

# 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

Felipe Naveca (✉ [felipe.naveca@fiocruz.br](mailto:felipe.naveca@fiocruz.br))

Fiocruz/ILMD <https://orcid.org/0000-0002-2888-1060>

**Valdinete Nascimento**

Fiocruz/ILMD

**Victor Souza**

Fiocruz/ILMD

**André Corado**

Fiocruz/ILMD

**Fernanda Nascimento**

Fiocruz/ILMD

**George Silva**

Fiocruz/ILMD

**Ágatha Costa**

Fiocruz/ILMD

**Débora Duarte**

Fiocruz/ILMD

**Karina Pessoa**

Fiocruz/ILMD

**Matilde Mejía**

Fiocruz/ILMD

**Maria Brandão**

Fiocruz/ILMD

**Michele Jesus**

Fiocruz/ILMD

**Luciana Gonçalves**

FVS-AM

**Cristiano da Costa**

FVS-AM

**Vanderson Sampaio**

Fundação de Medicina Tropical Dr Heitor Vieira Dourado <https://orcid.org/0000-0001-7307-8851>

**Daniel Barros**

FVS-AM

**Marineide Silva**

LACEN-AM

**Tirza Mattos**

LACEN-AM

**Gemilson Pontes**

Instituto Nacional de Pesquisas da Amazônia

**Ligia Abdalla**

UEA

**João Santos**

Hospital Adventista de Manaus

**Ighor Arantes**

Fiocruz/IOC

**Filipe Dezordi**

Fiocruz/IAM

**Marilda Siqueira**

Fiocruz/IOC

**Gabriel Wallau**

Fiocruz/IAM

**Paola Resende**

Fiocruz/IOC

**Edson Delatorre**

UFES

**Tiago Gräff**

Fiocruz/IGM

**Gonzalo Bello**

Fiocruz/IOC

---

## Article

### Keywords:

DOI: <https://doi.org/10.21203/rs.3.rs-275494/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The Northern Brazilian state of Amazonas is one of the most heavily affected country regions by the COVID-19 epidemic and experienced two exponential growing waves in early and late 2020. Through a genomic epidemiology study based on 250 SARS-CoV-2 genomes from different Amazonas municipalities sampled between March 2020 and January 2021 we revealed that the first exponential growth phase was driven mostly by the dissemination of lineage B.1.195 which was gradually replaced by lineage B.1.1.28. The second wave coincides with the emergence of the variant of concern (VOC) P.1 which evolved from a local B.1.1.28 clade in late November and rapidly replaced the parental lineage in less than two months. Our findings support that successive lineage replacements in Amazonas were driven by a complex combination of variable levels of social distancing measures and the emergence of a more transmissible VOC P.1 virus. These data provide unique insights to understanding the mechanisms that underlie the COVID-19 epidemic waves and the risk of disseminating SARS-CoV-2 VOC P.1 in Brazil and potentially worldwide.

## Text

The Amazonas state reported the first sequence confirmed SARS-CoV-2 case in Manaus, the state capital city, in March 2020 in a traveler returning from Europe <sup>1</sup>. By late February 2021, more than 306,000 laboratory-confirmed cases and more than 10,400 deaths had been reported <sup>2</sup>. The COVID-19 epidemic in the Amazonas was characterized by two exponentially growing curves of cases (Fig. 1a). Epidemiological data from syndromic surveillance of severe acute respiratory illness (SARI) and burials indicates the first wave of the epidemic started in March 2020 and peaked around early May 2020, when the number of cases dropped and then remained roughly stable from June to November 2020. However, in mid-December, the number of cases started to grow exponentially, establishing the second wave of the epidemic.

Recent findings bring attention to the emergence of a new SARS-CoV-2 variant of concern (VOC) in Manaus, designated P.1, also known as N501Y.V3. Lineage P.1 was first detected in four travelers returning from the Amazonas state on 2nd January 2021 <sup>3</sup> and soon later recognized as an emergent lineage in Manaus <sup>4</sup>. The VOC P.1 harbors 21 lineage-defining mutations, including 10 synapomorphic mutations in the Spike protein (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, T1027I). The emergence of P.1 was alleged as one of the putative causes of the second wave of COVID-19 in Manaus <sup>5</sup>. However, the precise relationship between circulating SARS-CoV-2 variants and epidemic dynamics in the Amazonas remains unclear due to the paucity of viral sequences sampled in this Brazilian state before December 2020.

## Results

**Evidence of successive SARS-CoV-2 lineage replacements in Amazonas.** To have a more in-depth understanding of the genetic diversity of the SARS-CoV-2 variants circulating in the Amazonas state since

the early epidemic phase, we generated 250 SARS-CoV-2 high-quality whole-genome sequences from individuals living in 25 municipalities of the Amazonas state between 16th March 2020 and 13th January 2021 (Figs. 1a and 1b). Viral sequences were generated at FIOCRUZ Amazônia, which is part of both the Amazonas state health genomics network and the official consortium “FIOCRUZ Genomics Network” of the Brazilian Ministry of Health, for diagnostics and genomic surveillance of the SARS-CoV-2 in Brazil (<http://www.genomahcov.fiocruz.br/>). Our genomic survey revealed that most sequences were classified into five lineages: B.1.1.28 (n = 93; 37%), P.1 (n = 60; 24%), B.1.195 (n = 49; 20%), B.1.1.33 (n = 29; 12%) and P.2 (n = 9; 4%) (Supplementary Table 1), with a changing temporal prevalence over time (Fig. 1c). The lineage B.1.195 was the most prevalent variant during the first exponential growth phase. However, its prevalence gradually decreased after the first epidemic peak in early May and was surpassed by lineage B.1.1.28. This lineage persisted as the most prevalent one from May to December 2020, when the second lineage replacement took place, coinciding with the second phase of exponential growth. The VOC P.1 was first detected on 4th December 2020 in Manaus and displayed an extremely rapid increase in prevalence up to January 2021.

To better estimate the temporal trajectory of the P.1 emergence in the Amazonas state in late 2020 and early 2021, we designed a real-time PCR assay to detect the deletion at orf1b (NSP6: S106del, G107del, F108del), which is a genetic signature of the VOCs (P.1, B.1.1.7 and B.1.351). Thus, we evaluate all SARS-CoV-2 positive samples available between 1st November 2020 and 31st January 2021 that were not sequenced. None of the SARS-CoV-2 positive samples genotyped by real-time PCR before 16th December was positive for the NSP6 deletion, supporting our sequencing results that indicate a very low prevalence of VOC P.1 before mid-December 2020 in Amazonas. However, positive samples for the NSP6 deletion were quite frequent in the second half of December 2020 and January 2021. The combined results of genomic sequencing and real-time PCR testing support the sharp increase of lineage P.1 prevalence in the Amazonas state, from 0% in November 2020 (n = 0/88), to 4% in 1st-16th December 2020 (n = 2/54), 45% in 17th-31st December 2020 (n = 104/232), and 73% in 1st-15th January 2021 (n = 119/162) (Fig. 1d).

**Major SARS-CoV-2 local clades drove the COVID-19 epidemic in the Amazonas state.** Time-scale maximum likelihood (ML) phylogeographic analyses support at least 36 importation events of the widespread Brazilian lineages B.1.1.28 (n = 20), P.2 (n = 8), and B.1.1.33 (n = 8) into the Amazonas state; mostly from the Southeastern (69%) and Southern (14%) Brazilian regions (Figs. 2a and 2b). Despite multiple seeding events, most B.1.1.28 and B.1.1.33 viruses from the Amazonas state branched in three highly supported (approximate likelihood-ratio test [aLRT] > 80%) local clades (Figs. 2a and 2b). The clades 28-AM-I and 28-AM-II comprised 49% (n = 46) and 26% (n = 24) of all B.1.1.28 sequences from Amazonas, respectively, and clade 28-AM-II further gave origin to lineage P.1. The clade 33-AM-I comprises 48% (n = 14) of all B.1.1.33 sequences from Amazonas here detected. Bayesian phylogeographic analysis of the B.1.195 sequences sampled worldwide, by contrast, support a single importation event of this lineage in the Amazonas state either from other Brazilian states (posterior probability [PSP] = 0.61) or from abroad (PSP = 0.39) (Fig. 2c). The clustering of B.1.195 Amazonian sequences in a single highly supported (PP = 0.92) monophyletic clade (195-AM) was resolved after incorporating a nine nucleotide deletion at Nsp1 (delta 640–648: K141, S142, F143) as an informative

trait in phylogenetic reconstruction using a Bayesian framework. Such deletion was the only molecular signature of clade 195-AM and was also detected in three sequences sampled outside Brazil that were nested within clade 195-AM: two from the Colombian Amazonian region and one from a Japanese traveler returning from Amazonas.

Analysis of the identified Amazonian clades revealed single lineage-defining mutations in clades 28-AM-I (C29284T), 28-AM-II (A6613G), and 33-AM-I (A28108C, ORF8:Q72P), relative to other B.1.1.28 and B.1.1.33 Brazilian sequences. Furthermore, only one signature mutation at the Spike (V1176F) distinguishes Amazonian lineages B.1.1.195 and B.1.1.28, which contrasts with the high number of mutations accumulated by the lineage P.1. A closer inspection of the genetic diversity within the Amazonian clade 28-AM-II further revealed a P.1-like sequence sampled in Manaus on 23rd December 2020 that branched basal to the P.1 lineage and also accumulated an unusually high number of genetic changes concerning other B.1.1.28 Brazilian sequences (Fig. 3a). The P.1-like sequence harbors 6/10 P.1 lineage-defining mutations in the Spike protein, including those at the RBD domain (K417T, E484K, and N501Y) (**Supplementary Table 1**). To investigate the events underlying P.1 and P.1-like lineages' emergence, we analyzed the intrahost SARS-CoV-2 variability in our samples. We found a total of 484 well-supported minor variants (MVs) with a predominance of non-synonymous ( $n = 368$ ) over synonymous ( $n = 105$ ) substitutions (Fig. 3b). However, no MVs corresponding to the Spike lineage-defining mutations of clades P.1 and P.1-like were identified (**Supplementary Table 2**). Furthermore, the frequency of MVs observed in samples taken during the early (March-September) and late (October-January) epidemic phases were comparable (**Extended Data Fig. 1**).

**Differences in the epidemic trajectory of major SARS-CoV-2 Amazonian clades.** Reconstruction of the spatiotemporal dissemination dynamic using a Bayesian phylogeographic approach supports that the early prevalent local clade 195-AM probably emerged in mid-March 2020 in the city of Manaus (**Supplementary Table 3**). Consequently, this lineage quickly spread from Manaus to several municipalities of the Manaus metropolitan region and inner Amazonas state (Figs. 4a and 4b). The origin of clades 28-AM-I and 28-AM-II was estimated to be in late March 2020 (**Supplementary Table 3**) in the area comprising Manacapuru, Iranduba, and Manaquiri municipalities, located on the Solimoes river, and from this area these clades were disseminated to Manaus and other inner municipalities (Figs. 4c to 4f). The clade 33-AM-I probably arose in late April (**Supplementary Table 3**) and remained restricted to the city of Manaus. Finally, our analyses indicate the VOC P.1 probably arose in Manaus around late November 2020 (**Supplementary Table 3**) and rapidly spread to other municipalities of the metropolitan region and also to municipalities located up to 1,100 km distant from Manaus, at the border with Peru, Colombia and Venezuela (Figs. 4e and 4f). These analyses further traced the most recent common ancestor of lineages P.1 and P.1-like to the Manaus city in late August (**Supplementary Table 3**).

We next applied the birth-death skyline (BDSKY) model to estimate the effective reproductive number ( $R_e$ ) of the Amazonian clades with more than 40 genomes. The estimated  $R_e$  trajectories matched the relative prevalence of lineages and social distancing metrics (Fig. 5) very closely. The  $R_e$  of clade 195-AM was high (2.6, 95% HPD: 1.6–3.8) in March, but displayed a steep decrease to 1.0 (95% HPD: 0.8–1.2) in April,

coinciding with an increase of social distancing above 50% in Manaus. Clade 28-AM-I, which was estimated to have emerged in Amazonas countryside municipalities, also presented a high  $R_e$  (2.1, 95% HPD: 1.2–3.4) in its initial spread, reducing to 0.9 (95% HPD: 0.7–1.2) in May, when the social distancing index reached 50% in the interior of Amazonas state. The increasing relative prevalence of clade 28-AM-I over clade 195-AM from April to June agrees with the estimated  $R_e$  differences during April. From June to August 2020, the  $R_e$  of clades 195-AM and 28-AM-I remained roughly stable around 1.0, as well as their relative prevalence. When the social distancing index decreased to below 40% in September 2020, the clade 195-AM became apparently extinct while the  $R_e$  of the clade 28-AM-I increased to 1.2 (95% HPD: 0.9–1.6) and then remained roughly stable above 1.0 up to the end of 2020, leading to an increasing prevalence of clade 28-AM-I between September and November 2020. The lineage P.1 arose in late November and displayed a high  $R_e$  of 2.6 (95% HPD: 1.5–4.5) during December 2020, becoming the predominant lineage. With the consequently increased social distance after the health system collapse, this VOC's  $R_e$  was estimated to decrease to 1.2 (95% HPD: 0.9–1.6) in late December and January.

**SARS-CoV-2 P.1 infections are associated with higher viral loads.** The estimated  $R_e$  trajectories of SARS-CoV-2 Amazonian lineages support that VOC P.1 could be more transmissible than earlier prevalent viral lineages circulating in Amazonas. To test whether this estimated  $R_e$  difference might reflect a distinct virological phenotype, we used the real-time RT-PCR cycle threshold (Ct) scores as a proxy of the viral load in the upper respiratory tract (URT) of P.1-positive and P.1-negative samples collected at similar time from the onset of symptoms (Fig. 6 and Supplementary Table 4). Our analysis revealed a significant (Mann-Whitney test,  $P < 0.0001$ ) lower median score in URT samples from P.1-infected (Ct = 19.8) compared with non-P.1-infected (Ct = 23.0) subjects, indicating that viral load was ~ 10-fold higher in P.1 infections than in non-P.1 infections. The Ct scores of P.1 infections were significantly lower than of non-P.1 infections in adult (18–59 years old) men ( $P = 0.0005$ ), adult women ( $P < 0.0001$ ), and old (> 59 years old) women ( $P = 0.0149$ ); but not significantly different in old men ( $P = 0.4624$ ). As expected, non-P.1 infected old men displayed higher viral load (Ct = 20.8) than adults male (Ct = 22.7) and female of all ages (Ct = 23.8). Old P.1-infected male, by contrast, displayed viral load (Ct = 19.8) similar to adult male (Ct = 19.4) and female of all ages (Ct = 20.1), suggesting that P.1-infected individuals above 18 years old could be equally infectious, irrespective of sex and age.

## Discussion

The present study is the most comprehensive SARS-CoV-2 genomic investigation performed to date in Amazonas, one of the most heavily hit Brazilian states by the COVID-19 pandemic. Our genomic analyses revealed that most Amazonian cases were driven by the successful dissemination of a few local viral clades that together comprise 77% of the 250 SARS-CoV-2 Amazonian genomes here sampled between March 2020 and January 2021. Early major SARS-CoV-2 Amazonian clades arose in Manaus or at the metropolitan region between mid-March and late April 2020 and were widely disseminated within the Amazonas state, reaching the most isolated inner localities. By contrast, we found almost no evidence of disseminating early local SARS-CoV-2 Amazonian lineages outside the state, supporting that Amazonas has not been a major hub of viral dissemination within Brazil during 2020. The low land

accessibility of major Amazonian cities from other Brazilian states, combined with the considerable reduction in tourism activities and air traffic during 2020, might have significantly reduced the chance of exportation of Amazonian SARS-CoV-2 variants to other Brazilian regions. However, travels during Christmas and New Year celebrations, combined with the emergence of a potentially more transmissible VOC P.1, might have changed this scenario.

Two SARS-CoV-2 lineage replacements characterized the COVID-19 epidemic in the Amazonas state during early and late 2020. The first lineage replacement started after the first epidemic peak and was a gradual process of nearly five months during which lineage B.1.1.28 progressively substituted lineage B.1.195 that was the dominant variant at the first epidemic wave. Our analyses suggest that epidemiological rather than virological factors probably drove this first lineage replacement. The SARS-CoV-2 Amazonian clades 28-AM-I and 28-AM-II, which became the dominant variants in the phase between peaks, displayed only single lineage-defining synonymous mutations, with few synapomorphic differences at the Spike protein concerning clade 195-AM and evolved at a relatively constant rate between April and November 2020. The most notable difference was that clade 195-AM arose in the city of Manaus, and its  $R_e$  was significantly reduced around mid-April when social distancing in Manaus increased to > 50%. Clades 28-AM-I and 28-AM-II, by contrast, arose outside the city of Manaus and the  $R_e$  of clade 28-AM-I remained above 1.0 until mid-May 2020, when social distancing outside the capital city increased to > 50%. When mitigation measures were relaxed, and the social distancing index fell below 40% in September 2020, the  $R_e$  of clade 28-AM-I returned to above 1.0, while the clade B.1.195 became extinct, completing the lineage replacement process. Thus, the lower social distancing observed in the Amazonas state's interior municipalities compared to Manaus was the probable driver of the first lineage replacement.

A study of blood donors conducted in Manaus estimated that the first wave of SARS-CoV-2 infected 76% (95% CI 67–98) of the city's population by October 2020, suggesting that herd immunity had already been reached<sup>6</sup>. Assuming that reinfection is rare and that immunity against infection did not significantly wane by December 2020, a second COVID-19 wave would not be expected so early. Several hypotheses were proposed to explain this unexpected second wave that resulted in the collapse of the health system in Manaus in December 2020 and January 2021<sup>5</sup>. Our findings support that non-pharmaceutical interventions (NPI) could explain a large portion of the epidemic dynamics in Amazonas. A drastic reduction in the median  $R_e$  (from 2.1–2.6 to 0.9-1.0) for Amazonian lineages B.1.195 and B.1.1.28 around April-May 2020, was reconstructed in our analysis. This is entirely consistent with the epidemic trajectories estimated from epidemiological modeling<sup>7,8</sup> and further coincides with the timing of implementation of NPIs that effectively increased social distancing in Amazonas. This evidence indicates that the first epidemic wave in Amazonas was brought under relative control by the implementation of NPIs, before herd-immunity. Our results also confirm that NPIs were not stringent enough to consistently reduce the  $R_e$  of SARS-CoV-2 Amazonian lineages to below 1.0 and that a stationary state of endemic community transmission was maintained from May to September 2020<sup>7,8</sup>.

September 2020 onwards, mitigation measures were relaxed, and the  $R_e$  of clade 28-AM-I returned to above 1.0. Notwithstanding, the second epidemic wave only started in December 2020, coinciding with the emergence of the VOC P.1 and the second lineage replacement event. Several complementary evidence supports that these events were probably driven by the emergence of a more transmissible VOC P.1 in a context of relaxed social distancing. First, the second lineage replacement event was an abrupt process as the VOC P.1 evolved from the local clade 28-AM-II around late November 2020, and it took less than two months to become the dominant variant. Second, the estimated median  $R_e$  of the VOC P.1 during December 2020 was 2.2 times higher than that estimated for clade 28-AM-I in the same period, indicating that P.1 could have been nearly two times more transmissible than the co-circulating B.1.1.28 parental lineage. Third, the level of SARS-CoV-2 RNA (estimated from the median Ct) in the URT samples from P.1 infections, particularly from adults (18–59 years old), was ~ 10-fold higher than the level detected in non-P.1 infections; suggesting that P.1-infected adult individuals are more infectious than those harboring non-P.1 viruses<sup>9,10</sup>. Phylodynamic modeling also indicates that NPIs implemented in Manaus since the late-December were effective to reduce the median  $R_e$  of the VOC P.1 ~ 50% (from 2.6 to 1.2), but probably failed to bring the epidemic under control ( $R_e < 1.0$ ), allowing the continued spread of this VOC in the Amazonas state.

Understanding the factors that drive the emergence and expansion of VOC harboring key mutations in the RBD of the Spike protein is of crucial importance. One hypothesis is that VOC evolved by the selective pressure of a large number of people having developed antibodies against SARS-CoV-2. Our study revealed no unusual pattern of intrahost viral variability in the Amazonian clades between April and December 2020, showing that the local emergence of VOC in heavily affected regions is an evolutionary event challenging to anticipate from the analysis of parental lineages. Concurrently, we identified a P.1-like virus in Manaus in December 2020 that harbors several of the P.1 lineage-defining mutations and probably shared a most recent common ancestor with lineage P.1 in September 2020. This finding revealed that the diversity of SARS-CoV-2 variants carrying mutations of concern at the Spike protein in Manaus could be larger than initially described and that those variants probably circulated for some time before the expansion of lineage P.1. Although only the lineage P.1 seems to have displayed a rapid dissemination so far, our findings alert for the potential spread of other P.1-related VOC in the Amazonas state and for the importance of a continuous molecular surveillance system to track the viral diversity in real-time in Brazil.

These findings collectively support those lineage replacements could be a recurrent phenomenon in the local evolution of SARS-CoV-2 driven by ecological and virological factors. Moreover, our findings indicate that NPIs deployed in the Amazonas state in April 2020 were sufficiently effective to reduce the  $R_e$  of early prevalent local SARS-CoV-2 clades. In contrast, NPIs were insufficient to bring the epidemic under control, allowing the establishment and local persistence of several endemic viral lineages and subsequent emergence of the VOC P.1 in late November 2020. The lack of efficient social distancing and other mitigation measures probably accelerated the early transmission of VOC P.1, while the high transmissibility of this VOC further fueled the rapid upsurge in SARS-CoV-2 cases and hospitalizations



observed in Manaus following its emergence. The feeble adoption of non-pharmaceutical interventions as it occurred in Amazonas and other Brazilian states represents a significant risk for the continuous emergence and dissemination of new variants. Implementing efficient mitigation measures combined with massive vaccination will be crucial to control the spread of SARS-CoV-2 VOCs in Brazil.

## Materials And Methods

**SARS-CoV-2 samples and ethical aspects.** We collected nasopharyngeal and pharyngeal swabs (NPS) from 644 residents in the Amazonas state which were positively tested by Real-Time PCR as a routine diagnostic for COVID-19 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>). Among those 644 NPS samples, 250 were submitted to nucleotide sequencing and other 394 were evaluated for P.1/VOCs by the real-time PCR developed in this study. Fiocruz/ILMD is one of the official labs for SARS-CoV-2 testing under the auspices of a network coordinated by the Amazonas State Health Foundation (FVS-AM) and the Brazilian Ministry of Health. This study was conducted as a request of the SARS-CoV-2 surveillance program of FVS-AM and was approved by the Ethics Committee of the Amazonas State University CAEE: 25430719.6.0000.5016.

**SARS-CoV-2 P.1/VOCs detection by Real-time PCR.** A total of 394 SARS-CoV-2 positive samples collected from 1st November to 13th January were submitted to a real-time PCR screening test designed for detection of VOCs that use a forward primer (P1/VOCs-FNF 5'- GGGTGATGCGTATTATGACATGGTTGG), a reverse primer (P1/VOCs-FNR 5'- CTAGCACCATCATCATAACAGTTCTTGC) and a probe (P1/VOCs-FNP 5'(ZEN)- TGGTTGATACTAGTTTGAAGCTAAAA) to detect the deletion in the ORF1b (NSP6: S106del, G107del, F108del) found in the three VOCs (P.1; B.1.1.7 and B.1.351), both primers were used at 300nM and probe at 150nM (final concentration) with TaqMan one-step Fast Virus master Mix (ThermoFisher Scientific). We validated this assay against 89 high-quality full SARS-CoV-2 genomes, 39 non-P.1 and 48 P.1 (further details in the **Supplementary Table 5**). All oligos used in this study were manufactured by IDT DNA (<https://www.idtdna.com/>). Because we do not detect circulation of VOCs B.1.1.7 and B.1.351 in the Amazonas state, we use the frequency of NSP6 deletion target among PCR positives as a reliable proxy for frequency of the VOC P.1.

**SARS-CoV-2 amplification and sequencing.** A total of 250 SARS-CoV-2 positive samples (122 male, 128 female; median age 43-year-old IQR 32-46) collected from residents of 25 out of 62 municipalities of the Amazonas state, including the capital Manaus, between 13th March and 13th January were subjected to amplification and NGS sequencing as previously described <sup>1</sup>, now with a reduced number of amplicons (nine instead of 15) of a mean average size of ~3,500bp <sup>2</sup>. Briefly, RNAs were extracted with Maxwell® RSC Viral Total Nucleic Acid Purification Kit (Promega) and then converted to cDNA with Superscript IV reverse transcriptase (ThermoFisher Scientific). Amplicons covering the entire SARS-CoV-2 genome were amplified with SuperFi II Green PCR master mix (ThermoFisher Scientific), precipitated with PEG 8000 (Promega), and quantified using a fluorimeter. Pooled amplicons of each sample were used to prepare

NGS libraries with Nextera XT and clustered with MiSeq Reagent Kit v2 (500-cycles) on 2 x 150 cycles paired-end runs (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 BBDuk tools (v37.25) and the NC\_045512.2 RefSeq as a template. Using this approach, we generated consensus sequences with mean depth coverage of 2,600X, excluding duplicate reads. Whole-genome SARS-CoV-2 consensus sequences were initially assigned to viral lineages according to the nomenclature proposed by Rambaut et al. (doi.org/10.1038/s41564-020-0770-5), using the Pangolin web application (<https://pangolin.cog-uk.io>) and later confirmed using phylogenetic analyses as explained below.

**Intra-host SARS-CoV-2 genomic variability.** Raw sequencing reads and primer sequences were removed with Trimmomatic 0.26<sup>3</sup> using default parameters. Reads that passed quality filtering were then mapped against the Wuhan SARS-CoV-2 reference genome (NC\_045512.2) using Bowtie2 software<sup>4</sup>. A .bed file was generated with bedtools v 2.15.0<sup>5</sup>, samtools<sup>6</sup> and vcftools v 0.1.13<sup>7</sup> using the internal parameter vcf-annotated (parameters-filter Qual = 20/MinDP = 100/SnpGap = 20, meaning that only nucleotide variants supported by reads with mapping quality  $\geq 20$  and at least 100 sequencing coverage depth would be retained in the intermediate variant call file). To characterize the viral intra-host population, we identified all minor variants (MVs) found in the samples, that is, highly supported nucleotides that are supported by 10 to 49% of the reads in a given position, and that was not added in the final majority consensus genome. We then replaced the nucleotide supported by the majority of the reads by the MVs in the consensus genome to evaluate the impact of the synonymous and nonsynonymous nucleotide variation between the major and minor variants. We performed the synonymous and nonsynonymous analysis using a R pipeline developed for SARS-CoV-2 (10.3389/fmicb.2020.01800).

**Discrete Maximum likelihood and Bayesian phylogeography.** All high quality (<1% of N) complete (>29 kb) SARS-CoV-2 genomes of lineages B.1.1.28 ( $n = 512$ ) and B.1.1.33 ( $n = 595$ ) sampled in Brazil and of lineage B.1.195 sampled worldwide ( $n = 110$ ), that were available on GISAID (<https://www.gisaid.org/>) as of January 13, 2021, were downloaded. SARS-CoV-2 complete genome sequences were aligned using MAFFT v7.467<sup>8</sup>. The B.1.1.28 and B.1.1.33 datasets were subject to maximum likelihood (ML) and phylogenetic analysis using IQ-TREE v1.6.12<sup>9</sup> under the GTR + F + I + G4 nucleotide substitution model, as selected by the ModelFinder application<sup>10</sup>, and the branch support was assessed by the approximate likelihood-ratio test based on the Shimodaira–Hasegawa-like procedure (SH-aLRT) with 1,000 replicates. Time-scaled phylogeographic ML phylogenetic trees of Brazilian B.1.1.28 and B.1.1.33 datasets were reconstructed using Treetime<sup>11</sup> with a fixed substitution rate ( $8 \times 10^{-4}$  substitutions/site/year) coupled with an ancestral character reconstruction (ACR) of epidemic locations using PASTML<sup>12</sup> with Marginal Posterior Probabilities Approximation (MPPA) with an F81-like model. A time-scaled Bayesian phylogeographic analysis was performed for the B.1.195 sampled worldwide using the Bayesian Markov Chain Monte Carlo (MCMC) approach implemented in BEAST 1.10<sup>13</sup> with BEAGLE library v3<sup>14</sup> to

improve computational time. Bayesian tree was reconstructed using the GTR + F + I + G4 nucleotide substitution model, the non-parametric Bayesian skyline (BSKL) model as the coalescent tree prior <sup>15</sup>, a strict molecular clock model with a uniform substitution rate prior ( $8 \times 10^{-4}$  substitutions/site/year) and a reversible discrete phylogeographic model <sup>16</sup> with a continuous-time Markov chain (CTMC) rate reference prior <sup>17</sup>. Three MCMC chains were run for 100 million generations and then combined to ensure stationarity and good mixing. Convergence (effective sample size > 200) in parameter estimates was assessed using TRACER v1.7 <sup>18</sup>. 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/>).

**Continuous Bayesian phylogeography.** The phylogenetic diffusion of the SARS-CoV-2 clades from Amazonas state was estimated with the heterogeneous relaxed random walk (RRW) model and a Cauchy distribution <sup>19</sup>, previously applied to SARS-CoV-2 in Brazil <sup>20</sup>, using BEAST v.1.10.4 as explained above. The viral spatio-temporal diffusion was analyzed and visualized in SPREAD <sup>21</sup> and further projected in maps generated with the QGIS software (<http://qgis.org>) using public access data downloaded from the GADM database (<https://gadm.org>).

**Effective Reproductive Number (Re) Estimation.** To estimate the Re of the Amazonian SARS-CoV-2 clades through time, we used the birth-death skyline (BDSKY) model <sup>22</sup> implemented within BEAST2 v2.6.2 <sup>23</sup>. 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 for the BDSKY model 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 six time-intervals (monthly) to the 195-AM clade, five time-intervals (bimonthly) to the 28-AM-I clade and two equal time-intervals to the P.1 clade. Time-intervals were defined from the date of the most recent sample up to the root of the tree. One MCMC chain was run for 20 million generations and then checked for stationarity and mixing, as explained above.

**Statistical analysis.** Descriptive statistics, test for normal distribution (D'Agostino & Pearson and Anderson-Darling), and the non-parametric Mann-Whitney test were used to compare the cycle threshold (Ct) of SARS-CoV-2 RT-PCR positive samples from the upper respiratory tract of patients infected with P.1 vs. non-P.1 viruses. Only Ct values from samples analyzed with the same RT-PCR diagnostic assay (e.g., the CDC assay) were compared. The threshold for statistical significance was set to  $P < 0.05$ . Graphics and statistical analyses were performed using GraphPad v9.01 and 9.02 (Prism Software, United States).

## Declarations

### 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, in particular, the

Japanese National Institute of Infectious Diseases (NIID) members Dr. Tsuyoshi Sekizuka, Dr. Kentaro Ito, Rina Tanaka and Masanori Hashino to publish the genomes. GISAID acknowledgment table containing sequences used in this study is shown in **Supplementary Table 6**. We would also wish to thank Dr. Nuno Faria from <https://www.caddecentre.org> for sharing their unpublished findings regarding the SARS-CoV-2 B.1.1.28 lineage. We also appreciate the support of Genomic Coronavirus Fiocruz Network members and the Respiratory Viruses Genomic Surveillance Network of the General Laboratory Coordination (CGLab) of the 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 support FAPEAM (PCTI-EmergeSaude/AM call 005/2020 and Rede Genômica de Vigilância em Saúde - REGESAM); Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant 403276/2020-9); Inova Fiocruz/Fundação Oswaldo Cruz (Grant VPPCB-007-FIO-18-2-30 - Geração de conhecimento). In the loving memory of all we have lost due COVID-19.

### Author contributions

FGN contributed to writing of the report, data analysis, laboratory management, and obtaining financial support. VN, VS, AC, FN, GS, AC, DD, KP, MM, MJB, MJ, and LG contributed to diagnostics and sequencing analysis. CFC, VS, DB, MS, TM, GP, LA, JHS contributed to patient and public health surveillance data. IA, FZD contributed to formal data analysis of sequence diversity. MMS, GLW, PCR, ED, TG, and GB contributed to formal data analysis, writing, and editing of the report.

### Competing interests

All authors have declared that no conflicts of interest exist.

### Data availability

All the SARS-CoV-2 genomes generated and analyzed in this study are available at GISAID (<https://www.gisaid.org/>), under the IDs EPI\_ISL\_792560; EPI\_ISL\_801386 - EPI\_ISL\_801403; EPI\_ISL\_811148; EPI\_ISL\_811149; EPI\_ISL\_833131 - EPI\_ISL\_833140; EPI\_ISL\_1034304 - EPI\_ISL\_1034306; EPI\_ISL\_1068078 - EPI\_ISL\_1068292.

## References

1. 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).
2. Amazonas., FVS. (ed Fundação em Vigilância e Saúde do Amazonas) (<http://www.fvs.am.gov.br/>, 2021).
3. Fujino, T. *et al.* Novel SARS-CoV-2 Variant Identified in Travelers from Brazil to Japan. *Emerg Infect Dis* **27**, doi:10.3201/eid2704.210138 (2021).

4. Faria, N. C., M.I.; Candido, D., Franco, L.A.M; Andrade, P; Coletti, T.; Silva, C.A.M, Sales, F.C, Manuli, E.R.; Aguiar, R.A; Gaburo N.; Camilo, C.C.; Fraiji, N.A.; Crispim, C.A.E.; Carvalho, M.P.S.S.; Rambaut, A.; Loman, N., Pybus, O.; Sabino, E.; on behalf of CADDE Genomic Network;. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virologica.org*, (2021).  
<<https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>>.
5. 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).
6. Buss, L. F. *et al.* Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371**, 288–292, doi:10.1126/science.abe9728 (2021).
7. Mellan, T. A. *et al.* Subnational analysis of the COVID-19 epidemic in Brazil. *medRxiv*, doi:10.1101/2020.05.09.20096701 (2020).
8. He, D., Artzy-Randrup, Y., Musa, S. S. & Stone, L. The unexpected dynamics of COVID-19 in Manaus, Brazil: Herd immunity versus interventions. *medRxiv*, doi:10.1101/2021.02.18.21251809 (2021).
9. Singanayagam, A. *et al.* Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.32.2001483 (2020).
10. Jaafar, R. *et al.* Correlation between 3790 qPCR positives samples and positive cell cultures including 1941 SARS-CoV-2 isolates. *Clin Infect Dis*, doi:10.1093/cid/ciaa1491 (2020).