medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Phylogenetic-based inference reveals distinct transmission dynamics of SARS-CoV-2 variant of concern Gamma and lineage P.2 in Brazil

Tiago Graf*[1], Gonzalo Bello*[2], Felipe Gomes Naveca[3], Marcelo Gomes[4], Vanessa Leiko Oikawa Cardoso[1], Alexandre Freitas da Silva[5], Filipe Zimmer Dezordi[5], Mirleide Cordeiro dos Santos[6], Katia Correa de Oliveira Santos[7], Erika Lopes Rocha Batista[8], Alessandro Leonardo alvares Magalhaes[8], Fernando Vinhal[9], Brazilian Ministry of Health COVID-19 Genomic Surveillance Network[10]**, Fabio Miyajima[11], Helisson Faoro[12], Ricardo Khouri[1], Gabriel Luz Wallau[5], Edson Delatorre[13], Marilda Mendonca Siqueira[14], Paola Cristina Resende[14]
* These authors contributed equally to this work.
** List of authors and affiliations (Appendix 1)

Affiliations
[1] Instituto Goncalo Moniz, Fundacao Oswaldo Cruz, Salvador, Brazil
[2] Laboratorio de AIDS e Imunologia Molecular, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil
[3] Laboratorio de Ecologia de Doencas Transmissiveis na Amazônia (EDTA), Leônidas e Maria Deane Institute, Fiocruz
[4] Grupo de Metodos Analiticos em Vigilancia Epidemiologica, Programa de Computacao Cientifica (PROCC), Fiocruz, Rio de Janeiro, Brazil
[5] Instituto Aggeu Magalhaes, Fundacao Oswaldo Cruz, Recife, Pernambuco, Brazil
[6] Instituto Evandro Chagas, Belem, Para
[7] Instituto Adolfo Lutz, Sao Paulo
[8] Secretaria de Saude de Aparecida de Goiania, Goias, Brazil
[9] Lab HLAGYN - Goiania, Goias, Brazil
[10] Brazilian Ministry of Health COVID-19 Genomic Surveillance Network, Av. Brasil, 4365 - Manguinhos, Rio de Janeiro - CEP: 21040-900
[11] Fundacao Oswaldo Cruz - Fiocruz Ceara
[12] Instituto Carlos Chagas (ICC), Fiocruz-PR
[13] Departamento de Biologia. Centro de Ciencias Exatas, Naturais e da Saude, Universidade Federal do Espirito Santo, Alegre, Brazil.
[14] Laboratory of Respiratory Viruses and Measles, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, RJ, Brazil.

Corresponding authors:
Tiago Graf: tiago.graf@fiocruz.br
Gonzalo Bello: gbellobr@gmail.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Abstract

The COVID-19 epidemic in Brazil experienced two major country-wide lineage replacements, the first driven by the lineage P.2, formerly classified as variant of interest (VOI) Zeta in late 2020 and the second by the variant of concern (VOC) Gamma in early 2021. To better understand how these SARS-CoV-2 lineage turnovers occurred in Brazil, we analyzed 11,724 high-quality SARS-CoV-2 whole genomes of samples collected in different country regions between September 2020 and April 2021. Our findings indicate that the spatial dispersion of both variants in Brazil was driven by short and long-distance viral transmission. The lineage P.2 harboring Spike mutation E484K probably emerged around late July 2020 in the Rio de Janeiro (RJ) state, which contributed with most (~50%) inter-state viral disseminations, and only became locally established in most Brazilian states by October 2020. The VOC Gamma probably arose in November 2020 in the Amazonas (AM) state, which was responsible for 60-70% of the inter-state viral dissemination, and the earliest timing of community transmission of this VOC in many Brazilian states was already traced to December 2020. We estimate that variant Gamma was 1.56-3.06 more transmissible than variant P.2 co-circulating in RJ and that the median effective reproductive number (Re) of Gamma in RJ and SP states (Re = 1.59-1.91) was lower than in AM (Re = 3.55). In summary, although the epicenter of the lineage P.2 dissemination in Brazil was the heavily interconnected Southeastern region, it displayed a slower rate of spatial spread than the VOC Gamma originated in the more isolated Northern Brazilian region. Our findings also support that the VOC Gamma was more transmissible than lineage P.2, although the viral Re of the VOC varied according to the geographic context.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Introduction

The COVID-19 epidemic in Brazil during 2020 was mostly driven by SARS-CoV-2 lineages B.1.1.28 and B.1.1.33 in most country regions, and by lineages B.1.1 and B.1.195 at some specific states [1-3]. In late 2020, two descendant sub-lineages of B.1.1.28 emerged in Brazil and replaced dominant lineages across the country. The first lineage turnover observed during late 2020 was associated with the emergence and dissemination of the lineage P.2, formerly defined Variant of Interest (VOI) Zeta, which harbors mutation E484K in the Spike protein [4]. The lineage P.2 was first detected in the Rio de Janeiro (RJ) state in October 2020 and became the most prevalent lineage in several Brazilian states from November 2020 to January 2021 [4-6]. The second country-wide lineage turnover event took place after the emergence of the Variant of Concern (VOC) Gamma in the Amazonas (AM) state in November 2020 [3,7]. The VOC Gamma harbors multiple mutations in the Spike protein and becomes the dominant lineage in all Brazilian regions by early 2021 [6,8-11].

Phylodynamic reconstruction of the spatiotemporal spread pattern of lineages B.1.1.28 and B.1.1.33 supports that densely populated and well-connected urban centers in the Southeastern region were the main sources of inter-regional SARS-CoV-2 spread during the first epidemic wave in 2020 in Brazil [2]. Mathematical modeling also supports that São Paulo (SP) and RJ were the main superspreader cities of SARS-CoV-2 during the first months of the pandemic in Brazil [12]. Even though lineage P.2 and VOC Gamma spread widely through Brazil since late 2020, the routes of regional dissemination and the onset date of community transmission across different states remains unclear. Furthermore, despite previous studies estimated that the effective reproductive number (Re) of the VOC Gamma was nearly two times higher than non-VOC co-circulating in the AM state [3,7,13], the Re of this VOC in other states and its relative transmissibility concerning the lineage P.2 has not been addressed yet.

Here, we used a large dataset of SARS-CoV-2 positive samples sequenced by the Brazilian Ministry of Health COVID-19 Genomic Surveillance Network (see Methods for details). Genomes retrieved from patients of all Brazilian states sampled between September 2020 and April 2021 were analyzed alongside publicly available SARS-CoV-2 sequences sampled throughout Brazil in the same period. Our findings show the changing prevalence of lineage P.2 and VOC Gamma in different country regions, reconstruct the pathways of inter-state spread, estimate the onset date of community transmission in such locations and compare the relative transmissibility of both co-circulating variants.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Results

Genomic surveillance of SARS-CoV-2 in Brazil: September 2020 to March 2021.

The COVID-19 epidemic in Brazil displayed two distinct phases over the period of investigation: a decreasing phase from around 900 deaths/day in early September to 300 deaths/day in mid-November 2020; followed by an increasing phase up to 3,000 deaths/day in late March 2021 (**Fig 1a**). A total of 11,724 high quality (<5% N) Brazilian SARS-CoV whole-genomes (29 kb) sampled between September 01st, 2020 and April 16th, 2021 were analyzed, including 5,472 generated by our collaborative genomic surveillance network (**Fig1b**) and 6,252 obtained from the EpiCoV database in GISAID (https://www.gisaid.org/), covering all Brazilian states (**Supplementary Table 1**). Based on this sampling, our genomic analysis revealed a changing pattern of SARS-CoV-2 lineage frequencies in the studied period with a high overall prevalence of lineages B.1.1.28 and B.1.1.33 during the decreasing phase, a subsequent predominance of lineage P.2 from December 2020 to January 2021, and the dominance of the VOC Gamma in February and March 2021 (**Fig 1c**). In addition, it is interesting to note that the overall proportion of local variants harbouring the mutation Spike:E484K, which comprises the VOC Gamma and the lineages P.2, N.9, N.10, sharply increased in Brazil from <1% in September 2020 to >90% in March 2021 (**Fig 1d**).

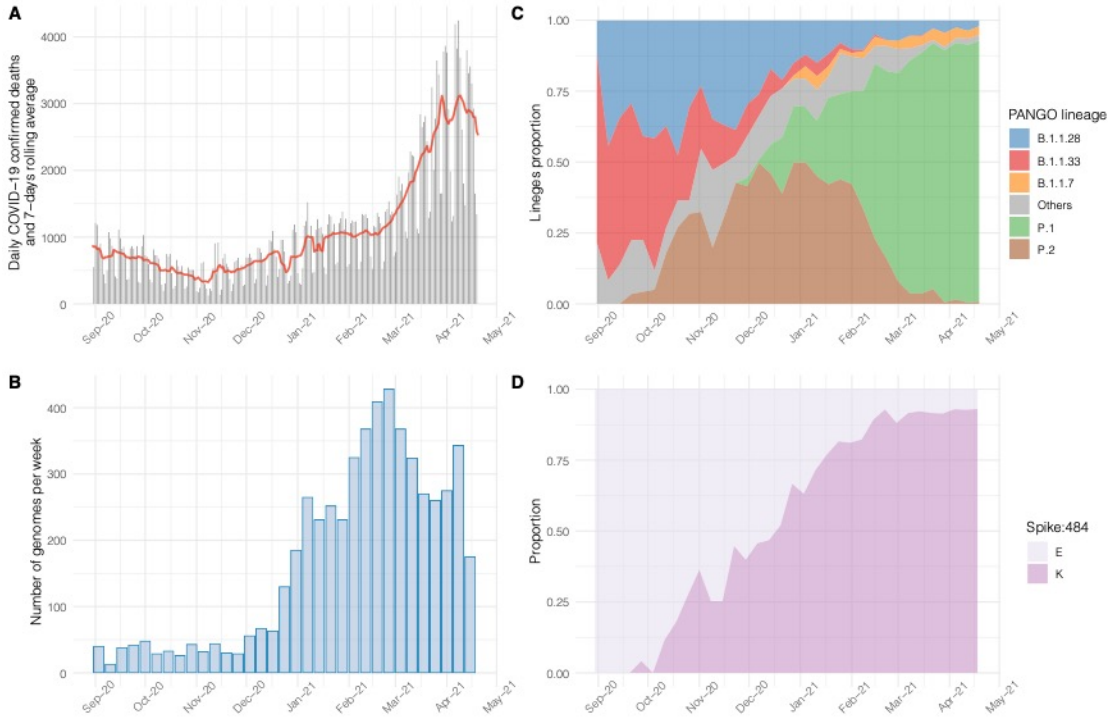

Figure 1. Genomic epidemiology of COVID-19 in Brazil from September 2020 to March 2021. a) Daily number of deaths due to COVID-19 and 7-days rolling average (line). b) Weekly number of genomes generated in this study. c) Proportion of lineages circulating in Brazil in the study period. d) Proportion of variants harboring Spike:484E/K in Brazil in the study period.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Timing the origin of variants Gamma and P.2 in Brazil.

To estimate the precise onset date of the lineage P.2 and the VOC Gamma, we first excluded all P.1 or P.2 sequences that did not display the full set of variant-defining mutations and those recently classified as Gamma-like [14]. Next, we assessed the temporal signal of the P.2 and Gamma datasets and excluded several outliers, including some of the sequences with the earliest recorded sampling dates (**Fig 2a**). The leaf-dating Bayesian molecular clock method confirmed that the earliest Gamma sequence recovered in November 2020 and three P.2 genomes sampled between April and August 2020 clearly violate the assumption of a molecular clock as displayed median expected collection dates much more recent than their recorded sampling dates (**Fig 2b**). This quality control process resulted in two final datasets of 1,357 P.2 and 2,477 Gamma Brazilian sequences sampled between 25th September 2020 - 26th March 2021 and 03rd December 2020 - 16th April 2021, respectively, mostly (>70%) generated by our collaborative genomic surveillance network. Finally, to estimate the time of the most recent common ancestor (T_{MRCA}) of each variant we generated spatiotemporal representative subsets containing 219 P.2 and 191 Gamma sequences. Bayesian molecular clock analyses traced the T_{MRCA} of the lineage P.2 to 28th July 2020 (95% HPD: 11th June-3rd September 2020) and of that of the VOC Gamma to 30th November 2020 (95% HPD: 14th November-3rd December 2020) (**Fig 2c**).

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

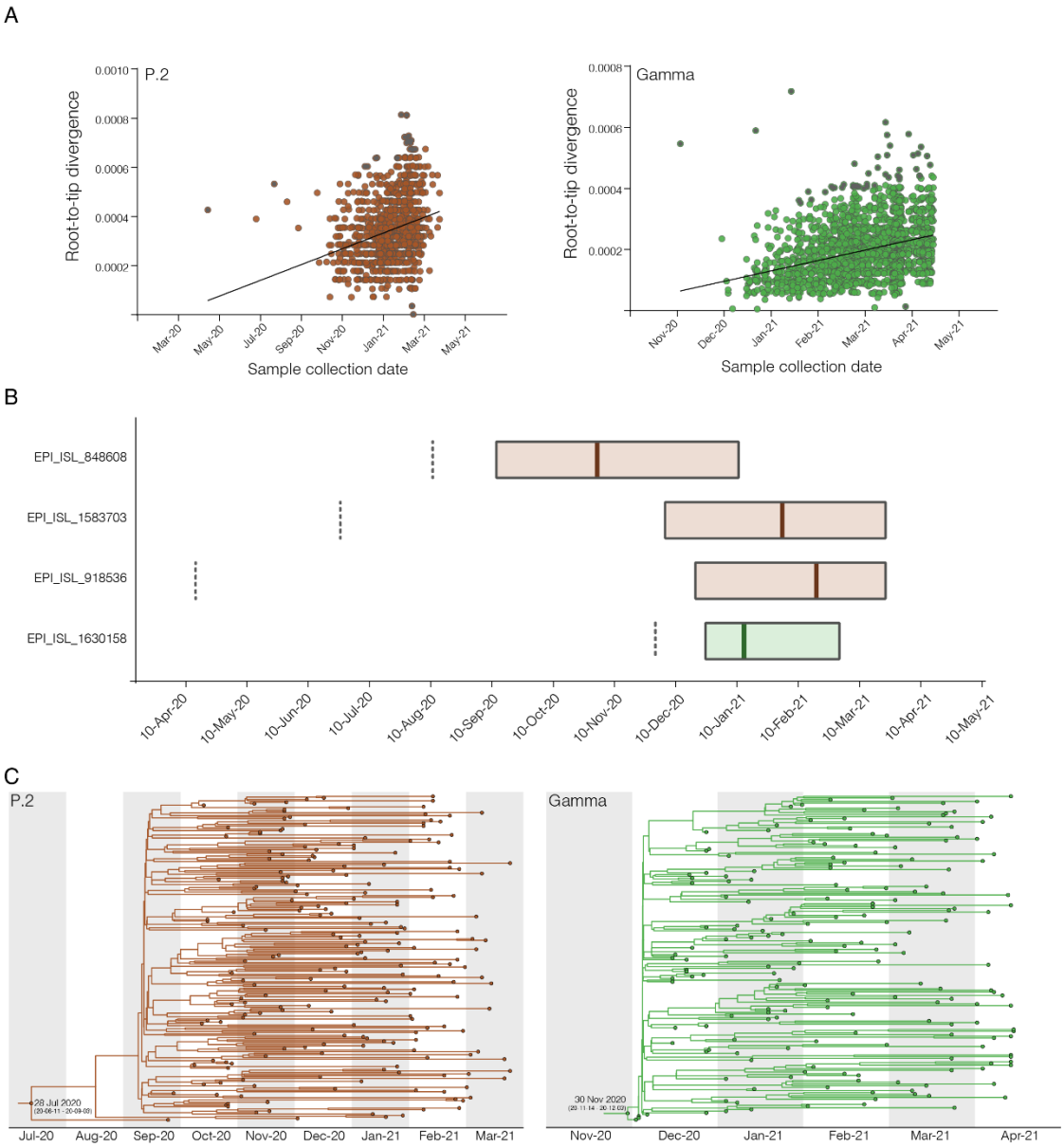

Figure 2. Temporal analysis of the P.2 and Gamma sequence datasets. a) Plots of the root-to-tip divergence against dates of sample collection for P.2 and Gamma datasets. Outliers excluded from the further analysis were colored gray. b) Leaf-dating Bayesian molecular clock of the four sequences excluded due to probably wrong collection date annotation. Reported collection dates are indicated by the dotted lines. Boxplots represent the median and the 95% highest posterior density interval of the estimated collection dates. c) Time-calibrated maximum clade credibility tree for P.2 and Gamma lineages spatial-temporal representative datasets. The median estimated onset dates (with 95% highest posterior density interval in parentheses) for each clade are shown on the root

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Inter-state spread and domestic dissemination of variants Gamma and P.2 in Brazil.

To describe the pathways that shaped the country-wide dissemination of the lineage P.2 and the VOC Gamma in Brazil, we performed a discrete Bayesian phylogeographic reconstruction. Bayesian analyses support that the lineage P.2 most probably (*PSP* = 0.71) arose in RJ (Southeastern region). This state was also the main epicenter of the regional spread of this lineage in Brazil, followed by the Southern states of Santa Catarina (SC) and Rio Grande do Sul (RS) and the Southeastern state of SP (**Fig 3a and Supplementary Video 1**). RJ contributed with ~50% of all inter-state disseminations of lineage P.2 between October 2020 and February 2021, while the four major spreading states (RJ, SP, SC and RS) combined accounted for >90% of all inter-state viral disseminations in that period (**Fig 3b**). The relative contribution of the secondary hubs SC, RS and SP to the overall dissemination of P.2 remained roughly constant between mid-October 2020 and late February 2021. The P.2 migrations from RJ to other states started in September 2020 but were mostly distributed between October 2020 and January 2021 (**Fig 3c**). The earliest onset date of community transmission of the lineage P.2 was traced to October 2020 in most Brazilian states **(Fig 3d**).

Bayesian analyses support that the VOC Gamma most probably (*PSP* = 1.0) arose in the AM (Northern region) and this state was also the main hub of Gamma spread throughout the country, followed by the Southeastern states of SP and RJ (**Fig 4a and Supplementary Video 2**). AM was responsible for 60-70% of the inter-state disseminations of the VOC Gamma between December 2020 and March 2021, while the three major spreading states combined accounted for >90% of all inter-state viral transmissions in that period (**Fig 4b**). The relative contribution of the secondary hubs (SP and RJ) to the overall dissemination of Gamma in Brazil increased from mid-December 2020 to late March 2021. The first transition events of VOC Gamma from AM into other Brazilian states occurred in December 2020 and reached a peak between January and February 2021 **(Fig 4c**); while the earliest timing of community transmission of this VOC in most Brazilian states was traced to December 2020 **(Fig 4d**). By early January 2021, this variant was already disseminated and established local transmission chains in nine out of 23 Brazilian states from all major country regions: North (Pará), Northeastern (Bahia, Maranhão and Sergipe), Southeastern (SP and RJ), Southern (SC and Paraná) and Central-Western (Goiás) regions.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

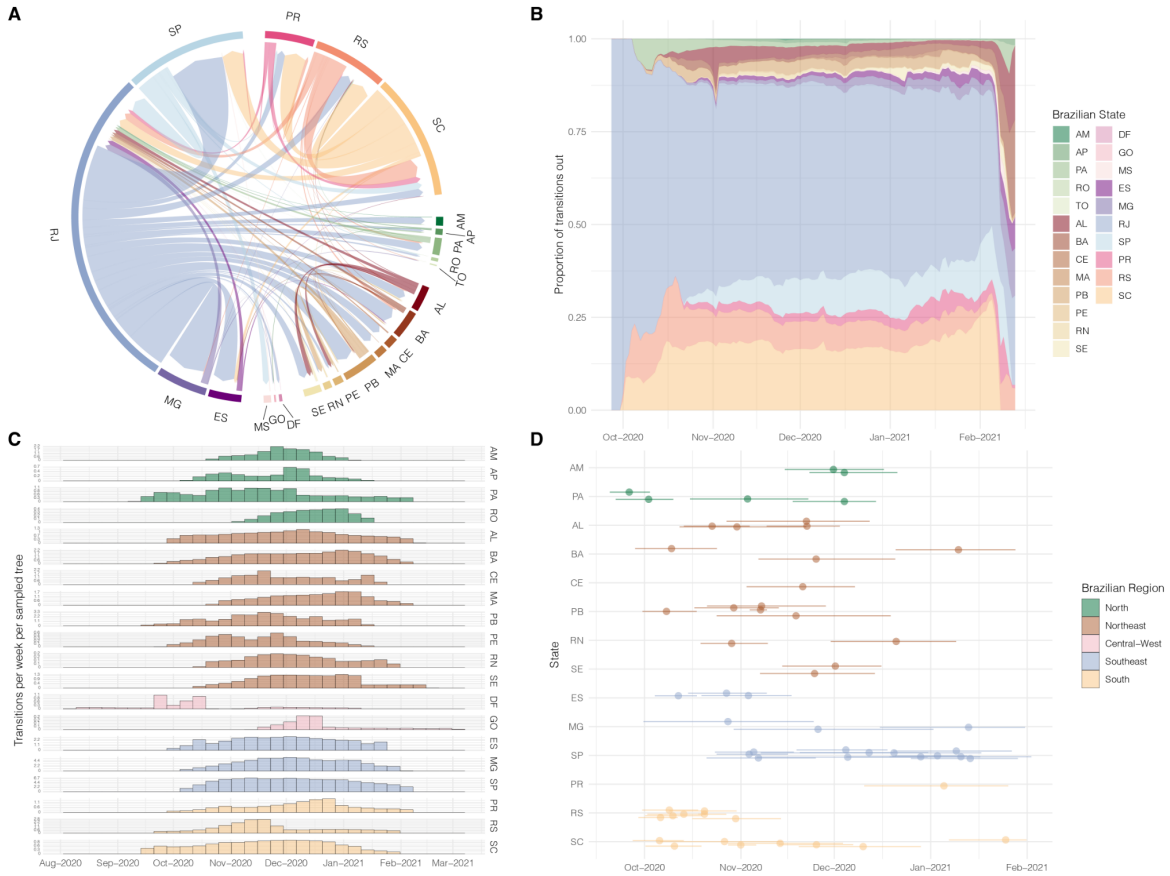

Figure 3. Spatio-temporal spread of SARS-CoV-2 lineage P.2 in Brazil. a) Inter-state circular migration flow plot as estimated from the Markov jumps analysis complementary to the phylogeographic model. Arrows indicate the direction of the migration and thickness is relative to the number of jumps. b) Contribution of each state in seeding P.2 to other locations through-time measured as the proportion of the total Markov per day. c) Time distribution (95% HPD) of jumps from RJ to other Brazilian states. d) T_{MRCA} of each highly supported (SH-aLRT > 80) state specific clades with more than three genomes.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

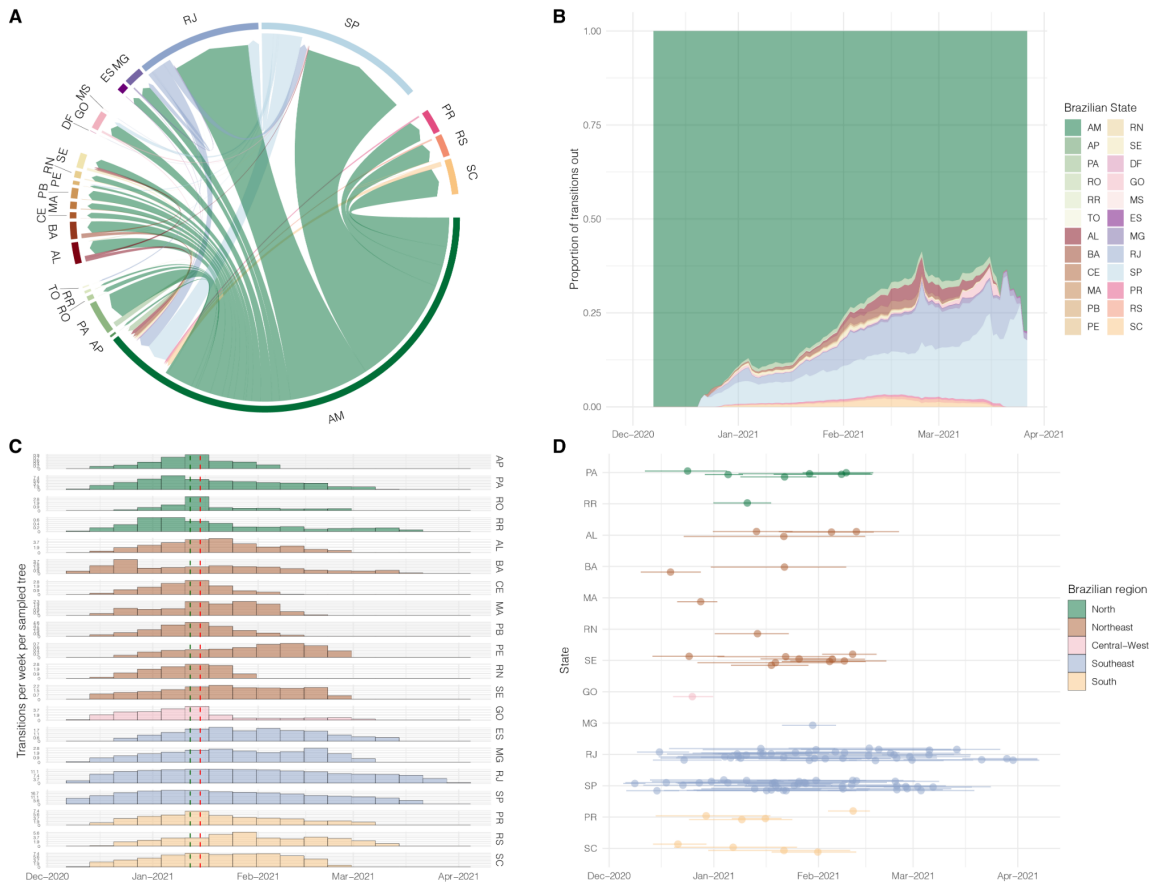

Figure 4. Spatio-temporal spread of SARS-CoV-2 VOC Gamma in Brazil. a) Inter-state circular migration flow plot as estimated from the Markov jumps analysis complementary to the phylogeographic model. Arrows indicate the direction of the migration and thickness is relative to the number of jumps. b) Contribution of each state in seeding Gamma to other locations through-time measured as the proportion of the total Markov jumps per day. c) Time distribution (95% HPD) of jumps from AM to other Brazilian states. The vertical green line shows the time of Gamma first notification and vertical red line shows the starting date of COVID-19 patients transferring from AM to other states. d) T_{MRCA} of each highly supported (SH-aLRT > 80) state specific clades with more than three genomes.

Estimating the Re of variants Gamma and P.2 in Brazil.

We applied the birth-death skyline (BDSKY) model to estimate the Re of P. 2 and Gamma at source locations by selecting all P.2 sequences from RJ ($P.2_{RJ}$) and all Gamma sequences from AM ($Gamma_{AM}$); and at secondary outbreaks by combining large (n > 10) highly supported (SH-aLRT>0.80) clades representative of the circulation in a single state. Only Gamma clusters from SP and RJ, and none for P.2, met the criteria for analysis of secondary outbreaks (see Methods). The temporal trajectory of the median Re of clade $P.2_{RJ}$ supports an initial expansion phase (Re > 1), followed by a phase of stabilization (Re ~ 1) and subsequent decrease (Re < 1) (**Fig 5a**). Interestingly, when clade $P.2_{RJ}$ reached the highest median Re in September 2020, the lineage P.2 was still undetected [4] and the number of SARI cases remained relatively constant in RJ (**Fig 5a**). When lineage P.2 became the dominant variant in RJ in December 2020, the clade $P.2_{RJ}$ displayed a median Re~1.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Similarly, to P.2 in RJ, in AM the clade $Gamma_{AM}$ showed an epidemic dynamic with three phases of expansion (Re > 1), stabilization (Re ~ 1) and subsequent decrease (Re < 1) (**Fig 5b**). When clade $Gamma_{AM}$ reached the highest median Re in December 2020, the relative prevalence of Gamma in Amazonas was low (<10%); but the number of SARI cases already started to increase (**Fig 5b**). When Gamma became the dominant variant in AM in January 2021, the clade $Gamma_{AM}$ also displayed a median Re~1. The initial expansion phase of major local Gamma clades in RJ ($Gamma_{RJ}$) and SP ($Gamma_{SP}$) from late December 2020 to late January 2021 occurred when Gamma was barely detected, and the number of SARI cases remained relatively constant in those states (**Fig 5a and 5c**). The clade $Gamma_{RJ}$ expanded in RJ during a period (December 2020 to March 2021) when the prevalence of clade $P.2_{RJ}$ was declining and the ratio $Re(Gamma_{RJ})/Re(P.2_{RJ})$ was 1.56 (December-January), 3.06 (February) and 2.33 (March) along the co-circulating period, supporting the higher transmissibility of the VOC Gamma concerning lineage P.2. Comparison of the median Re of VOC Gamma in different states revealed that clade $Gamma_{AM}$ displayed a much faster initial expansion (Re = 3.55) than clades $Gamma_{RJ}$ and $Gamma_{SP}$ (Re = 1.59 and 1.91, respectively). The growth rate of clade $Gamma_{AM}$, however, rapidly decreased in the second half of December 2020 (Re =1.40) and during early 2021 (Re =<1); while clades $Gamma_{RJ}$ and $Gamma_{SP}$ maintained longer phases of epidemic growth (Re = 1.45 - 2.01) until March 2021.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

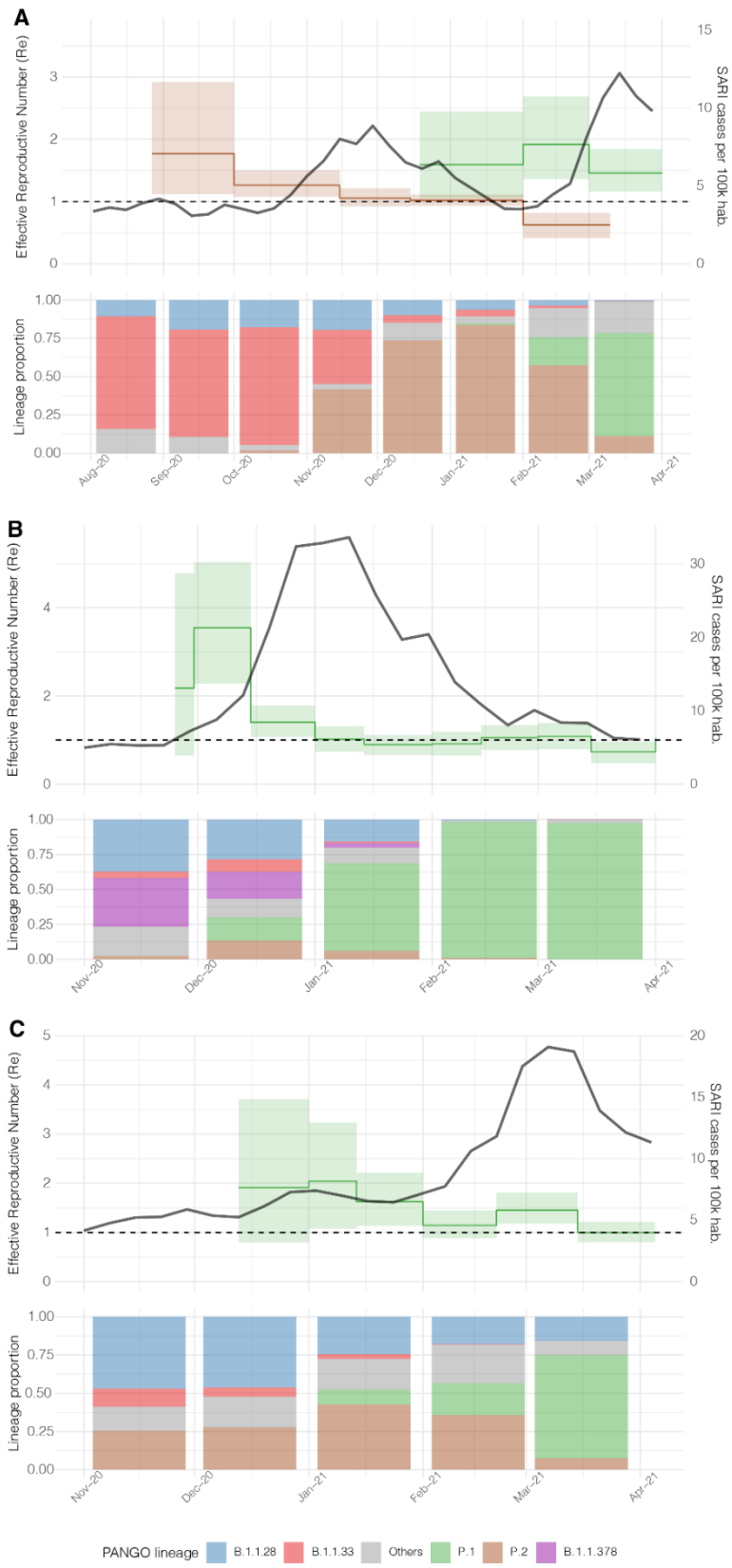

Figure 5. Phylodynamics analyses of variants P.2and Gamma in RJ (a), AM (b) and SP (c) states. Effective reproductive number (Re) [median (solid lines) and 95% HPD (high posterior density)], as estimated by Bayesian phylogenetics methods, is shown in green and brown for Gamma and P.2, respectively. Along with Re, SARI cases are shown as a black line. PANGO lineages proportions per month are shown in the bottom panel, colored according to the legend.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Discussion

We combine spatial and genomic data to reconstruct the spatiotemporal dynamics of the VOC Gamma and lineage P.2 in Brazil from September 2020 through April 2021, a study period that covers a decreasing phase of COVID-19 incidence from September to November 2020 and a subsequent increasing phase up to March-April 2021. Although both SARS-CoV-2 variants harbor the same mutation of concern S:E484K, they displayed different spatiotemporal dynamics probably shaped by the higher transmissibility of the VOC Gamma concerning lineage P.2.

Our study indicates that variants P.2 and Gamma probably arose in RJ and AM states around July and November 2020, respectively. This is consistent with initial estimations [4,7,15] and rejects the hypothesis of a much older origin of lineage P.2 in February 2020 and of VOC Gamma in August 2020. Our analyses suggest that some Gamma and P.2 Brazilian sequences with the earliest recorded sampling times available at the EpiCoV database in GISAID are probably the result of sample contamination or spelling errors in metadata entries. The inclusion of those problematic sequences, as well as of P.1 sequences belonging to Gamma-like clades [6], certainly pushed back the T_{MRCA} of Brazilian variants to unreliable old dates. The emergence of lineage P.2 in February 2020 is highly improbable because it overlaps with the estimated origin of the parental lineage B.1.1.28 [2], while the emergence of the VOC Gamma in August 2020 is inconsistent with the absence of this variant among hundreds of Amazonian SARS-CoV-2 positive samples genotyped between August and November 2020 [3].

Although the densely populated and well-connected states in the Southeastern region (RJ and SP) were the main sources of dissemination of lineage P.2 in Brazil, this variant spread at a slower rate than the VOC Gamma that was mainly disseminated from the poorly connected state of AM. According to our estimations, the earliest onset date of community transmission of lineage P.2 in most sampled states was traced to October 2020 (~3 months after its emergence), coinciding with the first detection of this variant in such locations[15], and it only became the most prevalent variant is some states around December 2020 (~5 months after its emergence). By comparison, the earliest timing of community transmission of the VOC Gamma in many Brazilian states was traced to December 2020 (~1 month after its emergence) and became the most prevalent variant in all Brazilian states by February 2021 (~3 months after its emergence). By the time when the VOC Gamma was first described on 12th January 2021 in Japanese travelers returning from the Amazonas state [16],

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

this variant had already established local transmission chains in at least nine out of 23 Brazilian states from all major country regions.

The rapid geographic spread of the VOC Gamma and the consistent fixation as the most prevalent SARS-CoV-2 lineage across all Brazilian regions during 2021 supports that this VOC is more transmissible than lineage P.2. The comparison of the Re of local clades of variants Gamma and P.2 that co-circulated in the RJ state from December 2020 to March 2021 allowed us to quantify the relative transmissibility of those lineages while controlling for confounding ecological factors such as local levels of immunity and social distancing measures. Our phylodynamics estimates indicate that the VOC Gamma displayed a median Re that was between 1.56 and 3.06 higher than that of lineage P.2 during the period of co-circulation. Notably, this result is consistent with a previous study conducted in the AM state that estimated that the median Re of the VOC Gamma was 2.2 times higher than that estimated for local B.1.1.28 clades that co-circulated during December 2020 [3]. These findings confirm the notion that the VOC Gamma was more transmissible than pre-existing SARS-CoV-2 lineages in Brazil, including the lineage P.2, in different local settings.

Even though our phylogeographic reconstructions support that RJ and SP were important hubs of inter-regional spread of SARS-CoV-2 during the epidemic wave in late 2020 and early 2021 as observed in early 2020 [2], they also pointed those other Brazilian states from the Northern (AM) and Southern (SC and RS) regions accounted for a substantial fraction of all inter-state disseminations of variants Gamma and P.2, respectively. Our findings also support that inter-state viral transmissions not only occurred between close geographic locations, but also between distant states of Northern, Southern and Northeastern regions. The diversification of geographic hubs of viral dispersion and the detection of long-distance viral migrations suggests that reactivation of domestic passenger air traffic since the second half of 2020 may have been an important driver of dissemination of variants P.2 and Gamma in Brazil. These findings also emphasize that implementation of more strict control measures for domestic air passengers, like the mandatory exigence of a negative SARS-CoV-2 RT-PCR test, could be an effective measure to reduce the dissemination rate of SARS-CoV-2 in Brazil [17].

Our findings revealed that variants P.2 and Gamma were disseminated for several weeks before being detected in a new location and their early phase of cryptic expansion is challenging to detect or predict by epidemiological surveillance systems. When the lineage P.2 reached the highest median Re in September 2020 in RJ, the mean daily number of SARS-CoV-2 cases in the state was decreasing and the incidence of SARI cases only started

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

to increase in mid-October 2020. Similarly, the VOC Gamma in RJ and SP probably started to expand in December 2020, well-before the first detection of this variant in January 2021 and the peak of SARI cases associated with Gamma in March 2021 in both states. We hypothesize that the unnoticed early expansion of variants P.2 or Gamma in many locations happened in an epidemic scenario where previous dominant lineages were declining, so the growth of daily SARI cases only became apparent when these new variants were already established as the most prevalent lineage. It may also be possible that initial circulation of variants P.2 and Gamma might have been driven by young people and SARI cases only increased when older individuals got infected.

Because the levels of immunity and social distancing varied over time among different Brazilian states, we compared the Re of the VOC Gamma in AM, RJ and SP between December 2020 and March 2021 to quantify the potential impact of ecological factors on Gamma spread. Although the Re estimates displayed large credibility intervals with considerable overlap, our data suggests that the VOC Gamma displayed a much faster initial expansion in AM (Re ~ 3.6) than in RJ or SP (Re ~ 1.5-2.0), but the growth phase in AM extends over a shorter time than in the Southeastern states. We speculate that the weaker mitigation measures implemented in AM with respect to RJ and SP at late 2020 and/or large-scale gatherings may have amplified the intrinsic higher transmissibility of the VOC Gamma in this Northern state [3]. The steep decrease of Re in AM in January 2021 might be the result of non-pharmaceutical interventions (NPIs) implemented after the health system collapse[3] and/or of high levels of population immunity estimated at ~70% in Manaus[18]. On the other hand, the larger number of susceptible individuals in the Southeastern region may have sustained a longer phase of Gamma expansion.

The most important limitation of our study was the uneven sampling among Brazilian states and throughout the time, which may have introduced some bias in phylogeographic analyses. Although the number of genomes analyzed in this study roughly follows the number of COVID-19 confirmed cases in Brazil (**Figure 1A and 1B**), the smaller sampling from September and December 2020 limited the potential of our analyzes to identify large state-specific P.2 clusters and to accurate estimates the earliest onset date of communitarian transmission of this variant in several locations. In addition, even though the datasets here analyzed comprise SARS-CoV-2 sequences from 24 out of 27 Brazilian states, the uneven distribution of these genomes among locations might have biased contributions to the overall inter-states' transmissions.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

In summary, this molecular epidemiological study showed that the COVID-19 epidemic in Brazil from September 2020 to March 2021 was characterized by the emergence and spread of the lineage P.2 and the VOC Gamma that were the most prevalent SARS-CoV-2 variants at different time periods in the country. The spatial dispersion of these variants in Brazil was driven by short and long-distance viral transmission and both variants circulated cryptically in several locations for some weeks before being detected. The VOC Gamma displayed a higher transmissibility than lineage P.2 which explained the faster rate of spatial spread of this variant and its establishment as the dominant SARS-CoV-2 lineage in Brazil within a few (~3) months after its emergence. Our findings also support that the transmissibility of the VOC Gamma varied according to the geographic context probably due to regional differences in social distancing measures and/or fraction of population previously infected. These findings also strengthen the need for ongoing genomic surveillance to provide early warnings about spread of emergent SARS-CoV-2 variants in Brazil.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Material and Methods

The genomic surveillance network and ethical aspects. The Brazilian Ministry of Health COVID-19 Genomic Surveillance Network is composed by Oswaldo Cruz Foundation (Fiocruz), Adolfo Lutz Institute (IAL), Evandro Chagas Institute (IEC), the Central Laboratories of each State of Brazil (LACENs) and other partners, such as Health Secretariat of Aparecida de Goiânia, Goiás. Saliva or nasopharyngeal swabs (NPS) samples were collected from suspect COVID-19 cases by sentinel hospitals or health care units and initially tested by real time RT-PCR as a routine diagnostic for SARS-CoV-2 by LACENs using any of the following different commercial assays: SARS-CoV2 (E/RP) (Biomanguinhos), Allplex 2019-nCoV Assay (Seegene) or an in-house protocol following the USA/CDC guidelines (https://www.fda.gov/media/134922/download). SARS-CoV-2 positive samples with cycling threshold (Ct) below 25 were randomly selected for genome sequencing. Alternatively, targeted cases, such as reinfection cases and fatal cases without comorbidities, were also selected for sequencing. Samples are then sent to one of the genomic reference laboratories participating in the network for whole genome sequencing. This study was approved by the Ethics Committee of the of the FIOCRUZ (CAAE: 68118417.6.0000.5248) and Amazonas State University (CAAE: 25430719.6.0000.5016), which waived the signed informed consent.

SARS-CoV-2 amplification and sequencing. The viral RNA was subjected to reverse transcription and PCR amplification using in-house protocols developed by COVID-19 Fiocruz Genomic Network [19,20] or the Illumina COVIDSeq Test (Illumina), including some primers to cover regions with dropout [14]. Normalized pooled amplicons of each sample were used to prepare NGS libraries with Nextera XT (FC-131-1096) and clustered with MiSeq Reagent Kit v2 (500-cycles - MS-102-2003) on 2 x 250 cycles (in-house protocols) or 2 x 150 cycles (MS-103-1002) paired-end runs. All sequencing data was collected using the Illumina MiSeq sequencing platforms and MiSeq Control software v2.6.2.1 (Illumina).

SARS-CoV-2 whole-genome consensus sequences and genotyping. FASTQ reads were generated by the Illumina pipeline at BaseSpace (https://basespace.illumina.com). All files were downloaded and imported into Geneious v10.2.6 for trimming and assembling using a customized workflow employing BBDuk and BBMap tools (v37.25) and the NC_045512.2 RefSeq as a template. Using a threshold of at least 50% to call a base. Consensus sequences were initially assigned to viral lineages according to the nomenclature proposed by Rambaut et al. [21], using the Pango Lineage software (https://pangolin.cog-uk.io) and later confirmed using phylogenetic analyses.

Dataset assembling. We complemented the P.1 and P.2 genomes dataset identified in our genomic surveillance by retrieving all Brazilian high quality (<5% N) whole-genomes (29 kb) identified as P.1 and P.2 from EpiCoV database in GISAID (https://www.gisaid.org) and that were deposited up to April 16th. Mutational composition of these sequences was analyzed in NextClade (https://clades.nextstrain.org) and due to highly variable patterns of substitutions, we have retained in the dataset only those genomes with the full-set of P.2 and Gamma lineages defining synapomorphies, thus excluding from the analysis all Gamma-like sequences (**Supplementary Table 2 and 3**).

Analysis of temporal signal. SARS-CoV-2 Gamma and P.2 complete genome sequences were aligned using MAFFT v7.467 [21] and subject to maximum likelihood (ML) phylogenetic analysis using IQ-TREE v2.1.2 [22] under the general time-reversible (GTR) model of nucleotide substitution with a gamma-distributed rate variation among sites, four rate categories (G4), a proportion of invariable sites (I) and empirical base frequencies (F)

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

nucleotide substitution model, as selected by the ModelFinder application [23]. The branch support was assessed by the approximate likelihood-ratio test based on the Shimodaira–Hasegawa-like procedure (SH-aLRT) with 1,000 replicates. The temporal signal of the P.1 and P.2 assembled datasets was assessed from the ML tree by performing a regression analysis of the root-to-tip divergence against sampling time using TempEst [24] and excluding outlier sequences that deviate more than 1.5 interquartile ranges from root-to-tip regression line, which included those Gamma and P.2 sequences with the oldest sampling dates. The expected collection dates of the oldest Gamma and P.2 sequences were next estimated by using the Bayesian leaf-dating method developed to estimate the age of sequences with unknown collection date [25]. For this analysis, the oldest Gamma and P.2 sequences were combined with sequences from Amazonas (Gamma epicenter) and RJ P.2 epicenter), respectively, and were analyzed with the BEAST v1.10 software [26] using a GTR+F+I+G4 nucleotide substitution model, an exponential growth coalescent model for the tree prior and a strict molecular clock with a uniform substitution rate prior (8 - 10 x 10^{-4} substitutions/site/year). To estimate the expected sampling dates of the oldest sequences we set a precision value of two years for the recorded date (year and month) of sampling and select the tip date option sampling uniformly from precision. Sequences for which the recorded collection date was not contained within the 95% High Posterior Density (HPD) interval of the leaf-age estimate were excluded. Markov Chain Monte Carlo (MCMC) was run sufficiently long to ensure convergence (effective sample size> 200) in all parameter estimates as assessed in TRACER v1.7 [27].

Estimating time-scaled phylogenetic trees. Due to the size of the datasets, we firstly estimated the time of the most recent common ancestor (T_{MRCA}) of each lineage by applying a full phylogenetic Bayesian analysis in downsampled datasets. To do so, 10 genomes per Brazilian state per week were randomly chosen using Augur [28], resulting in a dataset of 191 Gamma and 219 P.2 sequences. Alignments were generated using MAFFT v7.475 [21] and visually inspected in AliView [29]. Time-scaled phylogenetic trees were estimated in BEAST v.1.10 [26] under a GTR+F+I+G4 model. The evolutionary model was complemented with the non-parametric Bayesian skyline (BSKL) model as the coalescent tree prior [30] and a strict molecular clock model with a uniform substitution rate prior (8 - 10 x 10^{-4} substitutions/site/year). MCMC was run sufficiently long to ensure convergence (effective sample size> 200) in all parameter estimates as assessed in TRACER v1.7 [27]. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v1.10. ML and MCC trees were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The time-scaled tree for the full Gamma (N=2,477) and P.2 (N=1,357) datasets was performed using a modified version of BEAST (https://beast.community/thorney_beast) that makes feasible the analysis of big datasets. In summary the method alleviates most of the computational burden by fixing the tree topology and then re-scales the branch lengths based on a clock and coalescent models. To do so, ML phylogenetic trees constructed for both Gamma and P.2 datasets as explained above, were inputted in the BEAST xml file as starting and data trees and analyses were performed under a logistic coalescent prior, which outperformed [Bayes Factor (BF) > 3] the exponential prior in a Marginal Likelihood Estimation (MLE) of model fitness, and a strict molecular clock, as specified above. Additionally, we set a uniform prior on the root height estimates based on the 95% HPD determined in the BEAST analysis of the downsampled datasets. Four and three MCMC were run for 100 million generations and then combined to ensure stationarity and good mixing for the Gamma and P.2 datasets, respectively.

Discrete Bayesian phylogeography. A set of 1000 and 900 trees was randomly selected from the posterior distribution of trees resulting from the BEAST analysis of the full Gamma

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

and P.2 datasets, respectively. Sampling locations (Brazilian states) were used as traits in the phylogeographic model and reconstruction of the ancestral states was performed using a discrete symmetric model with BSSVS [31] on the posterior sampling of trees. SPREAD software [32] was used to identify the well-supported transition rates based (BF > 3). We complemented this analysis with Markov jump estimation of the number of location transitions throughout evolutionary history [33], which was used to explore directionality of the transitions. Viral migration history between all supported location transitions was then visualized in circular migration flow plots using the package "circlize" [34] available in R software (https://www.r-project.org). The temporal history of transitions from the lineage epicenter to other Brazilian states was summarized as histograms in one-week periods. Because many spatial transitions connected nodes with low support, we used a conservative approach to estimate the onset date of community transmission in each state from the T_{MRCA} of only highly supported (SH-aLRT > 80) clusters comprising at least three sequences from a given state and whose location root was most probably traced (*Posterior State Probability* [*PSP*] > 0.50) to that state. Additionally, we summarize P.2 and Gamma lineages spatial-temporal history alongside with SARI cases trajectories in Brazil, by generating an animation with baltic library (https://github.com/evogytis/baltic).

Effective Reproductive Number (Re) Estimation. To estimate the Re of lineages Gamma and P.2 through time in the source locations, we selected all Gamma sequences from Amazonas and all P.2 sequences from RJ whose state of the ancestral node in the phylogeographic trees remained in Amazonas or RJ, respectively. To estimate the Re of Gamma and P.2 dissemination outside the source locations, state specific clusters in the phylogenetic trees were selected based on the following criteria: a) high branch support (SH-aLRT > 80); b) cluster size bigger than 10 sequences; c) > 80% of sequences from a single state and origin of the cluster in the same state. Selected clusters were then combined and analyzed when the dataset comprised more than 50 sequences. Following these criteria, we could only analyze the phylodynamic of local dissemination of Gamma in the states of RJ and SP. The assembled datasets comprised 298 and 199 Gamma genomes for RJ and SP, respectively. For P.2 we could not analyze local dissemination outside the source region. The assembled datasets were then analyzed using the birth-death skyline (BDSKY) model [35] which reconstructs Re trajectories and is implemented within BEAST 2 v2.6.5 [36]. The sampling rate (d) was set to zero for the period before the oldest sample and estimated from the data afterward. The BDSKY prior settings were as follows: Become Uninfectious Rate (exponential, mean = 36); Reproductive Number (log normal, mean = 0.8, sd = 0.5); Sampling Proportion (beta, alpha = 1, beta = 100). Origin parameter was conditioned to the root height, and the Re was estimated in a piecewise manner over intervals defining to match state specific epidemiological trajectories. Molecular clock was as in the time-scaled trees analysis and the HKY+G4+F substitution model was used. MCMC chains were run until all relevant parameters reached ESS > 200, as explained above.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

REFERENCES

1 Resende, P. C., Delatorre, E., Gräf T., Mir D., Motta F.C., Appolinario L., Paixão A. C., Mendonça A. C.,, Ogrzewalska M., Caetano B., Wallau G. L., Docena C., Santos M. C., Ferreirra J., Sousa Junior E., Silva S., Fernandes S., Vianna L. A., Souza L., Ferro J. F, Nardy V., Santos C., Riediger I., Debur M., Croda J., Oliveira, W, Abreu A,, Bello G.. Siqueira M. M. Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B.1.1.33 during the early pandemic phase in Brazil. *Frontiers in Microbiology*, doi:10.3389/fmicb.2020.615280 (2020).

2 Candido, D. S. *et al*. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* **369**, 1255-1260, doi:10.1126/science.abd2161 (2020).

3 Naveca, F. G. *et al*. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nat Med* **27**, 1230–1238, doi:10.1038/s41591-021-01378-7 (2021).

4 Voloch, C. M. *et al*. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol*, doi:10.1128/JVI.00119-21 (2021).

5 FIOCRUZ. *Dashboard Rede Genômica Fiocruz*, <http://www.genomahcov.fiocruz.br/dashboard/> (2021).

6 Lamarca, A. P. *et al*. 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**, e0009835, doi:10.1371/journal.pntd.0009835 (2021).

7 Faria, N. R. *et al*. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815-821, doi:10.1126/science.abh2644 (2021).

8 Martins, A. F. *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. *Eurosurveillance* **26**, 2100276, doi:10.2807/1560-7917.ES.2021.26.12.2100276 (2021).

9 Silva, M. S. D. *et al*. Early detection of SARS-CoV-2 P.1 variant in Southern Brazil and reinfection of the same patient by P.2. *Rev Inst Med Trop Sao Paulo* **63**, e58, doi:10.1590/S1678-9946202163058 (2021).

10 Tosta, S. *et al*. Short Report: Early genomic detection of SARS-CoV-2 P.1 variant in Northeast Brazil. *PLoS Negl Trop Dis* **15**, e0009591, doi:10.1371/journal.pntd.0009591 (2021).

11 Dos Santos, C. A. *et al*. SARS-CoV-2 Genomic Surveillance in Northeast Brazil: Timing of Emergence of the Brazilian Variant of Concern P1. *J Travel Med*, doi:10.1093/jtm/taab066 (2021).

12 Nicolelis, M. A. L., Raimundo, R. L. G., Peixoto, P. S. & Andreazzi, C. S. The impact of super-spreader cities, highways, and intensive care availability in the early stages of the COVID-19 epidemic in Brazil. *Sci Rep* **11**, 13001, doi:10.1038/s41598-021-92263-3 (2021).

13 Coutinho, R. M. *et al*. Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. *medRxiv*, doi:10.1101/2021.03.03.21252706 (2021).

14 Naveca, F. *et al*. Spread of Gamma (P.1) sub-lineages carrying Spike mutations close to the furin cleavage site and deletions in the N-terminal domain drives ongoing transmission of SARS-CoV-2 in Amazonas, Brazil. *medRxiv*, doi:10.1101/2021.09.12.21263453 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

15 Resende, P. C. *et al*. Severe Acute Respiratory Syndrome Coronavirus 2 P.2 Lineage Associated with Reinfection Case, Brazil, June-October 2020. *Emerg Infect Dis* **27**, 1789-1794, doi:10.3201/eid2707.210401 (2021).

16 Fujino, T. *et al*. Novel SARS-CoV-2 Variant Identified in Travelers from Brazil to Japan. *Emerg Infect Dis* **27**, 1243-1245, doi:10.3201/eid2704.210138 (2021).

17 Murano, Y. *et al*. Impact of domestic travel restrictions on transmission of COVID-19 infection using public transportation network approach. *Sci Rep* **11**, 3109, doi:10.1038/s41598-021-81806-3 (2021).

18 Sabino, E. C. *et al*. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452-455, doi:10.1016/S0140-6736(21)00183-5 (2021).

19 Nascimento, V. A. D. *et al*. Genomic and phylogenetic characterisation of an imported case of SARS-CoV-2 in Amazonas State, Brazil. *Mem Inst Oswaldo Cruz* **115**, e200310, doi:10.1590/0074-02760200310 (2020).

20 Resende, P. *et al*. SARS-CoV-2 genomes recovered by long amplicon tiling multiplex approach using nanopore sequencing and applicable to other sequencing platforms. *bioRxiv* (2020). <https://doi.org/10.1101/2020.04.30.069039>.

21 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

22 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).

23 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* **14**, 587-589, doi:10.1038/nmeth.4285 (2017).

24 Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**, vew007, doi:10.1093/ve/vew007 (2016).

25 Shapiro, B. *et al*. A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol* **28**, 879-887, doi:10.1093/molbev/msq262 (2011).

26 Suchard, M. A. *et al*. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016, doi:10.1093/ve/vey016 (2018).

27 Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* **67**, 901-904, doi:10.1093/sysbio/syy032 (2018).

28 Huddleston, J. *et al*. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw* **6**, doi:10.21105/joss.02906 (2021).

29 Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276-3278, doi:10.1093/bioinformatics/btu531 (2014).

30 Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **22**, 1185-1192, doi:10.1093/molbev/msi103 (2005).

31 Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput Biol* **5**, e1000520, doi:10.1371/journal.pcbi.1000520 (2009).

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

32 Bielejec, F., Rambaut, A., Suchard, M. A. & Lemey, P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **27**, 2910-2912, doi:10.1093/bioinformatics/btr481 (2011).

33 O'Brien, J. D., Minin, V. N. & Suchard, M. A. Learning to count: robust estimates for labeled distances between molecular sequences. *Mol Biol Evol* **26**, 801-814, doi:10.1093/molbev/msp003 (2009).

34 Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812, doi:10.1093/bioinformatics/btu393 (2014).

35 Stadler, T., Kuhnert, D., Bonhoeffer, S. & Drummond, A. J. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A* **110**, 228-233, doi:10.1073/pnas.1207965110 (2013).

36 Bouckaert, R. *et al*. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15**, e1006650, doi:10.1371/journal.pcbi.1006650 (2019).

Acknowledgements

The authors wish to thank all the health care workers and scientists who have worked hard to deal with this pandemic threat, the GISAID team, and all the EpiCoV database's submitters. GISAID acknowledgment table containing sequences used in this study is shown in **Supplementary Table 4**. We also appreciate the support of the Fiocruz COVID-19 Genomic Surveillance Network (http://www.genomahcov.fiocruz.br/) members, the Respiratory Viruses Genomic Surveillance Network of the General Laboratory Coordination (CGLab), Brazilian Ministry of Health (MoH), Brazilian Central Laboratory States (LACENs), and the Amazonas surveillance teams for the partnership in the viral surveillance in Brazil.

Funding

Financial support was provided by Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) (PCTI-EmergeSaude/AM call 005/2020 and Rede Genômica de Vigilância em Saúde-REGESAM); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grant 402457/2020–0); CNPq/Ministério da Ciência, Tecnologia, Inovações e Comunicações/Ministério da Saúde (MS/FNDCT/SCTIE/Decit) (grant 403276/2020-9); Departamento da Ciência e Tecnologia (DECIT), Ministério da Saúde; Inova Fiocruz/Fundação Oswaldo Cruz (Grants VPPCB-007-FIO-18–2–30 and VPPCB-005-FIO-20–2–87), INCT-FCx (465259/2014–6) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (26/210.196/2020). F.G.N, G.L.W, G.B and M.M.S are supported by the CNPq through their productivity research fellowships (306146/2017–7, 303902/2019–1, 302317/2017–1 and 313403/2018-0, respectively). G.B. is also funded by FAPERJ (Grant number E-26/202.896/2018).

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Appendix 1

Authors members of Brazilian Ministry of Health COVID-19 Genomic Surveillance Network

Tirza Peixoto Mattos[1], Valdinete Alves Nascimento[2], Victor Souza[2], André de Lima Guerra Corado[2], Fernanda Nascimento[2], George Silva[2], Matilde Mejía[2], Maria Júlia Brandão[2], Ágatha Costa[2], Karina Pessoa[2], Michele Jesus[2], Luciana Fé Gonçalves[2], Cristiano Fernandes[3], Valnete Andrade[4], Luana Barbagelata[5], Ana Cecília Ribeiro Cruz[5], Andrea Costa[6], Lindomar dos Anjos Silva[6], Jucimária Dantas Galvão[7], Anderson Brandao Leite[8], Felicidade Mota Pereira[9], Thais Oliveira Costa[10], Joaquim Cesar Sousa Jr[10], Lidio Gonçalves Lima Neto[11], Haline Barroso[12], Dalane Loudal Florentino Teixeira[12], Joao Felipe Bezerra[13], Cássia Docena[14], Raul Emídio de Lima[14], Lilian Caroliny Amorim Silva[14], Gustavo Barbosa de Lima[14], Laís Ceschini Machado[14], Maria Eduarda Pessoa Lopes Dantas[15], Raíssa Liane Do Nascimento Pereira[15], Josélio Araújo[15], Cliomar A Santos[16], Rodrigo Ribeiro Rodrigues[17], André Felipe Leal Bernardes[18], Felipe Campos de Melo Iani[18], Beatriz Grinsztejn[19], Valdiléa G Veloso[19], Patricia Brasil[19], Anna Carolina Dias da Paixão[20], Luciana Reis Appolinario[20], Renata Serrano Lopes[20], Fernando do Couto Motta[20], Alice Sampaio Rocha[20], Taina Moreira Martins Venas[20], Elisa Cavalcante Pereira[20], Andrea Cony Cavalcanti[21], Leonardo Soares Bastos[22], Luis Fernando de Macedo Brigido[23], Mauro de Medeiros Oliveira[24], Michelle Orane Schemberger[24], Andreia Akemi Suzukawa[24], Irina Riediger[25], Maria do Carmo Debur[25], Richard Steiner Salvato[26], Tatiana Schäffer Gregianini[26], Darcita Buerger Rovaris[27], Sandra Bianchini Fernandes[27]

Affiliations

[1]Laboratorio Central de Saude Publica do Amazonas (LACEN-AM). Manaus, Amazonas, Brazil.; [2]Laboratório de Ecologia de Doenças Transmissíveis na Amazônia (EDTA), Leônidas e Maria Deane Institute, Fiocruz; [4]Fundação de Vigilância em Saúde do Amazonas - Dra. Rosemary Costa Pinto, Manaus, AM, Brazil; [5]Laboratório Central de Saúde Pública do Pará (LACEN-PA), Belém, Pará; ; [6]Instituto Evandro Chagas, Belem, Para; [7]Laboratorio Central de Saude Publica do Amapá (LACEN-AP). , Amapá, Brazil; [8]Laboratorio Central de Saude Publica do Amapá (LACEN-TO). , Tocantins, Brazil; [9]Laboratorio Central de Saude Publica de Alagoas (LACEN-AL). Maceio, Alagoas,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.24.21265116; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC 4.0 International license .

Brazil; [10]Laboratório Central de Saúde Pública Prof. Gonzalo Moniz (LACEN-BA); [11]Fundação Oswaldo Cruz - Fiocruz Ceará; [12]Laboratório Central de Saúde Pública do Maranhao (LACEN-MA) ; [13]Laboratório Central de Saúde Pública da Paraiba (LACEN-PB) ; [14]Universidade Federal da Paraíba (UFPB) ; [15]Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, Pernambuco, Brazil; [16]Universidade Federal do Rio Grande do Norte (UFRN) ; [16]Laboratorio Central de Saude Publica de Sergipe (LACEN-SE). Aracaju, Sergipe, Brazil; [17]Laboratório Central de Saúde Pública do Espirito Santo (LACEN-ES); [18]Laboratório Central de Saúde Pública de Minas Gerais (LACEN-MG) ; [19]Instituto Nacional de Infectologia Evandro Chagas (INI), Fiocruz, Rio de Janeiro, Brazil; [20]Laboratory of Respiratory Viruses and Measles, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, RJ, Brazil. ; [21]Laboratório Central de Saúde Pública do Rio de Janeiro (LACEN-RJ) ; [22]Grupo de Métodos Analíticos em Vigilância Epidemiológica, Programa de Computação Científica (PROCC), Fiocruz, Rio de Janeiro, Brazil; [23]Instituto Adolfo Lutz, São Paulo; [24]Instituto Carlos Chagas (ICC), Fiocruz-PR; [25]Instituto Carlos Chagas (ICC), Fiocruz-PR; [26]Laboratorio Central de Saude Publica de Parana (LACEN-PR). Curitiba, Parana, Brazil; [27]Laboratório Central de Saúde Pública do Rio Grande do Sul (LACEN-RS), Porto Alegre, Rio Grande do Sul, Brazil; [28]Laboratorio Central de Saude Publica do Estado de Santa Catarina (LACEN-SC), Florianopolis, Santa Catarina, Brazil.