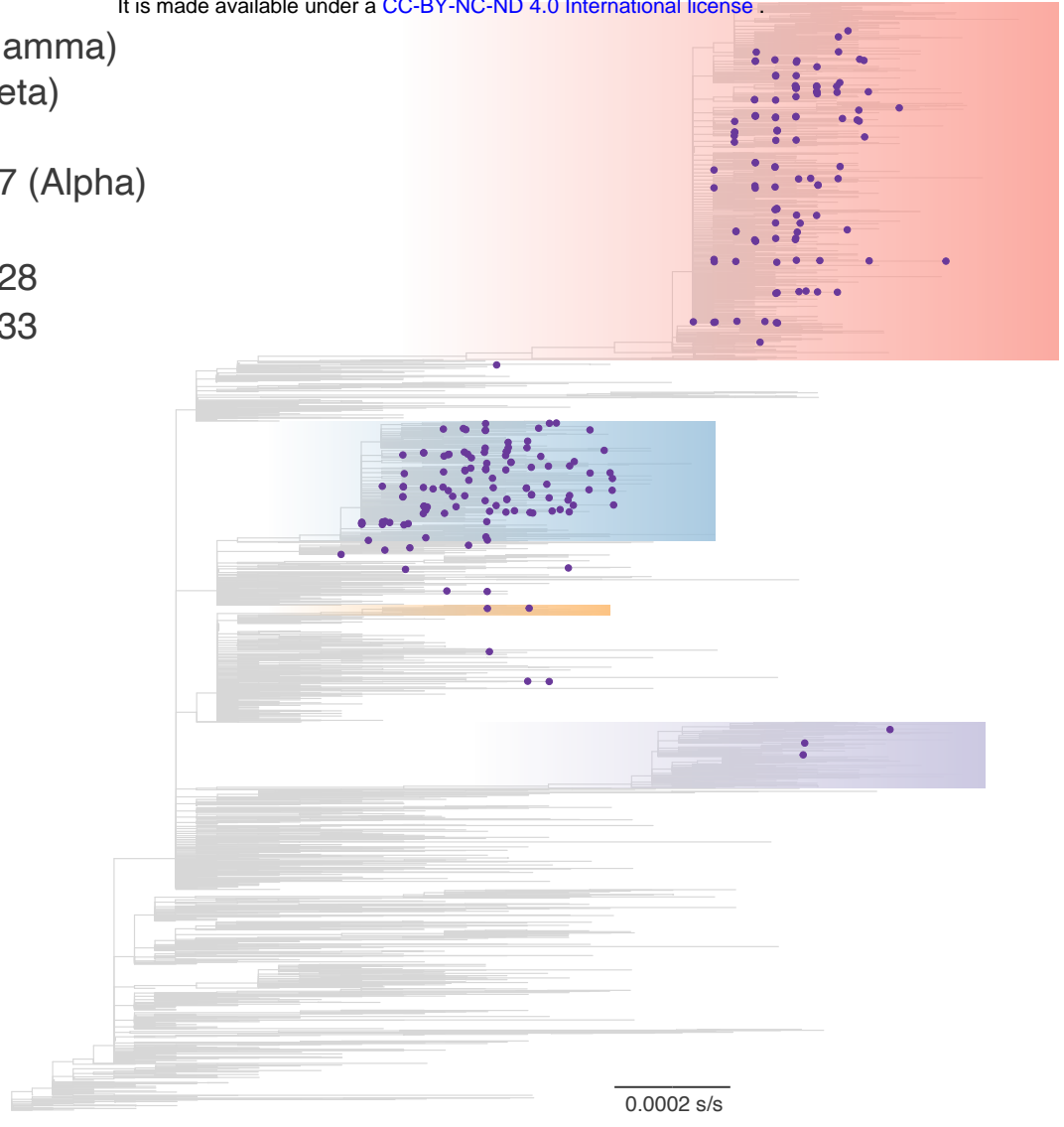
**Supplementary File S1:** BEAST maximum clade credibility trees and log files.

**Supplementary File S2:** Metadata and associated sequencing statistics for samples characterized in this study.

**Supplementary File S3:** GISAID acknowledgement table.

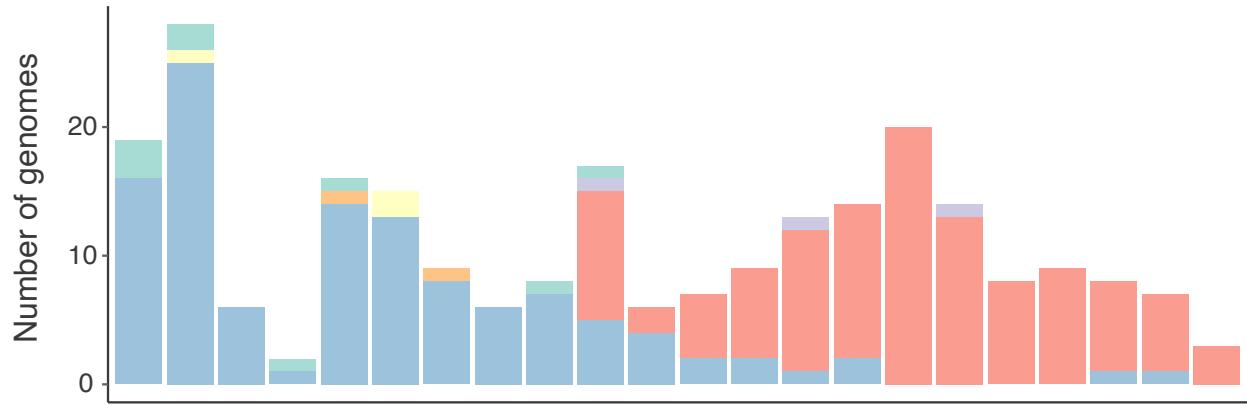
**A**

- P.1 (Gamma)
- P.2 (Zeta)
- N.9
- B.1.1.7 (Alpha)
- B.1.1.28
- B.1.1.33

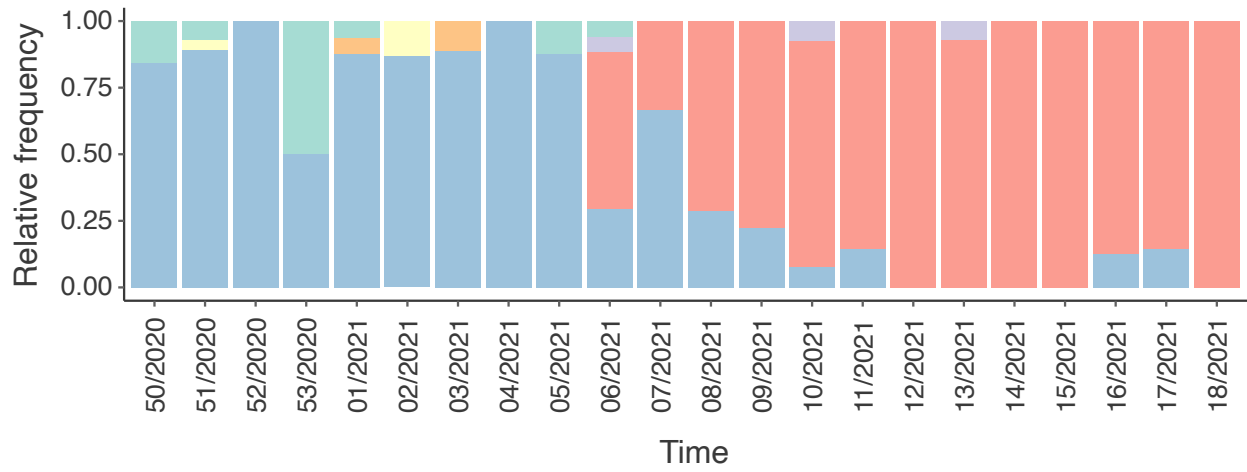


**B**

Lineage frequency per epidemiological week

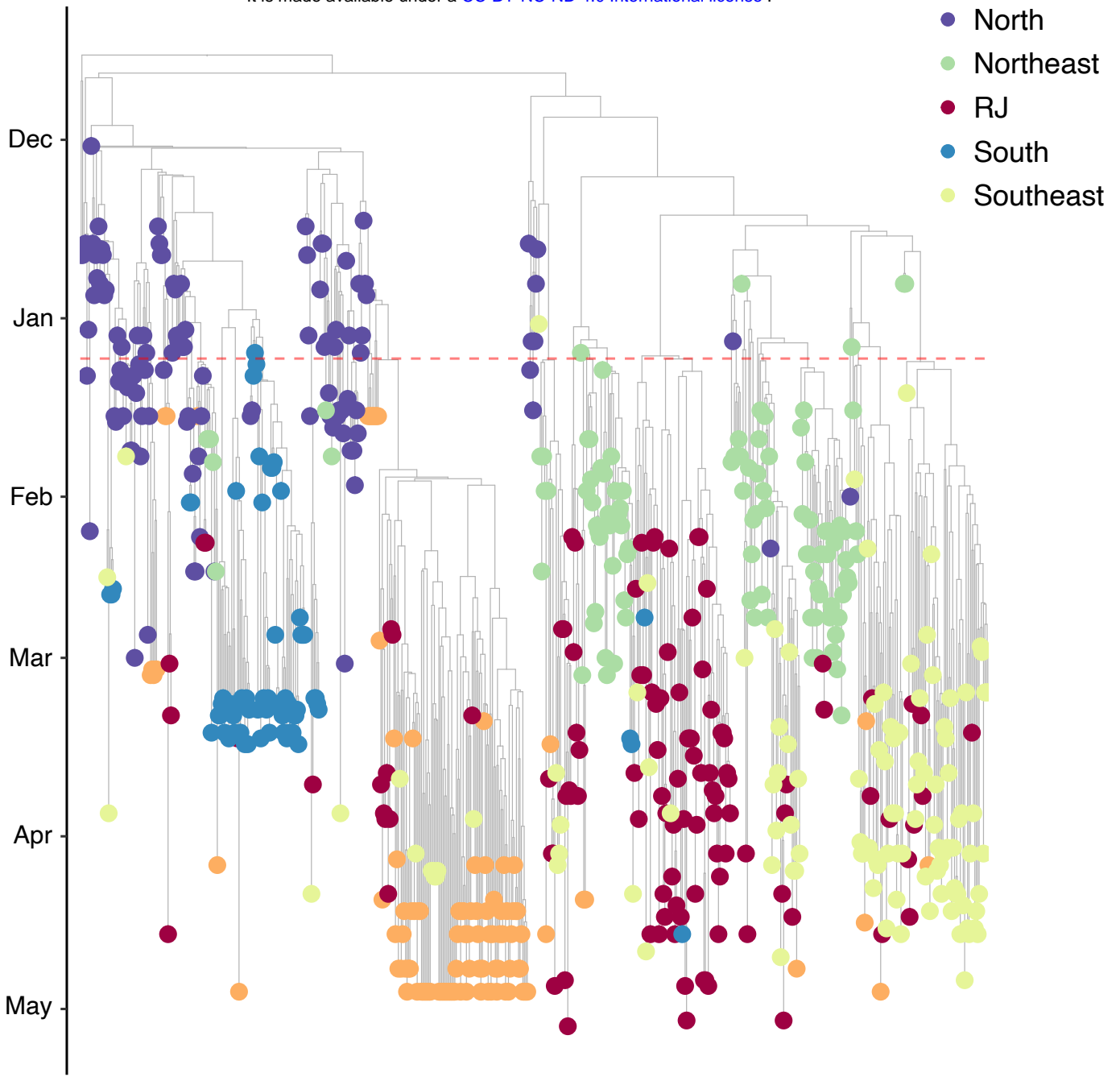


**C**



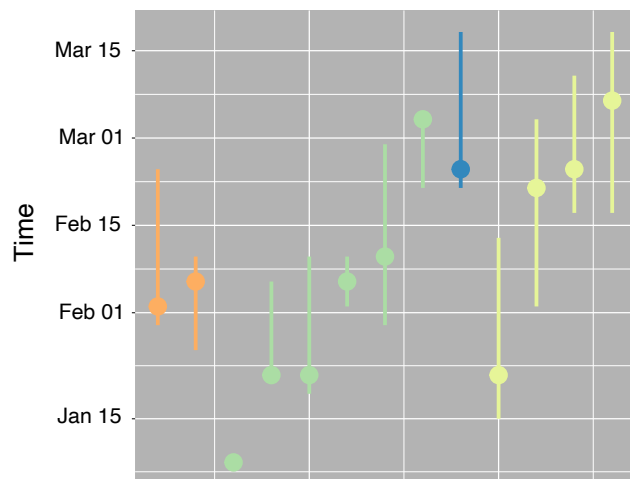


**A**



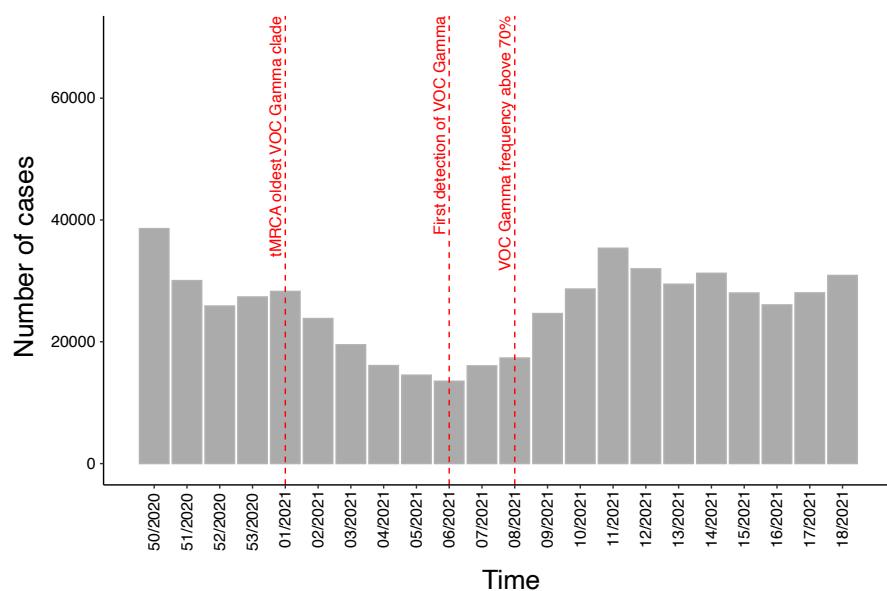
**B**

Introductions in Rio de Janeiro



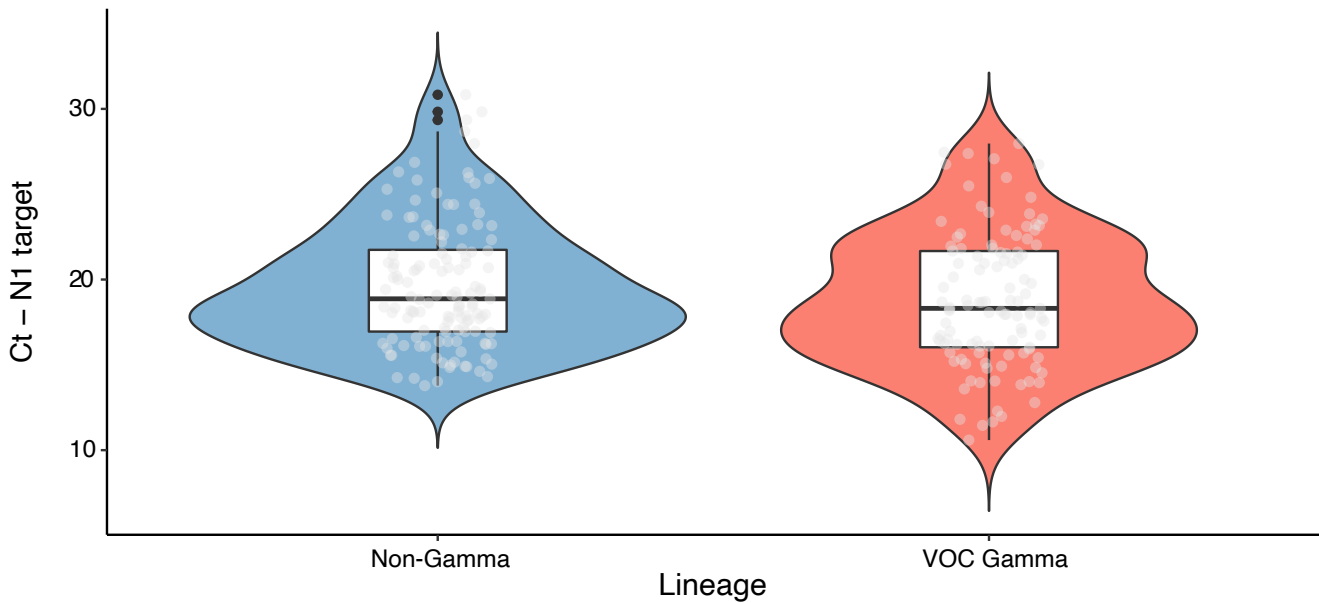
**C**

Number of cases in Rio de Janeiro municipality



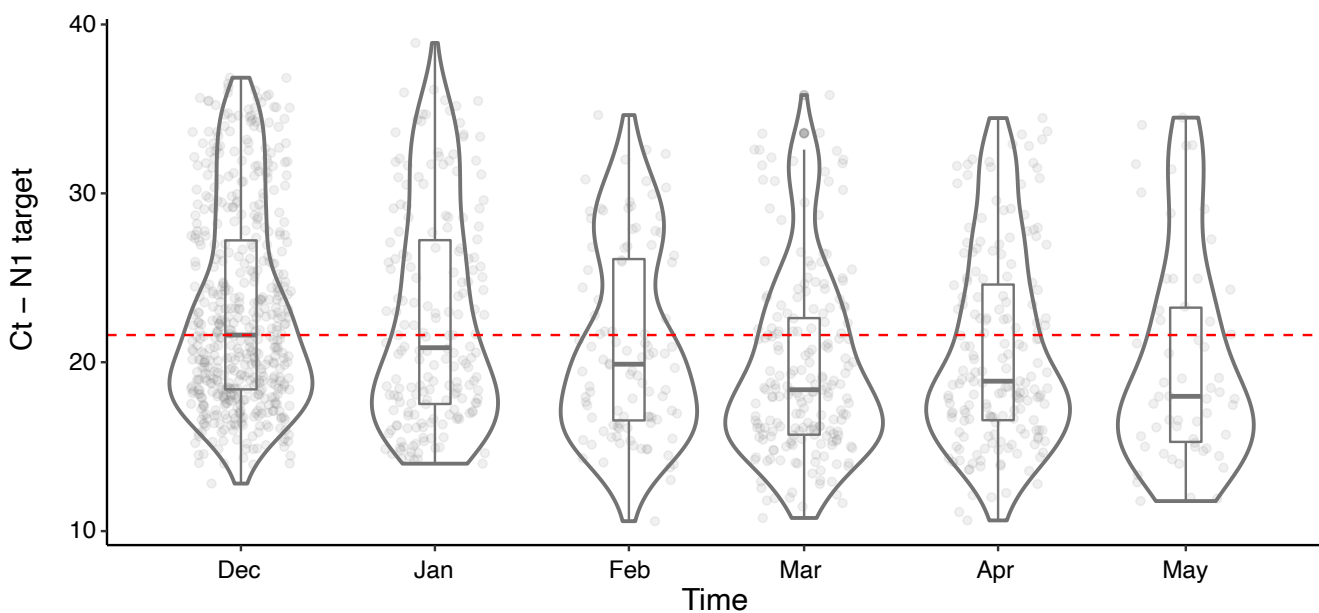
**A**

### Gamma vs non-Gamma samples



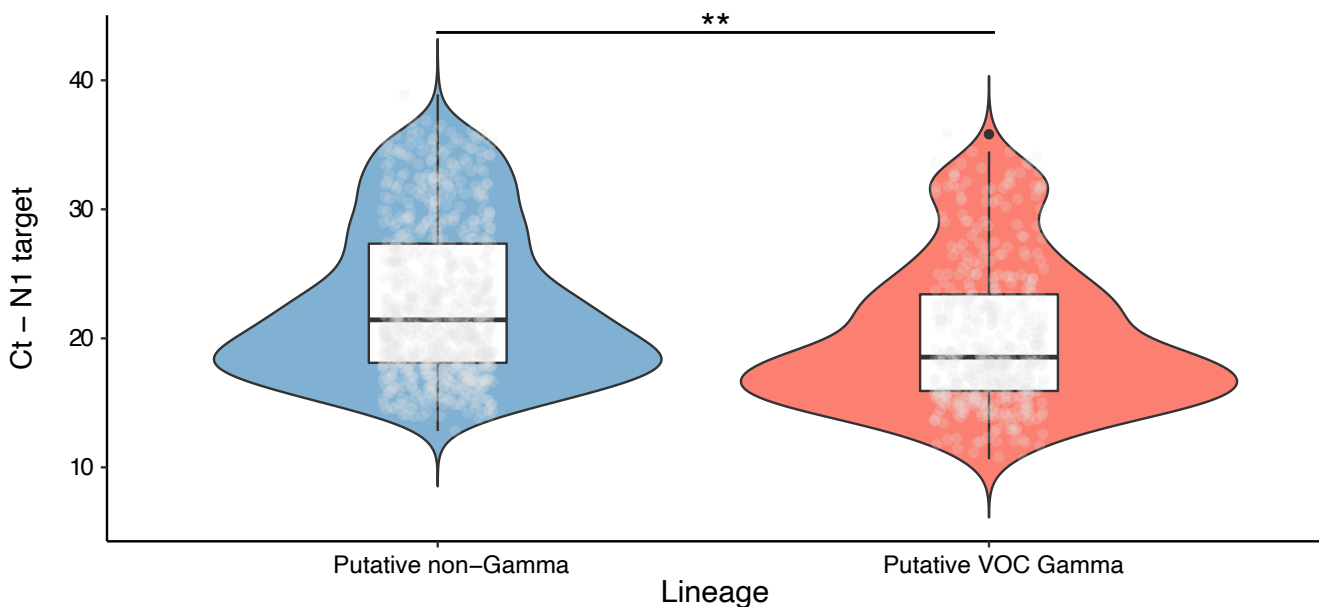
**B**

### Time series of Ct values - N1 target



**C**

### Difference of Ct values between putative Gamma vs non-Gamma samples



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

<b>Model</b>	<b>Parameters</b>	<b><i>p</i> - value</b>	<b><math>\beta</math></b>	<b>95% CI</b>
m1	Days of symptoms	<i>&lt;0.01</i>	<i>0.48</i>	<i>±0.16</i>
m2	Age	<i>&lt;0.01</i>	<i>-0.05</i>	<i>±0.04</i>
m3	Sex	0.14	-	-
M1	Lineage	0.09	-	-
M2	Lineage	0.33	-	-
	Days of symptoms	<i>&lt;0.01</i>	<i>0.47</i>	<i>±0.16</i>
M3	Lineage	0.07	-	-
	Age	<i>&lt;0.01</i>	<i>-0.05</i>	<i>±0.03</i>
M4	Lineage	0.29	-	-
	Days of symptoms	<i>&lt;0.01</i>	<i>0.46</i>	<i>±0.16</i>
	Age	<i>&lt;0.01</i>	<i>-0.05</i>	<i>±0.03</i>

Statistically significant values highlighted in italic.