PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil

Alessandra P. Lamarca[1], Luiz G. P. de Almeida[1], Ronaldo da Silva Francisco, Jr.[1], Lucymara Fassarella Agnez Lima[2], Kátia Castanho Scortecci[2], Vinícius Pietta Perez[3], Otavio J. Brustolini[1], Eduardo Sérgio Soares Sousa[4], Danielle Angst Secco[5], Angela Maria Guimarães Santos[5], George Rego Albuquerque[6], Ana Paula Melo Mariano[6], Bianca Mendes Maciel[6], Alexandra L. Gerber[1], Ana Paula de C. Guimarães[1], Paulo Ricardo Nascimento[7], Francisco Paulo Freire Neto[7], Sandra Rocha Gadelha[6], Luís Cristóvão Porto[5], Eloiza Helena Campana[4], Selma Maria Bezerra Jeronimo[7,8], Ana Tereza R. Vasconcelos[1]*

1 Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Petrópolis, Brazil, **2** Laboratório de Biologia Molecular e Genômica, Universidade Federal do Rio Grande do Norte, Natal, Brazil, **3** Laboratório de Endemias, Núcleo de Medicina Tropical, Centro de Ciências da Saúde, Universidade Federal da Paraíba, João Pessoa, Brazil, **4** Laboratório de Biologia Molecular, Centro de Ciências Médicas, Universidade Federal da Paraíba, João Pessoa, Brazil, **5** Laboratório de Histocompatibilidade e Criopreservação, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil, **6** Laboratório de Farmacogenômica e Epidemiologia Molecular, Universidade Estadual de Santa Cruz, Ilhéus, Brazil, **7** Instituto de Medicina Tropical do Rio Grande do Norte, Universidade Federal do Rio Grande do Norte, Natal, Brazil, **8** Departamento de Bioquímica, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal, Brazil

* atrv@lncc.br

Citation: Lamarca AP, de Almeida LGP, Francisco RdS, Jr., Lima LFA, Scortecci KC, Perez VP, et al. (2021) Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil. PLoS Negl Trop Dis 15(10): e0009835. https://doi.org/10.1371/journal.pntd.0009835

Editor: Colleen B. Jonsson, University of Tennessee Health Science Center College of Medicine Memphis, UNITED STATES

Received: March 11, 2021

Accepted: September 23, 2021

Published: October 13, 2021

Data Availability Statement: All assembled genomes are available in the GISAID database after registration. Accession IDs are indicated in S1 Table. RAW files were deposited in the NCBI database under the BioProject ID PRJNA752057.

Funding: This work was developed in the frameworks of Corona-ômica-RJ (FAPERJ = E-26/210.179/2020 http://www.faperj.br/) and Rede Corona-ômica BR MCTI/FINEP (FINEP =

Abstract

The sharp increase of COVID-19 cases in late 2020 has made Brazil the new epicenter of the ongoing SARS-CoV-2 pandemic. The novel viral lineages P.1 (Variant of Concern Gamma) and P.2, respectively identified in the Brazilian states of Amazonas and Rio de Janeiro, have been associated with potentially higher transmission rates and antibody neutralization escape. In this study, we performed the whole-genome sequencing of 185 samples isolated from three out of the five Brazilian regions, including Amazonas (North region), Rio Grande do Norte, Paraíba and Bahia (Northeast region), and Rio de Janeiro (Southeast region) in order to monitor the spread of SARS-CoV-2 lineages in Brazil in the first months of 2021. Here, we showed a widespread dispersal of P.1 and P.2 across Brazilian regions and, except for Amazonas, P.2 was the predominant lineage identified in the sampled states. We estimated the origin of P.2 lineage to have happened in February, 2020 and identified that it has differentiated into new clades. Interstate transmission of P.2 was detected since March, but reached its peak in December, 2020 and January, 2021. Transmission of P.1 was also high in December and its origin was inferred to have happened in August 2020. We also confirmed the presence of lineage P.7, recently described in the southernmost region of Brazil, to have spread across the Northeastern states. P.1, P.2 and P.7 are descended from the ancient B.1.1.28 strain, which co-dominated the first phase of the pandemic in Brazil with the B.1.1.33 strain. We also identified the occurrence of a new lineage descending from

01.20.0029.000462/20 http://www.finep.gov.br/; CNPq = 404096/2020-4 https://www.gov.br/cnpq/pt-br). A.T.R.V is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (303170/2017-4) and FAPERJ (E-26/202.903/20). R.S.F.J is a recipient of a graduate fellowship from CNPq. A.P.L is granted a post-doctoral scholarship (DTI-A) from CNPq. SMBJ was supported by Ministério da Educação https://www.gov.br/mec/pt-br (UFRN Covid Task Force), Conselho Nacional de Desenvolvimento Científico e Tecnológico (440893/2016-0) and by JBS https://jbs.com.br/. E.S.S.S was supported by Fundação de Amparo à Pesquisa do Estado da Paraíba – FAPESQ http://fapesq.rpp.br/ (003/2020) and Laboratórios de Campanha MCTI/FINEP (0494/20 - 01.20.0026.00). L.F.A.L was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior https://www.gov.br/capes/pt-br. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

B.1.1.33 that convergently carries the E484K mutation, N.9. Indeed, the recurrent report of many novel SARS-CoV-2 genetic variants in Brazil could be due to the absence of effective control measures resulting in high SARS-CoV2 transmission rates. Altogether, our findings provided a landscape of the critical state of SARS-CoV-2 across Brazil and confirm the need to sustain continuous sequencing of the SARS-CoV-2 isolates worldwide in order to identify novel variants of interest and monitor for vaccine effectiveness.

Author summary

Since its first detection in December 2019, the SARS-CoV-2 has evolved into more than a thousand recognized lineages. Several of these lineages are known to have higher transmissibility or better escape from the immune system. One of them is the P.1 lineage, also known as the variant of concern Gamma. It was first discovered in Brazil in December 2020 and has quickly replaced the previous lineage dominating the country, P.2. We used genomic data from SARS-CoV-2 samples that were collected in the first months of 2021 to analyze how P.1, P.2 and other lineages had spread across Brazil. Our research has identified that P.1 lineage was already present in several states of Brazil almost two months before its first detection in the state of Amazonas. Our work sheds light on the importance of continuous monitoring of SARS-CoV-2 lineages in historically understudied regions to early detect and control the spread of new variants of concern.

Introduction

More than a year after the first case of SARS-CoV-2 infection in Brazil, the country is in a catastrophic situation with 19 million cases of COVID-19 and 550,000 deaths (https://coronavirus.jhu.edu/map.html). The initially dominant lineages B.1.1.28 and B.1.1.33 [1] have been replaced first by the variant P.2 and later by the new variant of concern P.1 (Gamma) [2–4]. P.2 was firstly reported in November 2020 in samples from the state of Rio de Janeiro and was estimated to have first diverged in late July [5]. By December 2020, it was already prevalent in samples from the Brazilian states of Rio Grande do Sul Amazonas and Rio de Janeiro [3,6,7]. P.1 was first detected in the state of Amazonas in mid-December 2020, with a proposed emergence around November [3,4]. Both lineages evolved within the B.1.1.28 clade and convergently carried the E484K mutation in the receptor-binding domain (RBD) of the Spike protein. In addition to E484K, P.1 harbors the N501Y and K417T mutations in the RBD region. It is suggested that those three mutations allow SARS-CoV-2 to better escape from the host's immune response [8–10]. This hypothesis is supported by the explosive spread of P.1 cases across Brazil and reports of reinfection involving both P.1 and P.2 lineages [11,12].

During the first phase of the COVID-19 pandemic in Brazil, long-distance travel between large urban cities in southeastern states and less populated states from North and Northeast regions played an important role in the explosion of cases across the country [1,13,14]. Since mid-November, there has been a new surge in COVID-19 cases in Brazil, prompting the delimitation of a second phase of the pandemic in the country. This sharp increase in cases is attributed to the emergence of P.1 lineage, which has been reported in several cities in Brazil since December [15–20]. Unfortunately, lineage pervasiveness and genomic diversity are still unknown or outdated in several Brazilian states. If the aforementioned mutations in P.1 and

P.2 indeed promote escape from the host's immune response, this information is crucial to elaborate measures to slow nationwide and worldwide spread.

Monitoring P.1 lineage in Brazil is mainly executed by positive-PCR screening with mutation-targeted primers [18,21–23]. Although this strategy is valuable to estimate the relative frequency of the chosen variant in the screened population, it fails if the primer's target mutates. Furthermore, the exclusive use of targeted screening prevents monitoring the dispersal and prevalence of other lineages. In this context, sequencing of SARS-CoV-2 genomes systematically sampled from the population is decisive in identifying new variants. In order to reduce the knowledge gap regarding lineage distribution in Brazil during the second surge in COVID-19 cases, we performed an epidemiological and genomic survey by sequencing 185 new SARS-CoV-2 genomes from three Brazilian regions, including states of Amazonas (North region), Rio Grande do Norte, Paraíba, Bahia (all three in the Northeast region) and Rio de Janeiro (Southeast region). Samples were collected between December 2020 and February 2021.

Materials and methods

Ethics statement

The present study was approved by Ethical Review Board/Brazilian Commission of Ethical Study (Research Ethics Committee of: Universidade Federal Rio Grande do Norte—CAAE 36287120.2.0000.5537, CAAE 32049320.3.0000.5537, Universidade Federal da Paraíba—CAAE 30658920.4.3004.5183, Universidade Estadual do Rio de Janeiro—CAAE 30135 320.0.0000.5259 and Universidade Estadual de Santa Cruz—CAAE 39142720.5.0000.5526). 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.

Sample collection

In this work, a total of 185 participants were selected from Amazonas (4), Rio Grande do Norte (44), Paraiba (43), Bahia (58), and Rio de Janeiro (36) states, representing the Brazilian North, Northeast, and Southeast regions. Samples from Amazonas) were obtained from the four patients transferred to Paraíba in late January 2021 while samples from Rio de Janeiro, Rio Grande do Norte, Paraíba and Bahia were randomly selected among COVID-19 positive cases. These samples were collected from December 1st, 2020 through February 15th, 2021. Participants were divided into ninety-two males and 93 females, with age ranging between 11–90 years and with CT values between 8.70 and 29.00 (S1 Table). Nasopharyngeal swabs were obtained from each participant and SARS-CoV-2 infection was diagnosed by RT-PCR using CDC/EUA protocol [24], OneStep/COVID-19 (IBMP, Brazil) Allplex 2019-nCoV (Seegene, South Korea) or nCoVqRT-PCR kits (Biomanguinhos, Fiocruz, Rio de Janeiro).

Next-generation sequencing and bioinformatics analysis

cDNA synthesis and viral whole-genome amplification were carried out following the Artic Network protocol (https://artic.network/ncov-2019). Amplicon libraries were prepared using the Nextera DNA Flex kit (Illumina, USA). Sequencing was performed in a MiSeq System using MiSeq Reagent Kit v3 (Illumina, USA). Bioinformatic analysis was performed using an in-house pipeline for NGS data pre-processing, variant calling, and genome assembly as previously described [5,6,25]. Briefly, we first inspected the quality control of NGS raw read files in FASTQ format using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)

and removed low-quality, bad-formed and optical duplicates in 5' primer regions sequences with Trimmomatic v0.39 (parameters AVGQUAL = 25 and MINLEN = 100) [26], cutadapt v2.1 and clumpify v38.41 (https://sourceforge.net/projects/bbmap/), respectively. After that, the remaining reads were mapped to the Wuhan-Hu-1 (NC_045512.2) reference genome using BWA v0.7.17 [27]. The BAM files generated in the previous step were sorted and indexed using samtools v.1.11 [28]. We also used GATK v4.1.7.0 to perform the variant calling and filtration [29] and snpEff/SnpSiff v5.0e for VCF annotation. We then combined the list of variants identified in each sample to generate the consensus sequences with bcftools v.1.9 and bedtools v2.29.2 [30–32]. The raw sequencing files and the assembled genomes were submitted to the NCBI and GISAID public databases (NCBI BioProject ID: PRJNA752057, S1 Table)

Phylogenetic analyses

The evolutionary position of the newly sequenced genomes was inferred using 1441 sequences from Brazil and 70 from other countries, all of them obtained from the GISAID database on February 25th, 2021. The Brazilian background sequences were selected following the strategy described by Paiva et al. [33]. We modified this protocol by clustering aligned sequences with similarity of 0.99985 with CD-hit [34], keeping only the oldest record of each cluster and removing restrictions by country. Global sequences were added by selecting the sequence with the oldest sampling date in GISAID for each lineage found in the Brazilian background dataset. Genome sequence from Wuhan-Hu-1 (NC_045512.2) sample was then added as an outgroup. All sequence alignment steps were conducted using MAFFT with—auto and—addfragments parameters [35]. We used IQ-TREE2 [16] to infer the phylogeny of the final alignment. The substitution model GTR+F+I was selected with ModelFinder [36] using the global sequences as a proxy for the genomic diversity within the larger alignment. Clade support was estimated using 1,000 replicates of bootstrap. To confirm the monophyly of P.7 and N.9 clades, we have also reconstructed their phylogenies with an expanded dataset to include all available sequences in GISAID that share their characteristic mutations. The substitution models for these reconstructions were the GTR+F+I for the P.7 clade and the TIM2+F+I for N.9, both also selected using ModelFinder in IQ-TREE2.

We extracted P.1 and P.2 clades from the complete maximum likelihood phylogeny to infer divergence dates and ancestor spatial dispersion of both lineages with BEAST v1.10.4 [37]. After evaluating with TempEst [38] the correlation between root-to-tip distances and sampling dates (S1 Fig), we selected the strict clock model to date P.1 divergence and the lognormal uncorrelated clock for P.2 [39]. Models used in both analyses were Cauchy's relaxed random walk for geographic coordinates [40,41], the GTR+F+I substitution model and the exponential growth coalescent tree prior. All models were employed with default parameters. The MCMC was run through 10,000,000 steps with sampling every 10,000th and a 10% burn-in of the posterior results. We extracted ancestor location coordinates using the SERAPHIM package [42] in R software. Vector and raster map data used to plot dispersal routes were obtained from Natural Earth using the R package *rnaturalearth* and can be found on https://naturalearth.s3.amazonaws.com/10m_cultural/ne_10m_admin_1_states_provinces.zip.

To account for the impact of sampling bias across different Brazilian states on the inference of dispersal routes, we have created ten replicates of the previous analyses by resampling the available P.2 (new $n = 150$) and P.1 (50) genomes weighted by the ratio between the number of cases in each state and number of available genomes in GISAID from each state. Because we only used ingroup sequences, the strict clock model was employed on both datasets to infer their divergence dates. Except for this difference, Bayesian analyses were run with the same model and parameters previously described.

Results

The 185 newly sequenced genomes were assigned to 11 different lineages (Figs 1 and S2), with the majority belonging to P.1 (15.68%) and P.2 lineages (64.32%). Other lineages found were B.1.1.143 (4.32%), B.1.1.33 (3.24%), B.1.1.28 (2.70%), B.1.1.29 (2.70%), P.7 (2.16%), B.1 (1.62%), N.9 (1.08%), B.1.1.306 (0.54%), B.1.1.314 (0.54%), B.1.1.34 (0.54%) and B.1.212 (0.54%). The within-state relative lineages frequency revealed that P.2 was the most abundant lineage in Northeast and Southeast regions (S3 Fig). Among the Northeast states, Rio Grande do Norte showed the highest occurrence of the P1 lineage (34.1% of the sequences obtained). Whereas, in the neighboring state of Paraíba, P.2 was the most frequent lineage (51.2% in this study) since late November 2020 [9], and P1 was only reported in early January 2021 (9.3%). Three lineages are described for the first time in the state (B.1.1.29, B.1.1.34 and B.1.212).

We identified 794 single-nucleotide variants (SNVs) across the 185 genomes sequenced, of which 49% were missense substitutions, 45% synonymous and 6% in non-coding regions of the genome (S4 Fig). We found three nonsense mutations in ORF8 (n = 2) and ORF7a (n = 1)

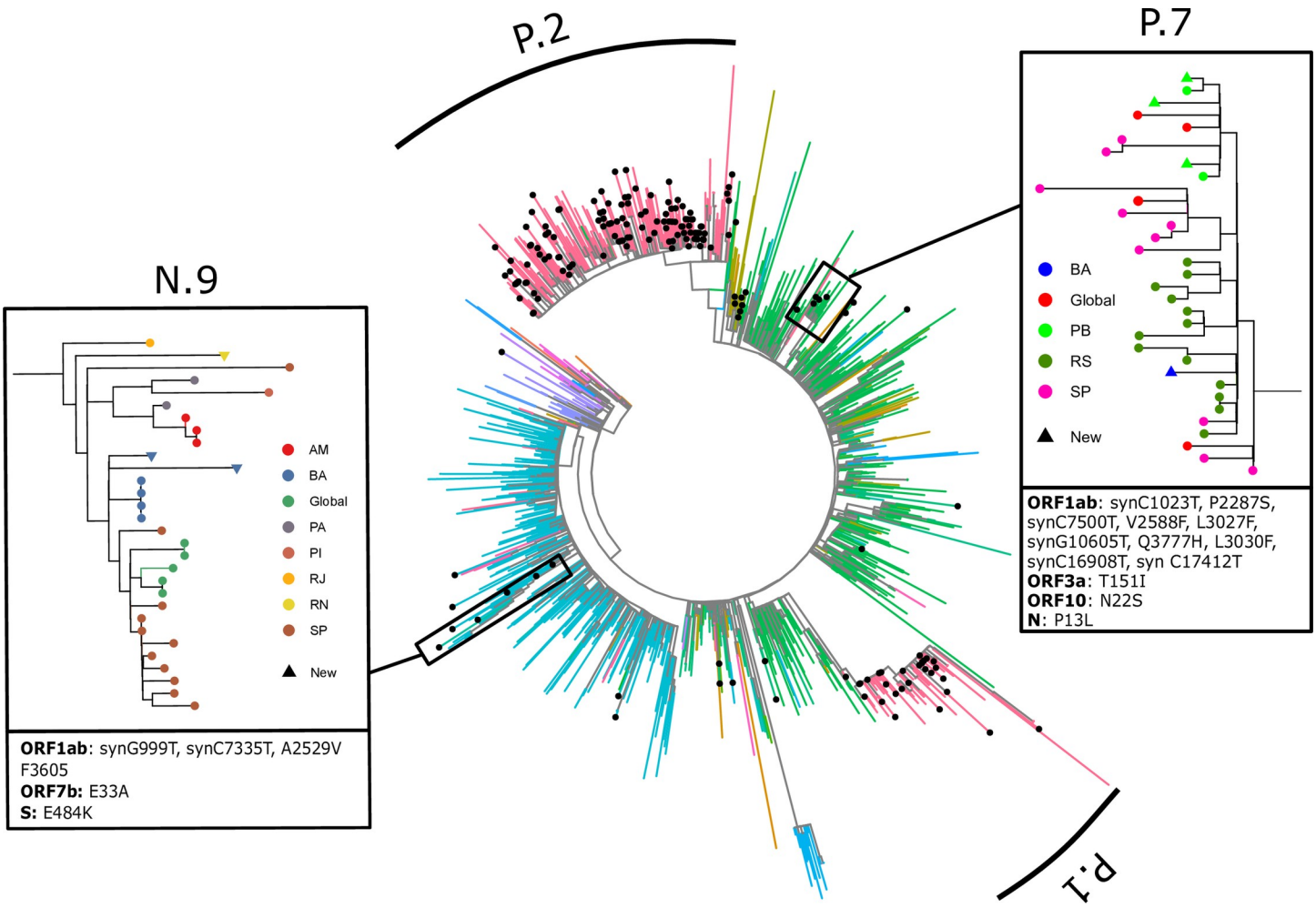

Fig 1. Phylogeny of 1696 SARS-CoV-2 genomes. Newly sequenced samples are signalized by a black point at the tip and lineages P.1 and P.2 are indicated by curved bars. In detail, the N.9 clade, originated from B.1.1.33 (blue branches in the larger tree), and the P.7 clade, originated from B.1.1.28 (green). Colors in smaller trees indicate the Brazilian state in which the sample were collected (AM: Amazonas, BA: Bahia, PA: Pará, PB: Paraíba, PI: Piauí, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RS: Rio Grande do Sul, SP: São Paulo). The boxes below the trees contain the characteristic mutations of each lineage.

https://doi.org/10.1371/journal.pntd.0009835.g001

in genomes from Rio de Janeiro, Rio Grande do Norte and Paraíba. We observed an elevated accumulation of mutation in the 3'UTR of the genome, mainly targeting ORF3 (subunits a, c and d), ORF9 (b and c), ORF8 and ORF7a (S2 Table). The nucleocapsid (N) protein and the subunit S1 of Spike protein showed the highest accumulation among the structural proteins of the SARS-CoV-2 genome. We found 16 SNVs targeting the receptor-binding domain (RBD) in S1, of which eight were missense variants, including K417T, N439K, L452R, S477R, E484K, N501Y, L518I, A522V.

A newly sequenced sample from the state of Rio de Janeiro was recovered as the first divergence within P.1 lineage (Fig 2A). Remarkably, this genome shows traces of intermediary evolution between B.1.1.28 and P.1, harboring 13 out of the 15 lineage-defining mutations according to Pango (https://cov-lineages.org/global_report_P.1.html). We did not observe two mutations (T20N, E92K) characteristic of P.1 clade. The evolutionary position of this new sample was confirmed by repeating the phylogenetic inference analysis with higher P.1 sampling and recovered the same results described here. This newly observed divergence pulls the estimated origin of P.1 lineage to mid-August 2020. Accordingly, interstate dispersal begins in September, with P.1 leaving the state of Amazonas to northeastern states of Rio Grande do Norte and Paraíba (Fig 2B). The divergence between previously sequenced P.1 would happen in mid-October, giving rise to the most common variant. The lineage was already widely distributed across the country by November, with transmission originating in several states, including a reintroduction from Rio Grande do Norte to Amazonas. Interstate transmission reaches its peak in December, with new dispersal routes and maintenance of previous ones. Resampling of the analyzed sequences recovers similar routes to the one obtained with the entire dataset (S5 Fig).

While P.1 sequences are very similar due to their recent origin, we observed a clear evolutive differentiation within P.2 lineage. The first within-clade diversification was estimated to have occurred in late February and the lineage went unreported until December [2], resulting in an uncontrolled transmission across the country (Fig 2B). The first introduction occurs simultaneously in São Paulo and Rio de Janeiro, followed by transmission to Brazil's southernmost state of Rio Grande do Sul. From then onwards, a multitude of dispersal routes is observed between states. Mirroring P.1 behavior, interstate transmission of P.2 was also most intense during December and extended into January 2021. Once again, the resampling of sequences resulted in similar routes to the ones inferred with the complete dataset (S5 Fig).

We have identified a monophyletic clade of 15 sequences containing the characteristic mutations of lineage N.9, including the E484K mutation. To confirm its monophyly, we have reconstructed this clade's phylogeny while further increasing the sampling of N.9 sequences to contain all genomes with the E484K available at GISAID (Fig 1). All these extra samples fall within the described clade. Therefore, the monophyly of the group was not disrupted by either the extensive B.1.1.33 outgroup sampling employed on the larger tree (64% bootstrap support) or by increasing the supposed ingroup (86%). Also noteworthy, one of the newly-sequenced samples from Bahia of to this clade is the single B.1.1.306 reported in this work and additionally harbors a previously undescribed N501Y mutation in this lineage. We hypothesized that this new combination of mutations within B.1.1.33 might be due to the Pango misclassification of this sample. We have also confirmed the monophyletic status (99% bootstrap support) of the proposed lineage P.7 [6], which emerged from B.1.1.28 in Brazil's southernmost state and is now also spread in the Northeast region. This result was, again, recovered even after increasing ingroup sampling with additional sequences available in GISAID. Finally, we report the occurrence of a single sample from the state of Rio Grande do Norte classified as B.1.1.29 that contains both E484K and N429K, uncharacteristic mutations in the lineage.

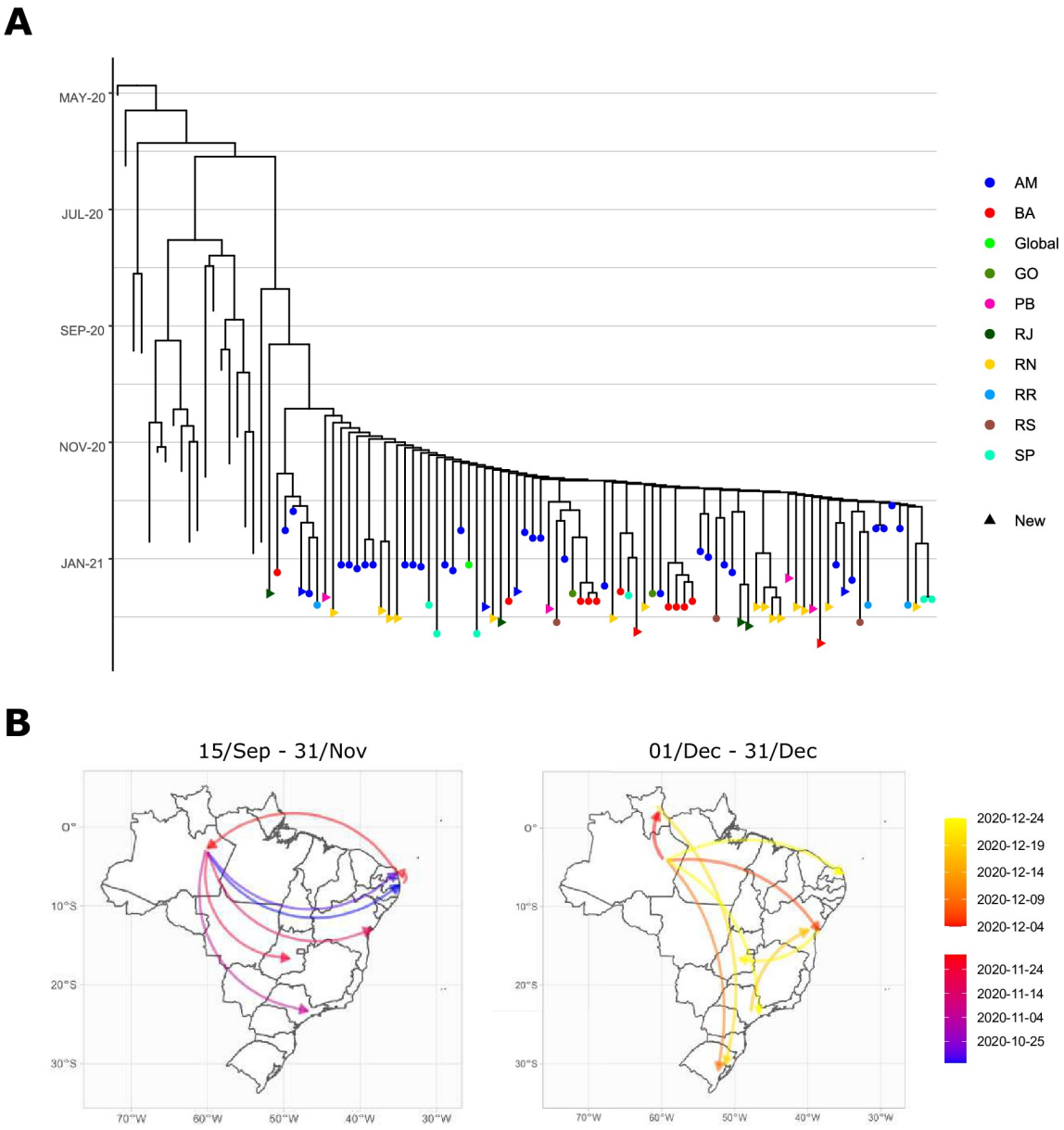

Fig 2. Divergence times within P.1 lineage (A) and its dispersal routes (B). Colors of tip points in the tree indicate the origin of P.1 samples (AM: Amazonas, BA: Bahia, GO: Goiás, PB: Paraíba, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RR: Roraima, RS: Rio Grande do Sul, SP: São Paulo), while tips without a point are sequences from other lineages. Colors of the arrows in the map indicate the date that each interstate transmission route initiated. Vector and raster map data were obtained from Natural Earth and can be found on https://naturalearth.s3.amazonaws.com/10m_cultural/ne_10m_admin_1_states_provinces.zip).

https://doi.org/10.1371/journal.pntd.0009835.g002

Discussion

The ongoing surge of SARS-CoV-2 in Brazil since the end of 2020 has turned the country into the epicenter of a very fast spread of new variants [3,4]. In the present work, we have conducted genomic surveillance of SARS-CoV-2 spread and evolution in historically undersampled regions of Brazil. We have reconstructed past interstate transmission routes across Brazil through phylodynamic analyses of P.1 and P.2 lineages. We have also inferred the origin of P.1, suggested to be the cause of a drastic resurgence in COVID-19 cases [43], to have

occurred around August. In contrast, phylogenetic analyses of P.2 indicate that the lineage originated in February 2020, when the virus was first reported in the country and is evolving into differentiated clades.

Our genomic surveillance has evaluated the frequency of lineages currently circulating in each sampled state. As expected, proximity to the Amazonas state seems to be correlated to the pervasiveness of P.1 lineage, as exemplified by the variation observed between Rio Grande do Norte, Paraíba and Rio de Janeiro. The relatively low frequency of P.1 and high frequency of P.2 in our sample from the south of the state of Bahia, a region distant from large airports, may shed light on a much more complex relation between traveling and viral dynamics rather than guilt by association (i.e., mere vicinity). Indeed, previous works suggest that viral spread in smaller or distant cities may happen in a first-come-first-get dynamic, with one lineage overtaking the population [44–46]. Beyond south Bahia cities, this can be seen on Amazonas, where all four samples were from P.1 lineage. Low viral diversity decreases the likelihood of recombination between lineages during a coinfection, which could create new combinations of mutations and more aggressive variants [6]. These results reinforce the importance of local and international traveling restrictions as a preventive measure to slow the spread of the virus [47], measures still not enforced in Brazil and many other countries. As an alternative, policies such as social distancing and early detection of more pathogenic variants could have curtailed the spread of P.1 and P.2 across states and unburdened the public health system [48–50].

Some lineages analyzed in this work require attention due to their evolutionary dynamics. First, we observed that P.2 lineage has differentiated in several subclades between April and September of 2020 (Fig 3), all of which are present in many states. The occurrence of P.2 subclades, in practice, means that the expected evolutionary course is for these subclades to evolve into whole new lineages with exclusive mutations. If uncontrolled, epidemiological parameters such as transmission rate, lethality, and immune response escape may vary within the lineage, hindering its containment [51]. Secondly, we have confirmed the spread since December of P.7 [6] to Paraíba, and Bahia's states, possibly from the Rio Grande do Sul. Not only has it crossed a continental distance, but P.7 has also been detected in England, Japan, and the Netherlands. Higher sampling and investigation of past and present transmission routes is urgent to stop further spread.

The occurrence of lineage N.9, derived from B.1.1.33, in northeastern Brazil has also been demonstrated. According to our evolutionary inference, this lineage may have originated in Rio de Janeiro and disseminated across Brazil. This variant was also detected in samples from the United States, Ireland and Singapore. Remarkably, we observed the convergent occurrence of E484K mutation in this new clade. This mutation was first detected in B.1.351 sequences from South Africa [52] but has now independently emerged in several lineages globally, including P.1 and P.2 [3–5]. Another example of convergent evolution is the single sequence classified as B.1.1.306, which carries not only the mutation E484K inherited from N.9, but also the N501Y variant on the Spike protein gene. N501Y mutation was firstly identified in B.1.1.7 lineage in the United Kingdom [53] and recently detected in the P.1 lineage [3,4]. Finally, the third newly-detected convergent event described in this work is the E484K and N439K variants in a sample of B.1.1.29 from Rio Grande do Norte. The N439K mutation was also first detected at B.1.1.7 samples from the United Kingdom.

Convergent mutations seem to play an essential role in the evolutionary dynamics of SARS-CoV-2. Intense selective pressure from the immune system against prolonged infections may promote intrahost variants with higher adaptive value [25,54–58]. Previous studies have shown that both N501Y and E484K have independently emerged in patients with persistent infection [54,59]. Indeed, all convergent mutations aforementioned are somehow associated with viral escape from immune system response: N439K has shown to escape immune escape

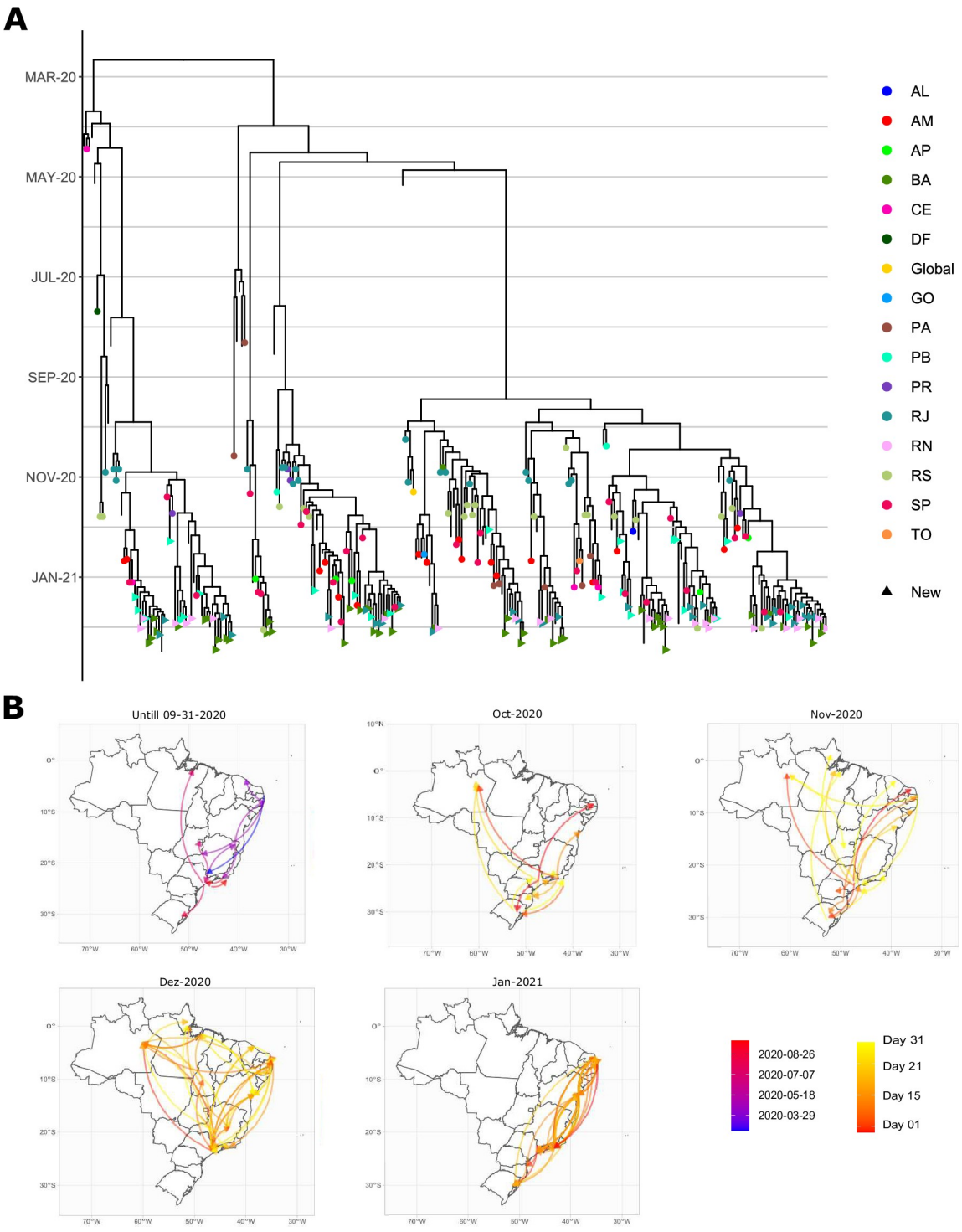

Fig 3. Divergence times within P.2 lineage (A) and its dispersal routes (B). Colors of tip points in the tree indicate the origin of P.2 samples (AL: Alagoas, AM: Amazonas, AP: Amapá, BA: Bahia, CE: Ceará, DF: Distrito Federal, GO: Goiás, PA: Pará, PB: Paraíba, PR: Paraná, RJ: Rio de Janeiro, RN: Rio Grande do Norte, RS: Rio Grande do Sul, SP: São Paulo, TO: Tocantins), while tips without a point are sequences from other lineages. Colors of the arrows in the map indicate the date that each interstate transmission route initiated. Vector and raster map data were obtained from Natural Earth and can be found on https://naturalearth.s3.amazonaws.com/10m_cultural/ne_10m_admin_1_states_provinces.zip.

https://doi.org/10.1371/journal.pntd.0009835.g003

from both polyclonal and monoclonal antibodies [60,61]; E484K has been associated with escape from both vaccines and previous infections [2,10,62–64]; and N501Y leads to increased binding specificity to the receptor and is associated with high transmissibility while also escaping immune response [65,66]. Altogether, the combination of these mutations raises the variant's fitness even higher, and increases the chance of the variant sequence becoming a new and dominant lineage [65]. Continuous monitoring of the convergent sequences here described is fundamental to follow their development and prevent spread in a worst-case-scenario.

Implementing suitable genomic surveillance approaches through sequencing samples selected randomly from PCR-positive tests is a powerful tool to monitor known and new variants across the country. It can guide the elaboration of efficient governmental policies that avoid the collapse of the national healthcare system, as happened in Brazil in the first months of 2021. Both targeted screening and random sampling methods are complementary and congruent to an adequate evaluation of the current pandemic status. Of note, the analyses conducted here are highly dependent on broad sequence sampling through both time and space, which requires both technical and human resources training. Consequently, genomic surveillance is undertaken only by a handful of laboratories, much less than needed to cover a continental-sized country such as Brazil efficiently. This causes a spatial sampling bias, which removes pieces of the historical puzzle that is the reconstruction of dispersal routes. Moreover, the underrepresented states are located in historically underfunded regions, exemplified by the North and Northeastern ones. Scientific collaborations, such as those conducted here, bypass regional barriers to monitor the advances of new and known lineages across states and foment an integrated analysis on the status of the pandemic in the country as a whole. Unfortunately, Brazil has become an open-air laboratory to the emergence and rapid dispersal of novel SARS-CoV-2 variants. Country-wide genomic surveillance is a significant step to better understand the origin and spread of new lineages.

Supporting information

S1 Fig. Correlation between root-to-tip distance and sequence sample dates. Samples P.1 lineage (red) evolved under the same clock dynamics that outgroup sequences (black), whereas P.2 (blue) do not obey the strict clock model.
(TIF)

S2 Fig. Evolutionary relationship between the 185 newly-sequenced genomes. Branches of the tree are colored according to the lineage the sequences are classified into.
(TIF)

S3 Fig. Frequency of SARS-CoV-2 lineages across Brazilian states. Barplot showing the relative frequency of the 13 lineages found in this study in Amazonas (North region), Rio Grande do Norte, Paraíba, Bahia (all three in the Northeast region), and Rio de Janeiro (Southeast region).
(TIF)

S4 Fig. Genomic characterization of SARS-CoV-2 mutations identified. Distribution of single-nucleotide variants (SNVs) found in the 185 genomes sequenced in this study. Each vertical line represents the relative variant frequency in the total number of genomes sequenced and its target protein products. The receptor-binding domain (RBD) highlighted in red showed the main mutations associated with P.2 and the variant of concern P.1. Density plot shows the accumulation of mutations across the SARS-CoV-2 genome.
(TIF)

S5 Fig. Dispersal routes of P.1 lineage inferred for the ten subsampled datasets. Colors of the arrows in the map indicate the date that each interstate transmission route initiated. Vector and raster map data were obtained from Natural Earth and can be found on https://naturalearth.s3.amazonaws.com/10m_cultural/ne_10m_admin_1_states_provinces.zip).
(TIF)

S6 Fig. Dispersal routes of P.2 lineage inferred for the ten subsampled datasets. Colors of the arrows in the map indicate the date that each interstate transmission route initiated. Vector and raster map data were obtained from Natural Earth and can be found on https://naturalearth.s3.amazonaws.com/10m_cultural/ne_10m_admin_1_states_provinces.zip.
(TIF)

S1 Table. Sample information.
(XLSX)

S2 Table. Frequency of mutations in SARS-CoV-2 genome.
(XLSX)

S3 Table. Acknowledgement for GISAID samples.
(PDF)

S1 File. Phylogenetic trees in Figs 1–3, written in NEXUS format.
(TXT)

Acknowledgments

We would like to thank all authors and the administrators of the GISAID database, allowing this genomic epidemiology study to be properly conducted. A full list of acknowledgment is available in S3 Table. A list acknowledging those from different institutions that participated in this study follows below:

Workgroup members

LABIMOL/ENDEMIAS/UFPB

Álisson Emannuel Franco Alves, Ana Beatriz Rodrigues dos Santos, Brena Ferreira dos Santos, Bruno Henrique Andrade Galvão, Daniela Letícia Torres da Silva, Fabio Marcel da Silva Santos, Fabrine Felipe Hilário, Gabriel Rodrigues Martins de Freitas, Marília Gabriela dos Santos Cavalcanti, Mayara Karla dos Santos Nunes, Moises Dantas Cartaxo de Abreu Pereira, Naiara Naiana Dejani, Sandrelli Meridiana de Fátima Ramos dos Santos Medeiros, Sergio Dias da Costa Junior, Talita Nayara Bezerra Lins, Wallace Felipe Blohem Pessoa

LAFEM/UESC

Mylene de Melo Silva, Renato Fontana, Renata Santiago Alberto Carlos, Galileu Barbosa Costa, Hllytchaikra Ferraz Fehlberg, Amanda Teixeira Sampaio Lopes, Íris Terezinha Santos de Santana, Fabrício Barbosa Ferreira, Luciano Cardoso Santos, Luane Etienne Barreto, Pérola Rodrigues dos Santos, Laíne Lopes Silva de Jesus, Thiago Silva Gonçalves, Gabriela Andrade Coelho Dias

POLICLÍNICA PIQUET CARNEIRO/UERJ

Kennedy Martins Kirk, Claudia Henrique da Costa, Rogerio Rufino, Claudia Henrique da Costa, Renata Salles Miraldi Oliveira, Gabriella da Silva Alves, Allan Motta Leal Pontes, Sandra

Pereira C. Vilas Boas, Vania Maria Almeida de Souza, Vinícius Miranda Porto, Jeane de Souza Nogueira.

DACT/CCS/CB/IMT/UFRN

Igor de Farias Domingos, Antonia Cláudia J. Câmara, Ivanise M. Moretti Rebecchi, Ana Cláudia G. Freire, Marcela A. Galvão Ururahy, Vivian N. Silbiger, André D. Luchessi, Antonnyo P. D. Lima, Thiala S. J. da Silva Parente, Carlos R. do Nascimento Brito, Kátia Castanho Scortecci, Susana M. Gomes Moreira, Daniella R. A. Martins Salha, Leonardo C. Ferreira, Waleska R. D. B. de Medeiros, Dayse Santos Arimateia, Arthur R. de Araújo Oliveira, Fernanda M. de Azevedo, João F. Rodrigues Neto, Glória Regina de Gois Monteiro, Paulo Ricardo P. do Nascimento, Ingryd Camara Morais, Francisco Paulo Freire Neto, Eliana L. Tomaz Nascimento, Iara Marques Medeiros, Ana Rafaela de Souza Timóteo.

LNCC/MCTI

Luciane Prioli Ciapina, Rangel Celso Souza, Éllen dos Santos Correa, Guilherme Cordenonsi da Fonseca, Vinícius Prata Klôh, Eduardo Wagner.

Author Contributions

Conceptualization: Ana Tereza R. Vasconcelos.

Formal analysis: Alessandra P. Lamarca, Luiz G. P. de Almeida, Ronaldo da Silva Francisco, Jr., Otavio J. Brustolini, Alexandra L. Gerber, Ana Paula de C. Guimarães.

Funding acquisition: Ana Tereza R. Vasconcelos.

Investigation: Alessandra P. Lamarca, Luiz G. P. de Almeida, Ronaldo da Silva Francisco, Jr.

Methodology: Alessandra P. Lamarca, Luiz G. P. de Almeida, Ronaldo da Silva Francisco, Jr., Otavio J. Brustolini, Alexandra L. Gerber, Ana Paula de C. Guimarães.

Project administration: Lucymara Fassarella Agnez Lima, Kátia Castanho Scortecci, Vinícius Pietta Perez, Sandra Rocha Gadelha, Luís Cristóvão Porto, Eloiza Helena Campana, Selma Maria Bezerra Jeronimo, Ana Tereza R. Vasconcelos.

Resources: Luiz G. P. de Almeida, Lucymara Fassarella Agnez Lima, Kátia Castanho Scortecci, Vinícius Pietta Perez, Otavio J. Brustolini, Eduardo Sérgio Soares Sousa, Danielle Angst Secco, Angela Maria Guimarães Santos, George Rego Albuquerque, Ana Paula Melo Mariano, Bianca Mendes Maciel, Alexandra L. Gerber, Ana Paula de C. Guimarães, Paulo Ricardo Nascimento, Francisco Paulo Freire Neto, Sandra Rocha Gadelha, Luís Cristóvão Porto, Eloiza Helena Campana, Selma Maria Bezerra Jeronimo, Ana Tereza R. Vasconcelos.

Software: Luiz G. P. de Almeida, Otavio J. Brustolini.

Supervision: Sandra Rocha Gadelha, Luís Cristóvão Porto, Eloiza Helena Campana, Selma Maria Bezerra Jeronimo, Ana Tereza R. Vasconcelos.

Validation: Alessandra P. Lamarca.

Visualization: Alessandra P. Lamarca, Ronaldo da Silva Francisco, Jr.

Writing – original draft: Alessandra P. Lamarca, Luiz G. P. de Almeida, Ronaldo da Silva Francisco, Jr., Lucymara Fassarella Agnez Lima, Kátia Castanho Scortecci, Vinícius Pietta

Perez, George Rego Albuquerque, Luís Cristóvão Porto, Eloiza Helena Campana, Selma Maria Bezerra Jeronimo, Ana Tereza R. Vasconcelos.

Writing – review & editing: Alessandra P. Lamarca, Luiz G. P. de Almeida, Ronaldo da Silva Francisco, Jr., Ana Tereza R. Vasconcelos.

References

1. Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science. 2020; 369: 1255–1260. https://doi.org/10.1126/science.abd2161 PMID: 32703910
2. Voloch CM, da Silva Francisco R Jr, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. J Virol. 2021. https://doi.org/10.1128/JVI.00119-21 PMID: 33649194
3. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021; 372: 815–821. https://doi.org/10.1126/science.abh2644 PMID: 33853970
4. Naveca FG, Nascimento V, de Souza VC, Corado A de L, Nascimento F, Silva G, et al. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. Nat Med. 2021; 27: 1230–1238. https://doi.org/10.1038/s41591-021-01378-7 PMID: 34035535
5. Voloch CM, da Silva Francisco R Jr, de Almeida LGP, Cardoso CC, Brustolini OJ, Gerber AL, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. J Virol. 2021. https://doi.org/10.1128/JVI.00119-21 PMID: 33649194
6. Francisco R da S Jr, Benites LF, Lamarca AP, de Almeida LGP, Hansen AW, Gularte JS, et al. Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. Virus Res. 2021; 296: 198345. https://doi.org/10.1016/j.virusres.2021.198345 PMID: 33631222
7. Francisco Junior R da S, Lamarca AP, de Almeida LGP, Cavalcante L, Machado DT, Martins Y, et al. Turnover of SARS-CoV-2 lineages shaped the pandemic and enabled the emergence of new variants in the state of Rio de Janeiro, Brazil. bioRxiv. medRxiv; 2021. https://doi.org/10.1101/2021.07.20.21260890
8. Vasques Nonaka CK, Miranda Franco M, Gräf T, Almeida Mendes AV, Santana de Aguiar R, Giovanetti M, et al. Genomic evidence of a SARS-CoV-2 reinfection case with E484K spike mutation in Brazil. Preprints. 2021. https://doi.org/10.20944/preprints202101.0132.v1
9. Resende PC, Bezerra JF, de Vasconcelos RHT, Arantes I, Appolinario L, Mendonça AC, et al. Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in Brazil, 2020. In: Virological [Internet]. 10 Jan 2021 [cited 8 Mar 2021]. Available: https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584
10. Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. Cell Host Microbe. 2021; 29: 44–57.e9. https://doi.org/10.1016/j.chom.2020.11.007 PMID: 33259788
11. Nonaka CKV, Franco MM, Gräf T, de Lorenzo Barcia CA, de Ávila Mendonça RN, de Sousa KAF, et al. Genomic Evidence of SARS-CoV-2 Reinfection Involving E484K Spike Mutation, Brazil. Emerg Infect Dis. 2021; 27: 1522–1524. https://doi.org/10.3201/eid2705.210191 PMID: 33605869
12. Naveca F, Costa C da, Nascimento V, Souza V, Corado A, Nascimento F, et al. Three SARS-CoV-2 reinfection cases by the new Variant of Concern (VOC) P.1/501Y.V3. Research Square. Research Square; 2021. https://doi.org/10.21203/rs.3.rs-318392/v1
13. Nicolelis MAL, Raimundo RLG, Peixoto PS, Andreazzi CS. The impact of super-spreader cities, highways, and intensive care availability in the early stages of the COVID-19 epidemic in Brazil. Sci Rep. 2021; 11: 13001. https://doi.org/10.1038/s41598-021-92263-3 PMID: 34155241
14. Ribeiro SP, Castro E Silva A, Dáttilo W, Reis AB, Góes-Neto A, Alcantara LCJ, et al. Severe airport sanitarian control could slow down the spreading of COVID-19 pandemics in Brazil. PeerJ. 2020; 8: e9446. https://doi.org/10.7717/peerj.9446 PMID: 32617196
15. Martins AF, Zavascki AP, Wink PL, Volpato FCZ, Monteiro FL, Rosset C, et al. Detection of SARS-CoV-2 lineage P.1 in patients from a region with exponentially increasing hospitalisation rate, February 2021, Rio Grande do Sul, Southern Brazil. Euro Surveill. 2021; 26. https://doi.org/10.2807/1560-7917.ES.2021.26.12.2100276 PMID: 33769251

16. Salvato RS, Gregianini TS, Campos AAS, Crescente LV, Vallandro MJ, Ranieri TMS, et al. Epidemiological investigation reveals local transmission of SARS-CoV-2 lineage P.1 in Southern Brazil. Research Square. Research Square; 2021. https://doi.org/10.21203/rs.3.rs-280297/v1

17. de Siqueira IC, Camelier AA, Maciel EAP, Nonaka CKV, Neves MCLC, Macêdo YSF, et al. Early detection of P.1 variant of SARS-CoV-2 in a cluster of cases in Salvador, Brazil. Int J Infect Dis. 2021; 108: 252–255. https://doi.org/10.1016/j.ijid.2021.05.010 PMID: 33989776

18. Adamoski D, de Oliveira JC, Bonatto AC, Wassem R, Nogueira MB, Raboni SM, et al. Large-scale screening of asymptomatic for SARS-CoV-2 variants of concern and rapid P.1 takeover, Curitiba, Brazil. bioRxiv. medRxiv; 2021. https://doi.org/10.1101/2021.06.18.21258649

19. Barbosa GR, Leão Moreira LV, Oliveira Justo AF, Perosa AH, Cunha Chaves AP, Bueno MS, et al. Rapid spread and high impact of the Variant of Concern P.1 in the largest city of Brazil. bioRxiv. medRxiv; 2021. https://doi.org/10.1016/j.jinf.2021.04.008 PMID: 33865897

20. Tosta S, Giovanetti M, Brandão Nardy V, de Oliveira da Silva LR, Gómez MKA, Lima JG, et al. Early genomic detection of SARS-CoV-2 P.1 variant in Northeast Brazil. medRxiv; 2021. https://doi.org/10.1371/journal.pntd.0009591 PMID: 34280196

21. Naveca F, Nascimento V, Souza V, Corado A, Nascimento F, Silva G, et al. COVID-19 epidemic in the Brazilian state of Amazonas was driven by long-term persistence of endemic SARS-CoV-2 lineages and the recent emergence of the new Variant of Concern P.1. Research Square. Research Square; 2021. https://doi.org/10.21203/rs.3.rs-275494/v1

22. Vogels CBF, Breban MI, Ott IM, Alpert T, Petrone ME, Watkins AE, et al. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. PLoS Biol. 2021; 19: e3001236. https://doi.org/10.1371/journal.pbio.3001236 PMID: 33961632

23. Lopez-Rincon A, Perez-Romero CA, Tonda A, Mendoza-Maldonado L, Claassen E, Garssen J, et al. Design of Specific Primer Sets for the Detection of B.1.1.7, B.1.351, P.1, B.1.617.2 and B.1.1.519 Variants of SARS-CoV-2 using Artificial Intelligence. bioRxiv. 2021. p. 2021.01.20.427043. https://doi.org/10.1101/2021.01.20.427043

24. CDC. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. In: FDA [Internet]. 2020 [cited 10 Mar 2021]. Available: https://www.fda.gov/media/134922/download

25. Voloch CM, da Silva F R, de Almeida LGP, Brustolini OJ, Cardoso CC, Gerber AL, et al. Intra-host evolution during SARS-CoV-2 persistent infection. medRxiv. 2020; 2020.11.13.20231217.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168

28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

29. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43: 491–498. https://doi.org/10.1038/ng.806 PMID: 21478889

30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

31. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011. pp. 2987–2993. https://doi.org/10.1093/bioinformatics/btr509 PMID: 21903627

32. Li H. Improving SNP discovery by base alignment quality. Bioinformatics. 2011; 27: 1157–1158. https://doi.org/10.1093/bioinformatics/btr076 PMID: 21320865

33. Paiva MHS, Guedes DRD, Docena C, Bezerra MF, Dezordi FZ, Machado LC, et al. Multiple Introductions Followed by Ongoing Community Spread of SARS-CoV-2 at One of the Largest Metropolitan Areas of Northeast Brazil. Viruses. 2020; 12. https://doi.org/10.3390/v12121414 PMID: 33316947

34. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012; 28: 3150–3152. https://doi.org/10.1093/bioinformatics/bts565 PMID: 23060610

35. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

36. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14: 587–589. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

37. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018; 4: vey016. https://doi.org/10.1093/ve/vey016 PMID: 29942656

38. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016; 2: vew007. https://doi.org/10.1093/ve/vew007 PMID: 27774300

39. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006; 4: e88. https://doi.org/10.1371/journal.pbio.0040088 PMID: 16683862

40. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. Mol Biol Evol. 2010; 27: 1877–1885. https://doi.org/10.1093/molbev/msq067 PMID: 20203288

41. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc Natl Acad Sci U S A. 2012; 109: 15066–15071. https://doi.org/10.1073/pnas.1206598109 PMID: 22927414

42. Dellicour S, Rose R, Faria NR, Lemey P, Pybus OG. SERAPHIM: studying environmental rasters and phylogenetically informed movements. Bioinformatics. 2016; 32: 3204–3206. https://doi.org/10.1093/bioinformatics/btw384 PMID: 27334476

43. Sabino EC, Buss LF, Carvalho MPS, Prete CA Jr, Crispim MAE, Fraiji NA, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. Lancet. 2021; 397: 452–455. https://doi.org/10.1016/S0140-6736(21)00183-5 PMID: 33515491

44. Kouriba B, Dürr A, Rehn A, Sangaré AK, Traoré BY, Bestehorn-Willmann MS, et al. First Phylogenetic Analysis of Malian SARS-CoV-2 Sequences Provides Molecular Insights into the Genomic Diversity of the Sahel Region. Viruses. 2020; 12. https://doi.org/10.3390/v12111251 PMID: 33147840

45. Viedma E, Dahdouh E, González-Alba JM, González-Bodi S, Martínez-García L, Lázaro-Perona F, et al. Genomic Epidemiology of SARS-CoV-2 in Madrid, Spain, during the First Wave of the Pandemic: Fast Spread and Early Dominance by D614G Variants. Microorganisms. 2021; 9. https://doi.org/10.3390/microorganisms9020454 PMID: 33671631

46. Githinji G, de Laurent ZR, Mohammed KS, Omuoyo DO, Macharia PM, Morobe JM, et al. Tracking the introduction and spread of SARS-CoV-2 in coastal Kenya. bioRxiv. medRxiv; 2020. https://doi.org/10.1101/2020.10.05.20206730

47. Chang M-C, Kahn R, Li Y-A, Lee C-S, Buckee CO, Chang H-H. Variation in human mobility and its impact on the risk of future COVID-19 outbreaks in Taiwan. BMC Public Health. 2021; 21: 226. https://doi.org/10.1186/s12889-021-10260-7 PMID: 33504339

48. Heck TG, Frantz RZ, Frizzo MN, François CHR, Ludwig MS, Mesenburg MA, et al. Insufficient social distancing may contribute to COVID-19 outbreak: The case of Ijuí city in Brazil. PLoS One. 2021; 16: e0246520. https://doi.org/10.1371/journal.pone.0246520 PMID: 33596229

49. Teslya A, Pham TM, Godijk NG, Kretzschmar ME, Bootsma MCJ, Rozhnova G. Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study. PLoS Med. 2020; 17: e1003166. https://doi.org/10.1371/journal.pmed.1003166 PMID: 32692736

50. McCombs A, Kadelka C. A model-based evaluation of the efficacy of COVID-19 social distancing, testing and hospital triage policies. PLoS Comput Biol. 2020; 16: e1008388. https://doi.org/10.1371/journal.pcbi.1008388 PMID: 33057438

51. Koma T, Adachi S, Doi N, Adachi A, Nomaguchi M. Toward Understanding Molecular Bases for Biological Diversification of Human Coronaviruses: Present Status and Future Perspectives. Front Microbiol. 2020; 11: 2016. https://doi.org/10.3389/fmicb.2020.02016 PMID: 32983025

52. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 2021; 592: 438–443. https://doi.org/10.1038/s41586-021-03402-9 PMID: 33690265

53. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. In: Virological [Internet]. 18 Dec 2020 [cited 10 Mar 2021]. Available: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

54. Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF, Qiu X, et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N Engl J Med. 2020; 383: 2291–2293. https://doi.org/10.1056/NEJMc2031364 PMID: 33176080

55. Martin DP, Weaver S, Tegally H, San EJ, Shank SD, Wilkinson E, et al. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2

selective landscape. bioRxiv. medRxiv; 2021. https://doi.org/10.1101/2021.02.23.21252268 PMID: 33688681

56. Zhang Y-Z, Holmes EC. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell. 2020; 181: 223–227. https://doi.org/10.1016/j.cell.2020.03.035 PMID: 32220310

57. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. 2021. https://doi.org/10.1126/science.abf6950 PMID: 33536258

58. Siqueira JD, Goes LR, Alves BM, de Carvalho PS, Cicala C, Arthos J, et al. SARS-CoV-2 genomic analyses in cancer patients reveal elevated intrahost genetic diversity. Virus Evol. 2021. https://doi.org/10.1093/ve/veab013 PMID: 33738124

59. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature. 2021. https://doi.org/10.1038/s41586-021-03291-y PMID: 33545711

60. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, da Silva Filipe A, Wojcechowskyj JA, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell. 2021; 184: 1171–1187.e20. https://doi.org/10.1016/j.cell.2021.01.037 PMID: 33621484

61. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell. 2020; 182: 1284–1294.e9. https://doi.org/10.1016/j.cell.2020.07.012 PMID: 32730807

62. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife. 2020; 9. https://doi.org/10.7554/eLife.61312 PMID: 33112236

63. Gaebler C, Wang Z, Lorenzi JCC, Muecksch F, Finkin S, Tokuyama M, et al. Evolution of antibody immunity to SARS-CoV-2. Nature. 2021. https://doi.org/10.1038/s41586-021-03207-w PMID: 33461210

64. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. bioRxiv. 2021. https://doi.org/10.1101/2021.01.15.426911 PMID: 33501451

65. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe. 2021. https://doi.org/10.1016/j.chom.2021.02.003 PMID: 33592168

66. Ali F, Kasry A, Amin M. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. Med Drug Discov. 2021; 100086.