Article

Turnover of SARS-CoV-2 Lineages Shaped the Pandemic and Enabled the Emergence of New Variants in the State of Rio de Janeiro, Brazil

Ronaldo da Silva Francisco Junior [1,†], Alessandra P Lamarca [1,†], Luiz G P de Almeida [1], Liliane Cavalcante [1], Douglas Terra Machado [1], Yasmmin Martins [1], Otávio Brustolini [1], Alexandra L Gerber [1], Ana Paula de C Guimarães [1], Reinaldo Bellini Gonçalves [1], Cassia Alves [2], Diana Mariani [2], Thais Felix Cruz [2], Isabelle Vasconcellos de Souza [3], Erika Martins de Carvalho [3], Mario Sergio Ribeiro [4], Silvia Carvalho [4], Flávio Dias da Silva [5], Márcio Henrique de Oliveira Garcia [5], Leandro Magalhães de Souza [6], Cristiane Gomes da Silva [6], Caio Luiz Pereira Ribeiro [5], Andréa Cony Cavalcanti [6], Claudia Maria Braga de Mello [4], Cláudio J. Struchiner [7], Amilcar Tanuri [2,‡] and Ana Tereza R Vasconcelos [1,*,‡]

[1] Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Petrópolis 25651-076, Brazil; ronaldoj@lncc.br (R.d.S.F.J.); pavanlamarca@gmail.com (A.P.L.); lgonzaga@lncc.br (L.G.P.d.A.); liliane.tavaresdefaria@gmail.com (L.C.); douglast@lncc.br (D.T.M.); yasmmin@lncc.br (Y.M.); tavinbio@lncc.br (O.B.); alegerber@lncc.br (A.L.G.); anapcg@lncc.br (A.P.d.C.G.); reinaldo.bellini@gmail.com (R.B.G.)

[2] Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-570, Brazil; cassia.alves@gmail.com (C.A.); mariani.diana@gmail.com (D.M.); thafe01@gmail.com (T.F.C.); atanuri1@gmail.com (A.T.)

[3] Unidades de Apoio ao Diagnóstico da COVID-19, Rio de Janeiro 21040-900, Brazil; isabelle.souza@fiocruz.br (I.V.d.S.); erika.decarvalho@fiocruz.br (E.M.d.C.)

[4] Secretaria Estadual de Saúde do Rio de Janeiro, Rio de Janeiro 20031-142, Brazil; mario_sesrj@yahoo.com.br (M.S.R.); silviac_carvalho@yahoo.com.br (S.C.); claudiamello.ses@gmail.com (C.M.B.d.M.)

[5] Secretaria Municipal de Saúde Rio de Janeiro, Rio de Janeiro 20211-901, Brazil; flaviodiass@outlook.com (F.D.d.S.); marciogarcia.smsrio@gmail.com (M.H.d.O.G.); oiac17@hotmail.com (C.L.P.R.)

[6] Laboratório Central de Saúde Pública Noel Nutels, Rio de Janeiro 20231-092, Brazil; magalhaessouza@yahoo.com (L.M.d.S.); crisarthurpharma@gmail.com (C.G.d.S.); dg@lacen.fs.rj.gov.br (A.C.C.)

[7] Fundação Getúlio Vargas, Rio de Janeiro 22250-900, Brazil; claustru@gmail.com

[*] Correspondence: atrv@lncc.br

[†] Both authors contributed equally to this work.

[‡] Both authors coordinated this work.

Abstract: In the present study, we provide a retrospective genomic epidemiology analysis of the SARS-CoV-2 pandemic in the state of Rio de Janeiro, Brazil. We gathered publicly available data from GISAID and sequenced 1927 new genomes sampled periodically from March 2021 to June 2021 from 91 out of the 92 cities of the state. Our results showed that the pandemic was characterized by three different phases driven by a successive replacement of lineages. Interestingly, we noticed that viral supercarriers accounted for the overwhelming majority of the circulating virus (>90%) among symptomatic individuals in the state. Moreover, SARS-CoV-2 genomic surveillance also revealed the emergence and spread of two new variants (P.5 and P.1.2), firstly reported in this study. Our findings provided important lessons learned from the different epidemiological aspects of the SARS-CoV-2 dynamic in Rio de Janeiro. Altogether, this might have a strong potential to shape future decisions aiming to improve public health management and understanding mechanisms underlying virus dispersion.

Keywords: surveillance; P.1.2; P.5; COVID-19; supercarriers; lineage replacement

1. Introduction

One and a half years have passed since the outbreak of the Severe Acute Respiratory Syndrome-related Coronavirus 2 (SARS-CoV-2) virus in late 2019 in Wuhan, China. Currently, SARS-CoV-2 has accumulated a myriad of mutations and diversified into more than a thousand recognized lineages [1]. While vaccination is advancing worldwide and different treatments are being tested, the imminent emergence of resistant variants haunts the scientific community. Lineages such as Gamma (P.1), Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) are already known to partially evade natural- and vaccine-induced antibody response as well as have an elevated transmission rate [2–4]. These Variants of Concern (VoC), named by the World Health Organization (WHO), have become widespread in the world, replacing other lineages [5–7].

The emergence of novel SARS-CoV-2 variants is particularly worrying in countries with slow vaccination, such as Brazil. Since March 2020, the COVID-19 pandemic has caused over 19 million cases and the loss of 530,000 lives so far. Although Brazil corresponds to approximately 2.7% of the total world population, it accounts for more than 12.5% of deaths globally. By July 2021, only 15% of the Brazilian population has completed the vaccination program recommended. This small number allows the virus to circulate freely within the population, enabling the emergence of mutations and further diversification into new lineages. Previous studies analyzing the spatial progression of the pandemic in Brazil revealed that the states of São Paulo and Rio de Janeiro were pivotal to the dissemination of lineages throughout the country, including VoCs [8–11]. Rio de Janeiro was also the background to the emergence and worldwide spread of the Variants of Interest (VoI) Zeta (P.2) [12].

The high diversity of SARS-CoV-2 lineages in Rio de Janeiro since the beginning of the pandemic and the presence of three VoC and one VoI is the perfect scenario to verify whether there is a relationship between lineage displacement and a surge in the number of cases. Here, we provide the first temporal analysis of the lineage dynamics in the state of Rio de Janeiro. By combining epidemiological and genomic data, we described the pandemic in three main phases in the state. In addition to genomes publicly available in GISAID, we sequenced 1927 new genomes from 91 out of the 92 cities in Rio de Janeiro, sampled periodically between March and June of 2021. Finally, we report the emergence of two new variants, P.1.2 and P.5, and reconstructed the within-country and worldwide dispersion.

2. Materials and Methods

2.1. Sampling, Genome Extraction, Sequencing, and Assembly

Samples from patients with SARS-CoV-2 positive nasopharyngeal RT-PCR were collected at the Noel Nutels Central Laboratory (LACEN-RJ) and Unidades de Apoio ao Diagnóstico da COVID-19 (UNADIG-RJ). The 1927 new genomes sequenced in this study were collected between 24 March 2021 and 6 June 2021 from 91 municipalities in Rio de Janeiro, Brazil. Patients were aged between 0 and 91 years old, being 47.69% male and 52.31% female. Extraction of the genetic material was performed at the Molecular Virology Laboratory (LVM-UFRJ) with QIAamp or MagMAX Viral/Pathogen Nucleic Acid Isolation kits and KingFisher automatic platform. Annealing of cDNA was conducted with 8.5 µL of viral RNA extracted from each sample. Libraries were constructed at the DFA/LNCC Genomics Unit with Illumina COVIDSeq Test (Illumina), according to the manufacturer's protocol. Purification was then conducted using 5 µL of each library combined, and the TapeStation (Agilent) system was used for quality control. We employed the NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles) to generate reads of 2 × 149 bp in the NextSeq (Illumina). Sequence analysis, consensus building, and variant calling were performed with DRAGEN COVID Lineage v3.5.1. The study was approved by the Ethics Committee (30161620.0.1001.5257 and 34025020.0.0000.5257). All the assembled genomes were then submitted to the GISAID database and made publicly available. Finally, the newly assembled genomes were classified according to the Pango Lineage classification using the PangoLEARN model database (v3.5.3).

2.2. Epidemiology and Viral Load Analysis

Epidemiological data were retrieved from the Centro de Informações Estratégicas e Resposta de Vigilância em Saúde (CIEVS-RJ) of the Secretaria de Saúde do Estado do Rio de Janeiro. We then estimated the time-varying reproduction numbers (Rt) from the Epidemic Curve using the EpiEstim package in R [13]. To verify whether a small group of samples can be overrepresented the Ct values in our dataset, we selected 1119 out of the 1927 samples sequenced and retrieved the cycle threshold (Ct) values of the target virus genes and the endogenous gene RNaseP. We performed a relative quantification of the viral loads (RQVL) using $2^{-\Delta CT}$. The dataset table with sample ID, sex, age, and RQVL columns was ordered by the RQLV values. We calculated the ratio of 2%, 10%, and 20% of the samples relative to the total sum of the RQVL in our dataset. We also performed a generalized linear model using the gamma distribution with the inverse link function to verify the significance of sex and age in the RQLV distribution using the stats package in R [14].

2.3. Evolutionary Analyses

To infer their phylogenetic tree, we retrieved all genomes from the state of Rio de Janeiro available in GISAID by 21 June 2021 that were longer than 29,000 bases, had high coverage (as defined by GISAID), and had complete collection dates (*n* = 3029). We used MAFFT 7.475 with –auto and –add fragments options [15,16] to align this dataset to the WH01 (EPI_ISL_406798) genome sampled in 2019 from Wuhan, China. The 3′ and 5′ untranslated regions were then trimmed. The subsequent alignment was used to construct the maximum likelihood tree using IQ-TREE 2 [17] with the GTR+F+I+G4 substitution model, selected with the ModelFinder algorithm [18]. Assuming a constant evolutionary rate of 8×10^{-4} substitutions/site/year, we rescaled the branch lengths to dates with Tree-Time [19]. Tips that deviated from the clocklike model are ignored by TreeTime (Figure S1A) and were removed from the final tree.

To reconstruct the evolution of the new P.1.2 lineage described in this work, we retrieved all sequences from this lineage available in GISAID by June 21 and aligned them to our newly sequenced samples. As an outgroup, we selected the two oldest P.1 sequences available from the state of Amazonas. We then reconstructed the maximum likelihood tree with IQ-TREE 2 and the GTR+F+I model selected with the ModelFinder algorithm. We then tested the sequences' conformity to the molecular clock model by analyzing the correlation between root-to-tip distances and tip dates and removed from the alignment all sequences that were not included within the 95% confidence interval of the regression (Figure S1B). As the two outgroup sequences selected were removed from the final dataset, we selected as new outgroups two sequences from the P.1 clade positioned closest to P.1.2 according to the maximum likelihood phylogeny of all sequences from Rio de Janeiro.

Divergence dates were estimated using a Bayesian approach with BEAST 1.10.4 [20]. Models selected for the analyses were the GTR+F+I with estimated base frequencies, the GMRF Skyride population model, and the strict clock model. The clock rate prior was defined as a uniform distribution with 6E-4 and 1E-3 as the minimum and maximum values. The MCMC was run through 200,000,000 steps with sampling every 10,000th and a burnin of 10%. Lowest ESS was 204, from the "treeLength" parameter. To evaluate the impact of biased sampling through time on population inference, we also applied the Bayesian Non-parametric Phylodynamic Reconstruction model with and without preferential sampling correction on the consensus tree of P.1.2 using the R package phylodyn [21].

The consensus tree was used as fixed topology and branch lengths to reconstruct the locality of its ancestor nodes. We used the Symmetric trait substitution model with the Bayesian Stochastic Search Variable Selection approach for rate estimation in BEAST 1.10.4. The MCMC chain was 10,000,000 steps with sampling every 10,000th and a burnin of 10%. The lowest ESS value was 166, from the "loc.indicators.Argentina.Paraguay" parameter. Ancestor locations inferred were then converted into coordinates and mapped using the SERAPHIM package [22] in R software.

We also reconstructed the evolutionary tree of the new lineage P.5. To accomplish this, we identified all the sequences available in the GISAID database by June 21 that carried the mutations that characterize the lineage. We aligned these sequences to the WH01 genome using MAFFT 7.475 with –auto and –addfragments options and trimmed the 3′ and 5′ untranslated regions. The maximum likelihood tree was then inferred using IQTREE with the GTR+F+I model, selected by the ModelFinder algorithm.

2.4. Structural Analysis of Spike Protein of P.1 and P.1.2 Lineages

The primary sequences of Spike protein from P.1 lineage and P.1.2 subclade (A262S) were submitted in the Swiss-Model web server (https://swissmodel.expasy.org/ accessed on 22 April 2021) to obtain the best templates and build their 3D structures. The stability analysis was performed using the Stability function from the FoldX tool [23]. Salt bridges were measured using the EBRIS web server [24]. The docking analysis was performed between P.1 and P.1.2 subclade structures in HDOCK [25]. We evaluated the Spike protein against ACE2 protein and four types of antibodies anti-Spike extracted from the CoV-AbDab database [26]. The human antibodies were selected based on their preferences for distinct binding locations in Spike: subunits S1, S2, S1–S2, S1-RBD, and S1-notRBD. The ACE2 structure was extracted from a complex reported in PDB with the RBD region. We compared the five complexes generated for each protein using the residues found in contact with ligand and belonging to the hotspot predicted area. In the first part of the docking analysis, we compared the amino acids contacts between the receptor (P.1/P.1.2) and the five ligands. The binding affinity of each complex was measured with the Prodigy tool [27]. We searched for hot spots in the receptors using the KFC (Knowledge-based Fast atomic density evaluation and Contacts) webserver [28]. The potential profiles considering P.1.2 subclade structure and P.1 were compared by PDB2QR [29] and APBS. Visualization and structure manipulation was conducted using PyMol.

3. Results

In early March 2020, Rio de Janeiro's Secretary of Health confirmed the first case of COVID-19 in the state. Since then, the SARS-CoV-2 virus has rapidly spread, causing more than 923,234 cases and 54,125 deaths in Rio de Janeiro only, with multiple events of the public health system collapsing. The highest prevalence of COVID-19 cases was observed among the central and north regions in the state (Supplementary Table S1). We also noticed an increased number of cases in the metropolitan region, known to have a high population density, workers commuting daily between neighboring cities, and an elevated number of interstate corporate travels. A statistically significant difference was found across the number of cases and deaths between females and males (OR 0.73, 95% CI: 0.71–0.74; p-value $< 2.2 \times 10^{-16}$; Figure S2A). Though more females got infected, male mortality was higher. Among confirmed cases, typically younger subjects ranging between 30 and 49 years old (y.o) represented the highest occurrence rate. Nevertheless, the mortality percentages surpassed 11% in patients aged >60 y.o reaching more than 79% in the oldest group (Figure S2B).

The epidemiological findings suggest three different exponentially growing phases associated with the disease in the state (Figure 1A). Firstly, the number of cases increased after the epidemic outbreak in March 2020, reaching a peak in May 2020 when it started to decline. In early November 2020, a period marked by the intensification of the pre-election campaigns, a new increase in the number of cases was detected and peaked around December 2020, rapidly dropping in January 2021. The third phase occurred right after the Carnival holiday on 16 February, with the maximum number of cases observed in March 2021. Estimating the real-time reproduction number (Rt) in the state assumed values between 0.78 and 2.40 during the epidemic, being close to 1–1.3 (Figure 1B). These values suggest a persistent transmission of SARS-CoV-2 but not in an explosive manner. Interestingly, we observed a 2-fold increase in the daily number of cases reported in the second [mean: 2509, sd: ±1177] and third [mean: 2505, sd: ±680] phases when compared

to the first one [mean: 1446, sd: ±709]. The pairwise comparison between the daily number of cases across the different phases revealed substantial differences among the first and the second phases (*p*-value: 1×10^{-13}, Wilcoxon test—Benjamini-Hochberg *p*-value correction), the first and the third phases (*p*-value: 2×10^{-16}), with no difference between the second and the third phases (*p*-value: 0.63).
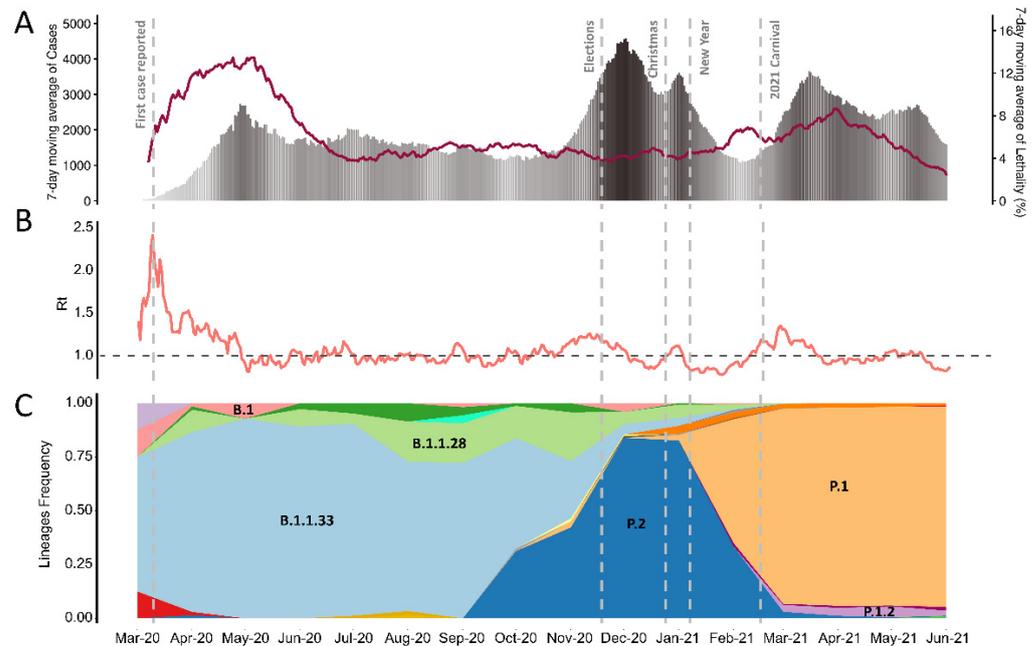

Figure 1. Epidemiological findings and successive lineage replacement during SARS-CoV-2 epidemic in the state of Rio de Janeiro. (**A**) The 7-days moving average of cases and lethality reported from 1 March 2020 to 6 June 2021. Dashed lines indicate the first reported case, elections, Christmas, new year, and carnival 2021, respectively. The degree in grayscale shows the increases in the number of cases in each phase. (**B**) Estimation of the real-time reproduction number (Rt). (**C**) Temporal distribution of the relative frequencies of SARS-CoV-2 lineages in the state of Rio de Janeiro. We only highlighted the VoC and VoI variants circulating in the state.

By comparing the mortality rate across the different phases, we observed successive increases in the mortality rate across the first and second phases (*p*-value: 2.2×10^{-4}) and the second and third stages (*p*-value: 3.2×10^{-5}). The third phase was also higher than the first one (*p*-value: 7.9×10^{-10}; Figure S2C). The differential mortality rate was also identified among the age groups (Figure S3, Table S2). The most recent phase seemed to be more lethal among subjects ranging between 20 and 69 y.o than the others. On the other hand, the mortality in the second phase was consistently decreased among young ages and increased in subjects older than 79 y.o (Figure S3).

After the introductory events in early March 2020, independent efforts to sequence SARS-CoV-2 genomes from Rio de Janeiro resulted in 3932 sequences publicly available in the GISAID database until June 2021. This number corresponded to an approximate overall average of four sequences per 1000 cases, reaching a peak of 1% of all reported cases sequenced in April and May 2021 (Table S3). We observed that the pandemic phases were characterized by successive replacement of SARS-CoV-2 lineages in the state (Figure 1C). The first phase was predominantly driven by the lineage B.1.1.33, the most prevalent strain circulating during eight months, from March to October 2020, being gradually replaced by P.2 during the second phase. In May 2020, B.1.1.33 achieved the higher relative frequency of approximately 92% of the genomes deposited in GISAID, coinciding with the peak reported during the first phase.

Moreover, the dominance of P.2, from November 2020 to January 2021, was preceded by a continuous increase in the frequency of its ancestor, B.1.1.28, that was already circulat-

ing in a minor frequency in the state since April 2020. The peak of reported cases observed in the second phase matches the highest prevalence of P.2 in December 2020. After this, P.1 became the dominant lineage reaching a frequency greater than 90% from March 2021, matching the peak of cases in the third phase. Therefore, the introduction and emergence of each lineage occurred concomitantly with the increase in the number of cases.

From the 1927 new genomes sequenced in this study, 92.42% were assigned to P.1 (Gamma) lineage. We then sought to investigate the distribution of SARS-CoV-2 viral loads within a population of 1119 samples randomly selected from the state. The cumulative distribution of the relative quantification of the viral loads (RQVL) showed that 2% ($n = 22$, group A) of the individuals carried 60% of the circulating virions in the period analyzed (Figure 2A). This value increases to 92.4% of the virions harbored by just 10% ($n = 112$, group B) of the population (Figure 2A). Furthermore, 20% ($n = 223$, group C) of our samples represented 98.2% of the viral loads of our dataset (Figure 2A). Next, we tested the contribution of age, sex, and the interaction between age and sex to the viral loads using a gamma-generalized linear model (GLM). Both groups (A, B, and C) had approximately half of the samples from each sex. Nevertheless, a higher relative viral load was reported among males within-group (Figure 2B). From the 22 samples (11 males and 11 females) present in group A, the viral load was 14.8% greater in males when compared to females. Among males in groups B and C, our model showed that subjects aged >40 y.o carried most of the population-circulating virions, and the older, the higher the viral load observed (Table S4; Figure S4). No statistical differences were observed among females age groups.

SARS-CoV-2 genomic surveillance in the state of Rio de Janeiro also allowed the identification of two new variants (P.5 and P.1.2) first reported in this study (Figure 3A). The new variant assigned as P.5 by PANGO descends from B.1.1.28, with the possible origin being the Brazilian state of São Paulo. P.5 harbors 12 nonsynonymous lineage-defining mutations (ORF1ab: T1637I, A3209V, Q3729K, and P4337L; S: F2L, Q14K, T95I, E484Q, and N501T; N: G215V) and by 21 June 2021 was restricted to Rio de Janeiro and São Paulo (Figure S5). P.1.2 derived from P.1 and was characterized by a constellation of five new mutations (ORF1ab: synC1912T, D762G and T1820I; ORF3a: D155Y; N: synC28789T) besides the P.1 lineage-defining mutations. This variant has already spread to Argentina, Australia, Chile, England, The Netherlands, Paraguay, Portugal, Spain, Uruguay, and the United States. The mutational signature of the lineages reported in this work is summarized in Table S5. We estimated the first divergence within P.1.2 to have occurred within the Brazilian state of São Paulo between late 2020 and early 2021 (Figure 3B). By the end of January, the lineage had already spread to all Brazilian states sampled, including an event of secondary dispersal (Goiás to Maranhão) and a reintroduction to São Paulo from Sergipe (Figure 3C). In late January, the lineage was first introduced in most countries; it arrived in Spain and Paraguay only in early March (Figure 3D). This early spread coincides with a rise in the effective population size of P.1.2 lineage (Figure S6A), though this value stabilizes around late February. Temporal bias does not seem to have influenced the inference of this fluctuation (Figure S6B,C).

P.1.2 was already diversified in a subclade that emerged in Rio de Janeiro around mid-February, being the subtype more frequently detected in the state. This subclade carries the additional mutations A262S in the Spike protein gene and the missense variant L83F in ORF3a. To further characterize the impact of A262S mutation, we performed a comparative structural analysis between the Spike protein from P.1 and the subclade of P.1.2. The 3-dimensional structure of P.1 and P.1.2 (A262S) was modeled by selecting the 7KRS model as the PDB template. This model has 3.2Å of crystal resolution with 99.09% and 99.14% of identity to P.1 and P.1.2, respectively. Quality assessment analysis showed that both predictions achieved 93.9% of atoms contained in P.1 and 93.84% in P.1.2. Despite the few differences in RMSD values (P.1: 0.11 and P.1.2: 0.14), stability analysis showed that P.1.2 is less stable than P.1 according to FoldX (Table S6). This difference was confirmed by estimating the number of salt bridges in both models. P.1 had more salt bridges

(n = 53) than P.1.2 (n = 48), demonstrating that P.1 folding is less flexible. We also investigated possible effects in the electrostatic surface caused by the change of an Alanine to Serine in position 262. Figure S7A,B shows that both proteins present the same potential surface. However, in the mutated region, the cavity becomes more neutral in P.1.2, possibly altering the binding site of the region. P.1.2 was more efficient than P.1 in interaction with different antibodies according to our affinity binding prediction analysis (Table S7; Figure S7). P.1 also performs more stable bonds with ACE2 (-18.2 kcal/mol) than P.1.2 (-17.8 kcal/mol) and is predicted to enter easier in the cell to establish virus infection. Furthermore, we detected an active binding site in the complex S1notRBD-P.1.2, a candidate to alter the contacts in the cavity (Figure S7).
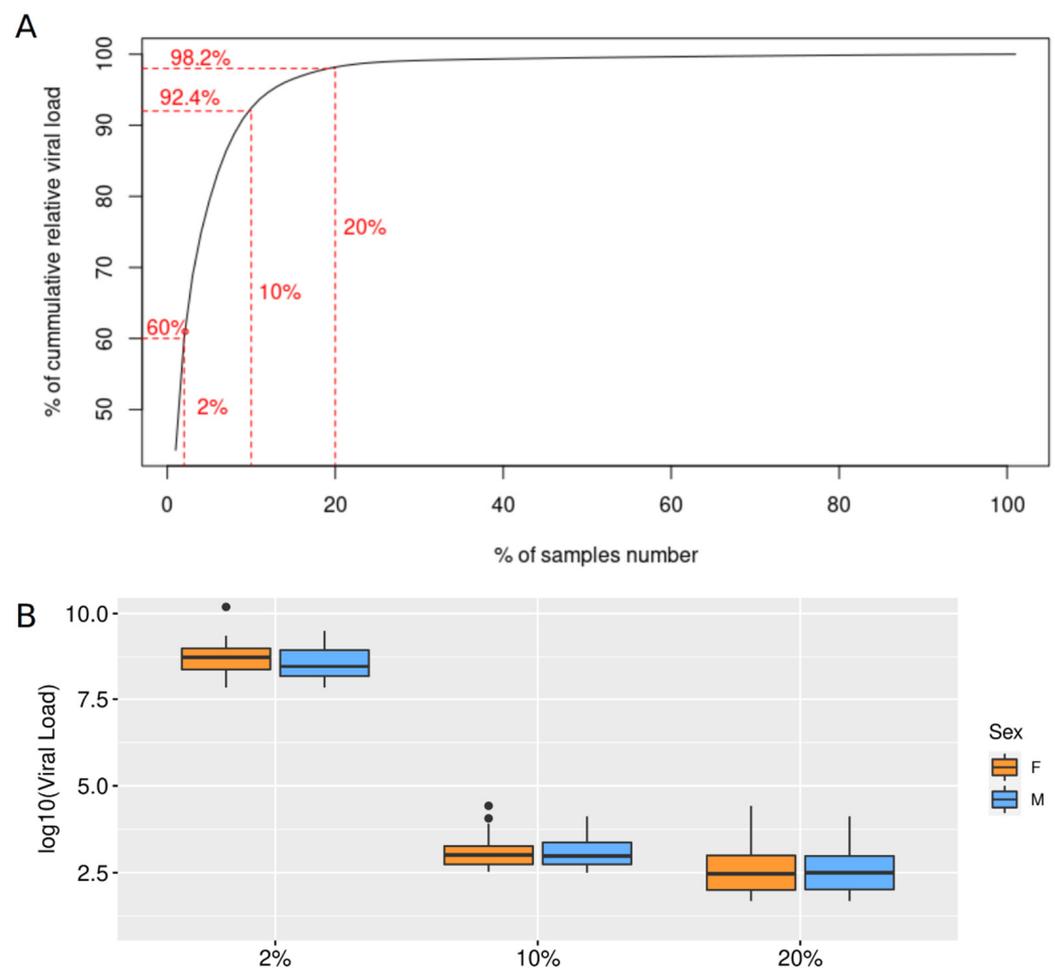

Figure 2. Distribution of the relative quantification of the SARS-CoV-2 viral loads (RQVL) within our population of 1119 samples. (**A**) Cumulative percentage of total circulating virions from nasopharynx RQVL within our population. (**B**) RQVL percentile comparative analysis between female (orange) and male (blue) in groups A (2%), B (10%), and C (20%), respectively. The gamma-generalized linear model showed a statistical significance difference among the sex of the supercarriers in each group, with a higher relative viral load reported among males, as demonstrated in Table S4.
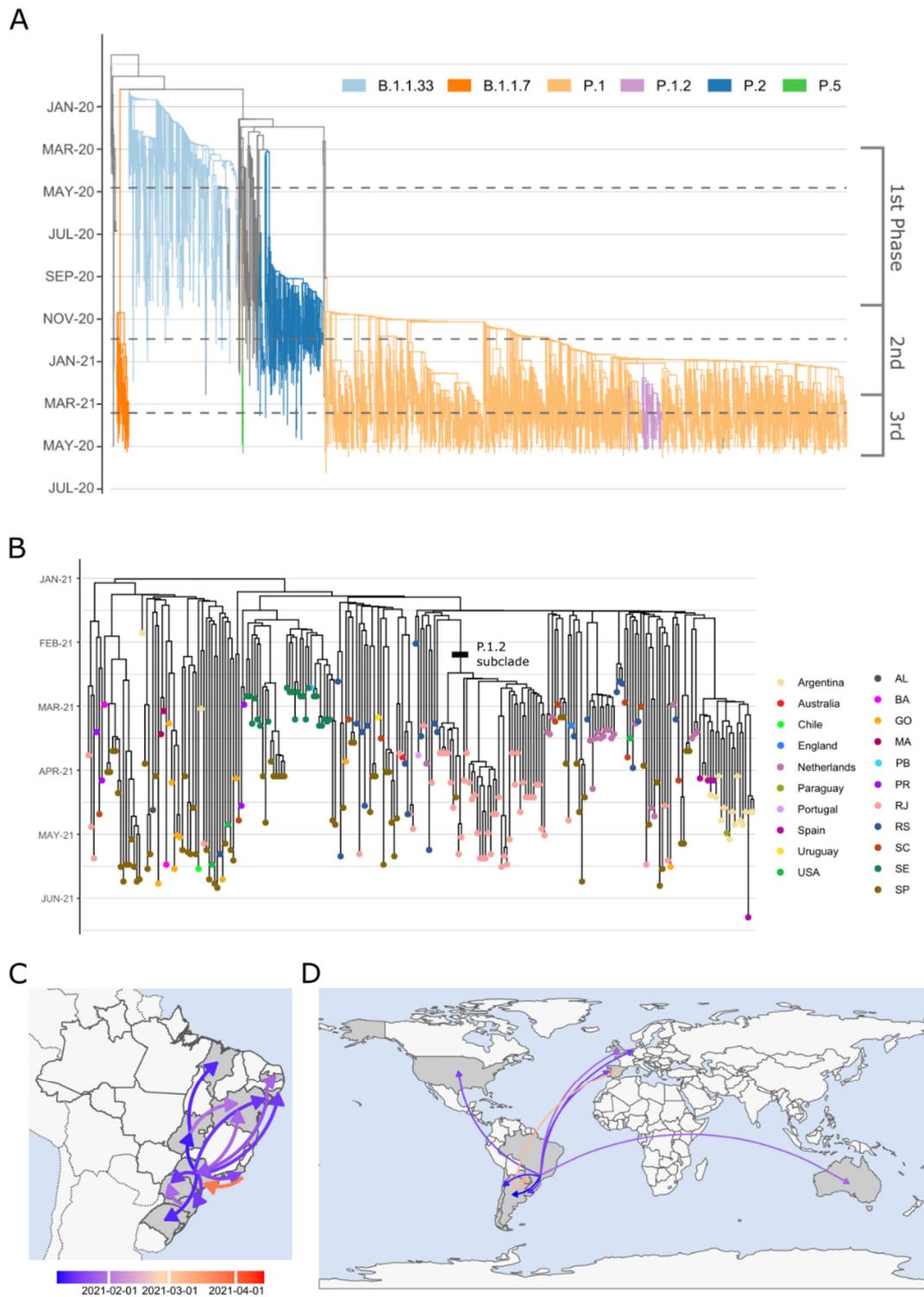
Figure 3. Evolutionary relationships between major lineages present in Rio de Janeiro and the evolution and dispersal of the new lineage P.1.2. (**A**) Time tree of 3029 genomes from the state of Rio de Janeiro, with lineages of interest highlighted by branch color. (**B**) Divergence dates between P.1.2 sequences indicate that the lineage originated at the first days of 2021. Black bar signals the subclade with A262S mutation in Spike. Colored dots show the location of sampling. (**C,D**) National and international transmission routes inferred for the dispersion of P.1.2. The color of the arrows corresponds to the date of the first dispersal event between locations.

4. Discussion

In this study, we gathered epidemiological and genomic data from the outbreak of SARS-CoV-2 until June 2021. We also conducted a thorough examination of the pandemic progression in the state of Rio de Janeiro, Brazil. Before this work, the state faced issues with keeping continuous genomic surveillance resulting in different proportions of positive samples being sequenced over time. To overcome the limitation in the number of sequences available for the state, we generated 1927 novel SARS-CoV-2 genomes sequenced biweekly between March and June 2021, sampled were originated from residents of 91 out of the 92 cities of Rio de Janeiro. These sequences represent ~50% of the genomes publicly available in GISAID from the state so far, notably reaching a sequencing ratio of ~1:100 cases reported during April and May 2021. We then analyzed lineage dynamics in Rio de Janeiro as well as the emergence and dispersion of new variants.

Three main phases defined the pandemic in the state of Rio de Janeiro, each of them represented by a different lineage. Remarkably, all phases occurred with distinct patterns of transmissibility and mortality. Whereas the daily number of cases increased during the second and third phases, more people died in the first and third ones. We hypothesized that a lack of knowledge about the disease and resources to treat patients in an early pandemic might explain such severity in the first phase [30–32]. After a year of pandemic, a significant amount of information had been already gathered about the virus and the disease, certainly improving patient care. We have observed a switch in the death profiles throughout the pandemic. While deaths were more restricted to old groups and people with comorbidities in the first phase, a substantial increase of young people was reported in the third wave. Prioritizing the vaccination of older adults might have had a strong impact on reducing the mortality of this group [33]. However, it does not explain the rise in mortality in younger ages. Thus, it is plausible to assume that the evolution of new viral phenotypes is correlated to the rejuvenation in mortality observed in the third stage [34–37].

The displacement of SARS-CoV-2 lineage across each phase suggests that the new successive lineage had a higher adaptive value than the last one [4,6]. Indeed, the three lineages progressively accumulated aggressive key mutations associated with higher transmissibility or evasion to the host immune system response: B.1.1.33 carried D614G, P.2 harbors D614G and E484K, and P.1 carries N501Y in addition to the two previously mentioned. All of these mutations occurred in Spike protein. The observed synchronicity between lineage substitution and a rise in the number of cases may reflect this increase in the adaptability of the virus to human biology. This dynamic raises a red flag to the aggressiveness of the next dominant lineage. Though lineage displacement is a well-documented aspect of pathogen evolution [38–40], the mechanisms and models underlying the influence of this event on the resurgence of local outbreaks of COVID-19 are still poorly understood. Nevertheless, displacement of lineages by B.1.1.7, B.1.351, B.1.617.2, and P.1 has been reported repeatedly worldwide [5,7,41–44].

We caution that direct association between each phase and the lineage distribution must be interpreted considering the temporal bias in genome sequencing, especially during the first phase (March–October 2020). Nevertheless, previous studies indicate that B.1.1.33 was widely disseminated throughout the country in the early pandemic phase [8,9], supporting our findings for Rio de Janeiro. Furthermore, the concentration of samples in the metropolitan region challenges the exact dispersion routes within-state, as also highlighted by Wilkinson and colleagues [45]. Although we consider the circulation of lineages an important component that characterized each phase, it does not fully explain the number of cases and deaths once the management of the epidemic also changed over time.

Therefore, the surge in cases could also have been triggered by the gradual flexibilization of social distancing measures [46]. As the number of cases never shrank to pre-pandemic levels in either phase, the slight increase in transmission rates caused by this flexibilization would be enough to create the exponential rise in the virus population. Though the "lineage" and the "mobility" hypotheses could be perceived as opposing nar-

ratives, we reiterate that both endogenous and exogenous factors control the threshold of a new surge in cases. The introduction of more aggressive lineages requires the enforcement of non-pharmacological interventions to control the local outbreak. So far, the only way to circumvent this relationship is to reduce viral transmission rates through vaccination. Further epidemiological modeling is still needed to evaluate better the influence of all mechanisms involved in this dynamic.

An important aspect of the disease dynamic revealed by this study was that viral supercarriers accounted for the overwhelming majority of the circulating virus in the state. Recent studies demonstrated that transmission of COVID-19 is more likely to occur in individuals with higher viral load [47–49]. Given that most of the samples in the period analyzed were assigned to P.1, such a mechanism could also shed light on the rapid spread of this VoC. Considering that the selection of RT-PCR positive samples for sequencing in our analysis is based on a pre-filtering of samples with Ct values < 27, the values of 60% and >90% of viral load of the supercarriers could represent an even more significant amount of the circulating virus in the population. Indeed, Yang et al., recently described a bigger fraction of viral load (90%) being carried by just 2% of the infected individuals [50], which corroborates our findings. Though the observed viral load in our cohort could be explained by each patient's different stages of infection, it is known that this measure is significantly variable between individuals [51,52]. Overall, supercarriers may play an important role in the dynamic of virus dissemination in the state.

It is well known that rises in effective population size increase molecular and phenotypic variation, accelerating diversification rates [53,54]. It is not surprising that the origin of P.1.2 is simultaneous to the countrywide propagation of P.1 in early January 2021 [10,55,56], when this rate is expected to be high. However, the new mutations observed in P.1.2 do not indicate that this lineage has a higher adaptive value than P.1. Indeed, structural protein modeling of its sub-clade suggests a more increased instability and better docking with antibodies than P.1. The fast dispersal of P.1.2 within Brazil could be attributed to an opportunistic event. As P.1 was introduced in Brazilian states at the end of 2020, it induced an immune response in the host population, which temporarily prevents reinfection and limits the number of available hosts. Therefore, the P.1.2 population could only grow before P.1 had not achieved such high frequency, as observed after January [39]. The same behavior could explain the emergence and spread across the state observed by the lineage P.5. A new opportunity for growth opens after infection-induced immune response fades in the host population, which can only be prevented through vaccination.

Finally, learning from past efforts to control the pandemic shapes future decisions towards improving national public health, especially when embracing the findings reported by the scientific community. Currently, as this manuscript is being written, the state is at imminent risk of experiencing the fourth phase of the COVID-19 epidemic due to the introduction of VoC delta (B.1.617.2), as recently reported by us [57]. Therefore, understanding the different epidemiological aspects controlling the SARS-CoV-2 dissemination, such as host mobility patterns and viral evolution, is essential to elaborate effective intervention measures that reduce the number of affected individuals. These evaluations are particularly vital in regions with high-density populations, which are prone to spread viruses worldwide.

genomes deposited in GISAID and number of cases from the state of Rio de Janeiro, Table S4: Gamma generalized-linear model results, Table S5: Mutations present in more than 75% of sequences of each lineage deposited in the GISAID database, summarized by the Lineage Comparison tool available at https://outbreak.info/, Table S6: Energy comparison between P.1 and sub-clade P.1.2 models using FoldX tool, Table S7: Number of contacts in common between P.1 and sub-clade P.1.2 in complex with the antibodies and ACE2. Table S8: GISAID acknowledgments.

Author Contributions: Conceptualization: C.M.B.d.M., C.J.S., A.T. and A.T.R.V.; methodology: R.d.S.F.J., A.P.L., O.B., Y.M., C.J.S and A.T.R.V.; software: R.d.S.F.J., A.P.L., L.G.P.d.A., D.T.M., Y.M. and R.B.G.; validation: L.C., A.L.G., A.P.d.C.G., C.A., D.M. and T.F.C.; resources: I.V.d.S., E.M.d.C., M.S.R., S.C., F.D.d.S., M.H.d.O.G., L.M.d.S., C.G.d.S., C.L.P.R., A.C.C. and C.M.B.d.M.; writing—original draft preparation: R.d.S.F.J., A.P.L. and D.T.M.; writing—review and editing, C.J.S. and A.T.R.V.; visualization, R.d.S.F.J., A.P.L., D.T.M., O.B. and Y.M.; supervision, C.J.S., A.T. and A.T.R.V.; project administration: A.T. and A.T.R.V.; funding acquisition: A.T.R.V. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of the Universidade Federal do Rio de Janeiro (protocol code, CAAE: 30161620.0.1001.5257 and 34025020.0.0000.5257).

Informed Consent Statement: Research protocol was approved without informed consent in accordance with Brazilian National Health Council's Resolution 510/2016. All samples were residual COVID-19 clinical diagnostic samples de-identified before receipt by the researchers.

Data Availability Statement: The data generated in this study is publicly available in Gisaid (www.gisaid.org) the identifiers of each sequence are reported in Table S8.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Rambaut, A.; Holmes, E.C.; O'Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. Addendum: A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology. *Nat. Microbiol.* **2021**, *6*, 415. [CrossRef]
2. Davis, C.; Logan, N.; Tyson, G.; Orton, R.; Harvey, W.; Haughney, J.; Perkins, J.; Peacock, T.P.; Barclay, W.S.; Cherepanov, P.; et al. Reduced Neutralisation of the Delta (B.1.617.2) SARS-CoV-2 Variant of Concern Following Vaccination. *bioRxiv* **2021**. [CrossRef]
3. Goes, L.R.; Siqueira, J.D.; Garrido, M.M.; Alves, B.M.; Pereira, A.C.P.M.; Cicala, C.; Arthos, J.; Viola, J.P.B.; Soares, M.A.; INCA COVID-19 Task Force. New Infections by SARS-CoV-2 Variants of Concern after Natural Infections and Post-Vaccination in Rio de Janeiro, Brazil. *Infect. Genet. Evol.* **2021**, *94*, 104998. [CrossRef]
4. Garcia-Beltran, W.F.; Lam, E.C.; St Denis, K.; Nitido, A.D.; Garcia, Z.H.; Hauser, B.M.; Feldman, J.; Pavlovic, M.N.; Gregory, D.J.; Poznansky, M.C.; et al. Multiple SARS-CoV-2 Variants Escape Neutralization by Vaccine-Induced Humoral Immunity. *Cell* **2021**, *184*, 2372–2383.e9. [CrossRef]
5. Bolze, A.; Cirulli, E.T.; Luo, S.; White, S.; Wyman, D.; Dei Rossi, A.; Cassens, T.; Jacobs, S.; Nguyen, J.; Ramirez, J.M., III; et al. Rapid Displacement of SARS-CoV-2 Variant B.1.1.7 by B.1.617.2 and P.1 in the United States. *bioRxiv* **2021**. [CrossRef]
6. Chen, Z.; Chong, K.C.; Wong, M.C.S.; Boon, S.S.; Huang, J.; Wang, M.H.; Ng, R.W.Y.; Lai, C.K.C.; Chan, P.K.S. A Global Analysis of Replacement of Genetic Variants of SARS-CoV-2 in Association with Containment Capacity and Changes in Disease Severity. *Clin. Microbiol. Infect.* **2021**, *27*, 750–757. [CrossRef] [PubMed]

7. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Detection of a SARS-CoV-2 Variant of Concern in South Africa. *Nature* **2021**, *592*, 438–443. [CrossRef] [PubMed]

8. Resende, P.C.; Delatorre, E.; Gräf, T.; Mir, D.; Motta, F.C.; Appolinario, L.R.; da Paixão, A.C.D.; da Fonseca Mendonça, A.C.; Ogrzewalska, M.; Caetano, B.; et al. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. *Front. Microbiol.* **2020**, *11*, 615280. [CrossRef]

9. Candido, D.S.; Claro, I.M.; de Jesus, J.G.; Souza, W.M.; Moreira, F.R.R.; Dellicour, S.; Mellan, T.A.; du Plessis, L.; Pereira, R.H.M.; Sales, F.C.S.; et al. Evolution and Epidemic Spread of SARS-CoV-2 in Brazil. *Science* **2020**, *369*, 1255–1260. [CrossRef] [PubMed]

10. Lamarca, A.P.; de Almeida, L.G.P.; da Silva Francisco, R.; Lima, L.F.A.; Scortecci, K.C.; Perez, V.P.; Brustolini, O.J.; Sousa, E.S.S.; Secco, D.A.; Santos, A.M.G.; et al. Genomic Surveillance of SARS-CoV-2 Tracks Early Interstate Transmission of P.1 Lineage and Diversification within P.2 Clade in Brazil. *medRxiv* **2021**. [CrossRef]

11. Moreira, F.R.R.; D'arc, M.; Mariani, D.; Herlinger, A.L.; Schiffler, F.B.; Rossi, Á.D.; de Carvalho Leitão, I.; dos Santos Miranda, T.; Cosentino, M.A.C.; de Paula Tôrres, M.C.; et al. Epidemiological Dynamics of SARS-CoV-2 VOC Gamma in Rio de Janeiro, Brazil. *medRxiv* **2021**. [CrossRef]

12. Voloch, C.M.; da Silva Francisco, R.; de Almeida, L.G.P.; Cardoso, C.C.; Brustolini, O.J.; Gerber, A.L.; Guimarães, A.P.D.C.; Mariani, D.; da Costa, R.M.; Ferreira, O.C.; et al. Genomic Characterization of a Novel SARS-CoV-2 Lineage from Rio de Janeiro, Brazil. *J. Virol.* **2021**, *95*, 10. [CrossRef]

13. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *Am. J. Epidemiol.* **2013**, *178*, 1505–1512. [CrossRef]

14. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 20 July 2021).

15. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]

16. Katoh, K.; Frith, M.C. Adding Unaligned Sequences into an Existing Alignment Using MAFFT and LAST. *Bioinformatics* **2012**, *28*, 3144–3146. [CrossRef]

17. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef] [PubMed]

18. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]

19. Sagulenko, P.; Puller, V.; Neher, R.A. TreeTime: Maximum-Likelihood Phylodynamic Analysis. *Virus Evol.* **2018**, *4*, vex042. [CrossRef] [PubMed]

20. Drummond, A.J.; Rambaut, A. BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evol. Biol.* **2007**, *7*, 214. [CrossRef]

21. Karcher, M.D.; Palacios, J.A.; Lan, S.; Minin, V.N. Phylodyn: An R Package for Phylodynamic Simulation and Inference. *Mol. Ecol. Resour.* **2017**, *17*, 96–100. [CrossRef]

22. Dellicour, S.; Rose, R.; Faria, N.R.; Lemey, P.; Pybus, O.G. SERAPHIM: Studying Environmental Rasters and Phylogenetically Informed Movements. *Bioinformatics* **2016**, *32*, 3204–3206. [CrossRef]

23. Buß, O.; Rudat, J.; Ochsenreither, K. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Comput. Struct. Biotechnol. J.* **2018**, *16*, 25–33. [CrossRef] [PubMed]

24. Sarakatsannis, J.N.; Duan, Y. Statistical Characterization of Salt Bridges in Proteins. *Proteins* **2005**, *60*, 732–739. [CrossRef] [PubMed]

25. Yan, Y.; Tao, H.; He, J.; Huang, S.-Y. The HDOCK Server for Integrated Protein–protein Docking. *Nat. Protoc.* **2020**, *15*, 1829–1852. [CrossRef] [PubMed]

26. Raybould, M.I.J.; Kovaltsuk, A.; Marks, C.; Deane, C.M. CoV-AbDab: The Coronavirus Antibody Database. *Bioinformatics* **2021**, *37*, 734–735. [CrossRef]

27. Xue, L.C.; Rodrigues, J.P.; Kastritis, P.L.; Bonvin, A.M.; Vangone, A. PRODIGY: A Web Server for Predicting the Binding Affinity of Protein-Protein Complexes. *Bioinformatics* **2016**, *32*, 3676–3678. [CrossRef]

28. Darnell, S.J.; LeGault, L.; Mitchell, J.C. KFC Server: Interactive Forecasting of Protein Interaction Hot Spots. *Nucleic Acids Res.* **2008**, *36*, W265–W269. [CrossRef]

29. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667. [CrossRef]

30. Ciceri, F.; Ruggeri, A.; Lembo, R.; Puglisi, R.; Landoni, G.; Zangrillo, A. COVID-BioB Study Group Decreased in-Hospital Mortality in Patients with COVID-19 Pneumonia. *Pathog. Glob. Health* **2020**, *114*, 281–282. [CrossRef]

31. Horwitz, L.I.; Jones, S.A.; Cerfolio, R.J.; Francois, F.; Greco, J.; Rudy, B.; Petrilli, C.A. Trends in Covid-19 Risk-Adjusted Mortality Rates in a Single Health System. *bioRxiv* **2020**. [CrossRef]

32. Sun, Q.; Qiu, H.; Huang, M.; Yang, Y. Lower Mortality of COVID-19 by Early Recognition and Intervention: Experience from Jiangsu Province. *Ann. Intensive Care* **2020**, *10*, 33. [CrossRef]

33. Victora, C.; Castro, M.C.; Gurzenda, S.; de Medeiros, A.C.; França, G.; Barros, A.J.D. Estimating the Early Impact of Vaccination against COVID-19 on Deaths among Elderly People in Brazil: Analyses of Routinely-Collected Data on Vaccine Coverage and Mortality. *bioRxiv* **2021**. [CrossRef]

34. de Souza, F.S.H.; Hojo-Souza, N.S.; da Silva, C.M.; Guidoni, D.L. Second Wave of COVID-19 in Brazil: Younger at Higher Risk. *Eur. J. Epidemiol.* **2021**, *36*, 441–443. [CrossRef] [PubMed]

35. Bono, L.M.; Gensel, C.L.; Pfennig, D.W.; Burch, C.L. Competition and the Origins of Novelty: Experimental Evolution of Niche-Width Expansion in a Virus. *Biol. Lett.* **2013**, *9*, 20120616. [CrossRef]

36. Freitas, A.R.R.; Beckedorff, O.A.; Cavalcanti, L.P.D.G.; Siqueira, A.M.; Castro, D.B.; Costa, C.F.D.; Lemos, D.R.Q.; Barros, E.N.C. The Emergence of Novel SARS-CoV-2 Variant P.1 in Amazonas (Brazil) Was Temporally Associated with a Change in the Age and Gender Profile of COVID-19 Mortality. *SSRN Electron. J.* **2021**. [CrossRef]

37. De Oliveira, M.H.S.; Lippi, G.; Henry, B.M. Sudden Rise in COVID-19 Case Fatality among Young and Middle-Aged Adults in the South of Brazil after Identification of the Novel B.1.1.28.1 (P.1) SARS-CoV-2 Strain: Analysis of Data from the State of Parana. *bioRxiv* **2021**. [CrossRef]

38. Hanley, K.A.; Nelson, J.T.; Schirtzinger, E.E.; Whitehead, S.S.; Hanson, C.T. Superior Infectivity for Mosquito Vectors Contributes to Competitive Displacement among Strains of Dengue Virus. *BMC Ecol.* **2008**, *8*, 1. [CrossRef]

39. Yuen, J.E. Modelling Pathogen Competition and Displacement– Phytophthora Infestans in Scandinavia. *Eur. J. Plant Pathol.* **2012**, *133*, 25–32. [CrossRef]

40. Wolf, Y.I.; Viboud, C.; Holmes, E.C.; Koonin, E.V.; Lipman, D.J. Long Intervals of Stasis Punctuated by Bursts of Positive Selection in the Seasonal Evolution of Influenza A Virus. *Biol. Direct* **2006**, *1*, 34. [CrossRef]

41. Campbell, F.; Archer, B.; Laurenson-Schafer, H.; Jinnai, Y.; Konings, F.; Batra, N.; Pavlin, B.; Vandemaele, K.; Van Kerkhove, M.D.; Jombart, T.; et al. Increased Transmissibility and Global Spread of SARS-CoV-2 Variants of Concern as at June 2021. *Eurosurveillance* **2021**, *26*, 2100509. [CrossRef]

42. Di Giallonardo, F.; Puglia, I.; Curini, V.; Cammà, C.; Mangone, I.; Calistri, P.; Cobbin, J.C.A.; Holmes, E.C.; Lorusso, A. Emergence and Spread of SARS-CoV-2 Lineages B.1.1.7 and P.1 in Italy. *Viruses* **2021**, *13*, 794. [CrossRef]

43. Sallam, M.; Mahafzah, A. Molecular Analysis of SARS-CoV-2 Genetic Lineages in Jordan: Tracking the Introduction and Spread of COVID-19 UK Variant of Concern at a Country Level. *Pathogens* **2021**, *10*, 302. [CrossRef]

44. Challen, R.; Dyson, L.; Overton, C.E.; Guzman-Rincon, L.M.; Hill, E.M.; Stage, H.B.; Brooks-Pollock, E.; Pellis, L.; Scarabel, F.; Pascall, D.J.; et al. Early Epidemiological Signatures of Novel SARS-CoV-2 Variants: Establishment of B.1.617.2 in England. *bioRxiv* **2021**. [CrossRef]

45. Wilkinson, E.; Giovanetti, M.; Tegally, H.; San, J.E.; Lessels, R.; Cuadros, D.; Martin, D.P.; Zekri, A.-R.N.; Sangare, A.K.; Ouedraogo, A.-S.; et al. A Year of Genomic Surveillance Reveals How the SARS-CoV-2 Pandemic Unfolded in Africa. *medRxiv* **2021**. [CrossRef]

46. Morato, M.M.; Bastos, S.B.; Cajueiro, D.O.; Normey-Rico, J.E. An Optimal Predictive Control Strategy for COVID-19 (SARS-CoV-2) Social Distancing Policies in Brazil. *Annu. Rev. Control* **2020**, *50*, 417–431. [CrossRef] [PubMed]

47. He, X.; Lau, E.H.Y.; Wu, P.; Deng, X.; Wang, J.; Hao, X.; Lau, Y.C.; Wong, J.Y.; Guan, Y.; Tan, X.; et al. Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19. *Nat. Med.* **2020**, *26*, 672–675. [CrossRef] [PubMed]

48. Marks, M.; Millat-Martinez, P.; Ouchi, D.; Roberts, C.H.; Alemany, A.; Corbacho-Monné, M.; Ubals, M.; Tobias, A.; Tebé, C.; Ballana, E.; et al. Transmission of COVID-19 in 282 Clusters in Catalonia, Spain: A Cohort Study. *Lancet Infect. Dis.* **2021**, *21*, 629–636. [CrossRef]

49. Bjorkman, K.K.; Saldi, T.K.; Lasda, E.; Bauer, L.C.; Kovarik, J.; Gonzales, P.K.; Fink, M.R.; Tat, K.L.; Hager, C.R.; Davis, J.C.; et al. Higher Viral Load Drives Infrequent SARS-CoV-2 Transmission between Asymptomatic Residence Hall Roommates. *medRxiv* **2021**. [CrossRef]

50. Yang, Q.; Saldi, T.K.; Gonzales, P.K.; Lasda, E.; Decker, C.J.; Tat, K.L.; Fink, M.R.; Hager, C.R.; Davis, J.C.; Ozeroff, C.D.; et al. Just 2% of SARS-CoV-2−positive Individuals Carry 90% of the Virus Circulating in Communities. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104547118. [CrossRef]

51. Kissler, S.M.; Fauver, J.R.; Mack, C.; Olesen, S.W.; Tai, C.; Shiue, K.Y.; Kalinich, C.C.; Jednak, S.; Ott, I.M.; Vogels, C.B.F.; et al. Viral Dynamics of Acute SARS-CoV-2 Infection. *medRxiv* **2021**. [CrossRef]

52. Zou, L.; Ruan, F.; Huang, M.; Liang, L.; Huang, H.; Hong, Z.; Yu, J.; Kang, M.; Song, Y.; Xia, J.; et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N. Engl. J. Med.* **2020**, *382*, 1177–1179. [CrossRef]

53. Frankham, R. Relationship of Genetic Variation to Population Size in Wildlife. *Conserv. Biol.* **1996**, *10*, 1500–1508. [CrossRef]

54. Stevens, M.H.H.; Sanchez, M.; Lee, J.; Finkel, S.E. Diversification Rates Increase with Population Size and Resource Concentration in an Unstructured Habitat. *Genetics* **2007**, *177*, 2243–2250. [CrossRef] [PubMed]

55. Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.S.; Mishra, S.; Crispim, M.A.E.; Sales, F.C.S.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [CrossRef] [PubMed]

56. Naveca, F.G.; Nascimento, V.; de Souza, V.C.; Corado, A.L.; Nascimento, F.; Silva, G.; Costa, Á.; Duarte, D.; Pessoa, K.; Mejía, M.; et al. COVID-19 in Amazonas, Brazil, Was Driven by the Persistence of Endemic Lineages and P.1 Emergence. *Nat. Med.* **2021**. [CrossRef] [PubMed]

57. de Almeida, L.G.P.; Lamarca, A.L.; Fracisco Junior, R.S.; Cavalcante, L.; Gerber, A.L.; Guimarães, A.P.C.; Machado, D.T.; Alves, C.; Mariani, D.; Cruz, T.F.; et al. Genomic Surveillance of SARS-CoV-2 in the State of Rio de Janeiro, Brazil: Technical Briefing. Available online: https://virological.org/t/genomic-surveillance-of-sars-cov-2-in-the-state-of-rio-de-janeiro-brazil-technical-briefing/683 (accessed on 14 July 2021).