

# Temporal spread and evolution of SARS-CoV-2 in the second pandemic wave in Brazil

Jonas M. Wolf<sup>1,2</sup> | Diéssy Kipper<sup>3</sup> | Gabriela R. Borges<sup>2</sup> | André F. Streck<sup>3</sup> |  
Vagner R. Lunge<sup>1,2,4</sup>

<sup>1</sup>Laboratório de Diagnóstico Molecular, Programa de Pós-Graduação em Biologia Celular e Molecular Aplicada à Saúde, Universidade Luterana do Brasil, ULBRA, Canoas, Rio Grande do Sul, Brazil

<sup>2</sup>Laboratório de Diagnóstico Molecular, Universidade Luterana do Brasil, Canoas, Rio Grande do Sul, Brazil

<sup>3</sup>Laboratório de Diagnóstico em Medicina Veterinária, Universidade de Caxias do Sul, UCS, Caxias do Sul, Rio Grande do Sul, Brazil

<sup>4</sup>Simbios Biotecnologia, Cachoeirinha, Rio Grande do Sul, Brazil

## Correspondence

Jonas Michel Wolf, Laboratório de Diagnóstico Molecular - ULBRA, Av. Farroupilha, 8001 - Bldg 22 - Room 312, 92425-900/Canoas, RS, Brazil.  
Email: [jonasmwolf@gmail.com](mailto:jonasmwolf@gmail.com)

## Funding information

Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul

## Abstract

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic spread rapidly and this scenario is concerning in South America, mainly in Brazil that presented more than 21 million coronavirus disease 2019 cases and 590 000 deaths. The recent emergence of novel lineages carrying several mutations in the spike protein has raised additional public health concerns worldwide. The present study describes the temporal spreading and evolution of SARS-CoV2 in the beginning of the second pandemic wave in Brazil, highlighting the fast dissemination of the two major concerning variants (P.1 and P.2). A total of 2507 SARS-CoV-2 whole-genome sequences (WGSs) with available information from the country (Brazil) and sampling date (July 2020–February 2021), were obtained and the frequencies of the lineages were evaluated in the period of the growing second pandemic wave. The results demonstrated the increasing prevalence of P.1 and P.2 lineages in the period evaluated. P.2 lineage was first detected in the middle of 2020, but a high increase occurred only in the last trimester of this same year and the spreading to all Brazilian regions. P.1 lineage emerged even later, first in the North region in December 2020 and really fast dissemination to all other Brazilian regions in January and February 2021. All SARS-CoV-2 WGSs of P.1 and P.2 were further separately evaluated with a Bayesian approach. The rates of nucleotide and amino acid substitutions were statistically higher in P.1 than P.2 ( $p < 0.01$ ). The phylodynamic analysis demonstrated that P.2 gradually spread in all the country from September 2020 to January 2021, while P.1 disseminated even faster from December 2020 to February 2021. Skyline plots of both lineages demonstrated a slight rise in the spreading for P.2 and exponential growth for P.1. In conclusion, these data demonstrated that the P.1 (recently renamed as Gamma) and P.2 lineages have predominated in the second pandemic wave due to the very high spreading across all geographic regions in Brazil at the end of 2020 and beginning of 2021.

## KEYWORDS

dissemination, pandemic, SARS coronavirus

## 1 | INTRODUCTION

Brazil is the Latin American country with the largest number of deaths due to coronavirus disease 2019 (COVID-19), recording more than 590 000 deaths and more than 21 million reported cases, updated on September 2021 (<https://coronavirus.jhu.edu/map.html>). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and COVID-19 burden have been highly variable across the country, with North Brazil being the first and worst-affected region.<sup>1</sup> A difference in the fatality rates of SARS-CoV-2 infection across the country was observed in the first pandemic wave, possibly related to the diverse demographic composition, the control of the measures adopted to reduce viral spreading, and the occurrence of different viral lineages.<sup>2-8</sup>

SARS-CoV-2 has been extensively studied in the last year and different mutations in the viral genome have been demonstrated in the lineages disseminating across the world.<sup>9,10</sup> This genetic diversity has impacted on pathogenicity and transmissibility of the virus<sup>11,12</sup> as well as it is an additional concern in the development of appropriate vaccines.<sup>13</sup> Also, some SARS-CoV-2 variants could be associated with mild clinical outcomes.<sup>14,15</sup>

Since the emergence of the SARS-CoV-2, the combination between the unprecedented number of cases and more than 500 000 genomes quickly sequenced allowed the identification of hundreds of circulating viral variants.<sup>16</sup> Several variants of interest (VOIs) were already identified and they have differential virological characteristics, associated to improved transmission, high pathogenicity, and immunological escape.<sup>17</sup> Four main VOIs (B.1.1.7, B.1.351, P.1, and P.2) carrying several mutations in the receptor-binding domain (RBD) of the spike (S) protein raise more concerns about their potential to shift the viral dynamics and to impact public health before the emergence of the variant Delta.<sup>5,16,18-23</sup> These VOIs were isolated for the first time in different countries of the world, such as UK (B.1.1.7), South Africa (B.1.351), and Brazil (P.1 and P.2).<sup>12,18,21,23,24</sup> They are potentially associated with (I) increased transmissibility, (II) escape from neutralizing antibodies, (III) propensity for re-infection, and (IV) increased affinity for the human ACE2 receptor. Three of them were recently redefined as variants of concern (VOCs) due to the high dissemination worldwide: B.1.1.7 (Alpha), B.1.351 (Beta), and P.1 (Gamma).<sup>20,25</sup>

In Brazil, three main SARS-CoV-2 clades were introduced from March to April 2020: G, GR, and GH.<sup>26-28</sup> This first pandemic wave was mostly driven by B.1.195 lineage, which was subsequently replaced by B.1.1.28.<sup>19</sup> Later, P.2 lineage (B.1.1.28 ancestral lineage) was detected in Rio de Janeiro by genome sequencing in October 2020 and it was estimated to have emerged in July 2020.<sup>29</sup> P.1 lineage (B.1.1.28.1 ancestral lineage) was first identified in January 2021 in Brazilian travelers who arrived in Japan.<sup>30</sup> After, it was also identified in several states in the country, mainly in the North region.<sup>5,7,24</sup> This new pandemic wave resulted in an exponential growth in hospitalizations and deaths since mid-December 2020.<sup>5-8</sup> Therefore, the main objective of this study was to understand the temporal spread and evolution of SARS-CoV-2 in the second pandemic wave,

with a special focus on the phylodynamics of P.1 (Gamma variant) and P.2 lineages in Brazil, through a Bayesian approach with SARS-CoV-2 whole genome sequences (WGSs) obtained from July 2020 to February 2021.

## 2 | METHODS

### 2.1 | Data collection

SARS-CoV-2 WGSs with available information from country (Brazil) and the month/year of sampling were obtained in Global Initiative on Sharing All Influenza Data (GISAID) (<https://www.epicov.org/epi3/frontend#502179>) from July 2020 to February 2021. Two independent data sets were assembled with all obtained WGS of each lineage: P.1 (Table S1) and P.2 (Table S2). The remaining WGSs of other lineages are described in Table S3. All lineages' classifications were confirmed by BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Putative recombination events were verified using the Recombination Detection Program version 4 (RDP4) software<sup>31</sup> with the default settings using the algorithms RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, 3Seq, and LARD. The beginning and end breakpoints of the potential recombinant sequences were also defined by the RDP4 software. Putative recombinant events were considered significant when  $p \leq 0.01$  was observed for the same event using four or more algorithms. Sequences that presented putative recombination events, clonal dissemination, and missing record of the collection were excluded from the evolutionary analysis.

### 2.2 | Mutation analysis

Alignments were performed for P.1 and P.2 lineages by MAFFT v7<sup>32</sup> and visually inspected with AliView v1.26. We identified nucleotide and amino acids substitutions for P.1 and P.2 lineages using Nextclade v0.8.1 (<https://clades.nextstrain.org/>) and Pangolin v2 (<https://github.com/covlineages/pangolin>).

### 2.3 | Root-to-tip regression

Alignments of P.1 and P.2 data sets were performed by MAFFT v7<sup>32</sup> and visually inspected with AliView v1.26. The best-fitting nucleotide substitution (HKY) model was selected using a hierarchical likelihood ratio, Akaike information criterion, and Bayesian information criterion tests with Model Finder in IQ-TREE webserver.<sup>33</sup> SARS-CoV-2 WGSs maximum likelihood phylogenetic tree were inferred according to the best-fitting model using IQ-TREE webserver (<http://iqtree.cibiv.univie.ac.at/>). Statistical supports for internal branches in the phylogeny were assessed by bootstrapping (1000 replicates) and the approximate likelihood ratio test.<sup>34</sup> The resulting tree was used to further obtain root-to-tip regressions in TempEst v1.5<sup>35</sup> by selecting the root position that maximized the correlation coefficient. The

root-to-tip versus divergence plot of the full data sets showed a correlation between sampling time and genetic distance to the root of the ML tree of the available sequences ( $R^2 = 0.66$  for P.1 and  $R^2 = 0.78$  for P.2), suggesting moderate temporal signals and the possibility to calibrate a reliable molecular clock.

## 2.4 | Bayesian coalescent inference

Time-scaled phylogenetic tree estimation was performed using BEAST/BEAGLE v2.5 software.<sup>36</sup> The best-fitting nucleotide substitution (HKY) model with gamma site distribution was selected using a hierarchical likelihood ratio, Akaike information criterion, and Bayesian information criterion tests with Model Finder in IQ-TREE webserver (<http://iqtree.cibiv.univie.ac.at/>). For each run of 500 million of Monte Carlo Markov Chains (MCMC), the marginal likelihood was estimated via path sampling (PS) and stepping stone (SS) methods and the resulting Bayes Factors (BF) (ratio of marginal likelihoods) used to select the best-fitting clock/demographic model. The models can be compared to evaluate the strength of evidence against the null hypothesis (H0) defined in the following way:  $2\ln BF < 2$  indicates no evidence against H0; 2–6, weak evidence; 6–10: strong evidence, and more than 10 very strong evidence.<sup>37</sup> Both SS and PS estimators indicated the uncorrelated lognormal molecular clock as the best-fitted model to the data sets under analysis (Bayes Factor > 10 for both data sets). MCMC analysis was performed and the maximum likelihood estimations of the obtained trees were compared using a BF to select the best model and parameter values. BF analysis showed that the strict clock fitted the data significantly better than other clocks ( $2\ln BF > 500$  for both data sets). BF analysis showed that the Bayesian skyline plot (BSP) was better than other models ( $2\ln BF > 100$  for both data sets). MCMC was run for 500 million generations to ensure stationary and adequate effective sample size for all statistical parameters. Tracer v1.6 software<sup>38</sup> was used to diagnose MCMC, to adjust initial burn-in, and to perform the Skyline demographic reconstruction. Uncertainty in parameter estimates was evaluated in the 95% highest posterior density (HPD95%) interval. TreeAnnotator v1.8.2 was used to summarize the maximum clade credibility (MCC) tree from the posterior distribution of trees and the MCC tree was visualized and edited in FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

## 2.5 | Statistical analysis

Statistical analyses were performed using the SPSS (Statistical Package for Social Sciences, version 23.0) software. The normality of the quantitative variables was evaluated by visual inspection of histograms and by Shapiro–Wilk and Kolmogorov–Smirnov tests. Nucleotide and amino acid mutations of P.1 and P.2 lineages were represented by the mean, standard deviation, minimum and maximum values. These data were represented by a bar chart. The Student's *t*-test for independent samples was used to compare the mean

mutations between P.1 and P.2 lineages. All estimates were bilateral with a significance level of 5% ( $p < 0.05$ ).

## 3 | RESULTS

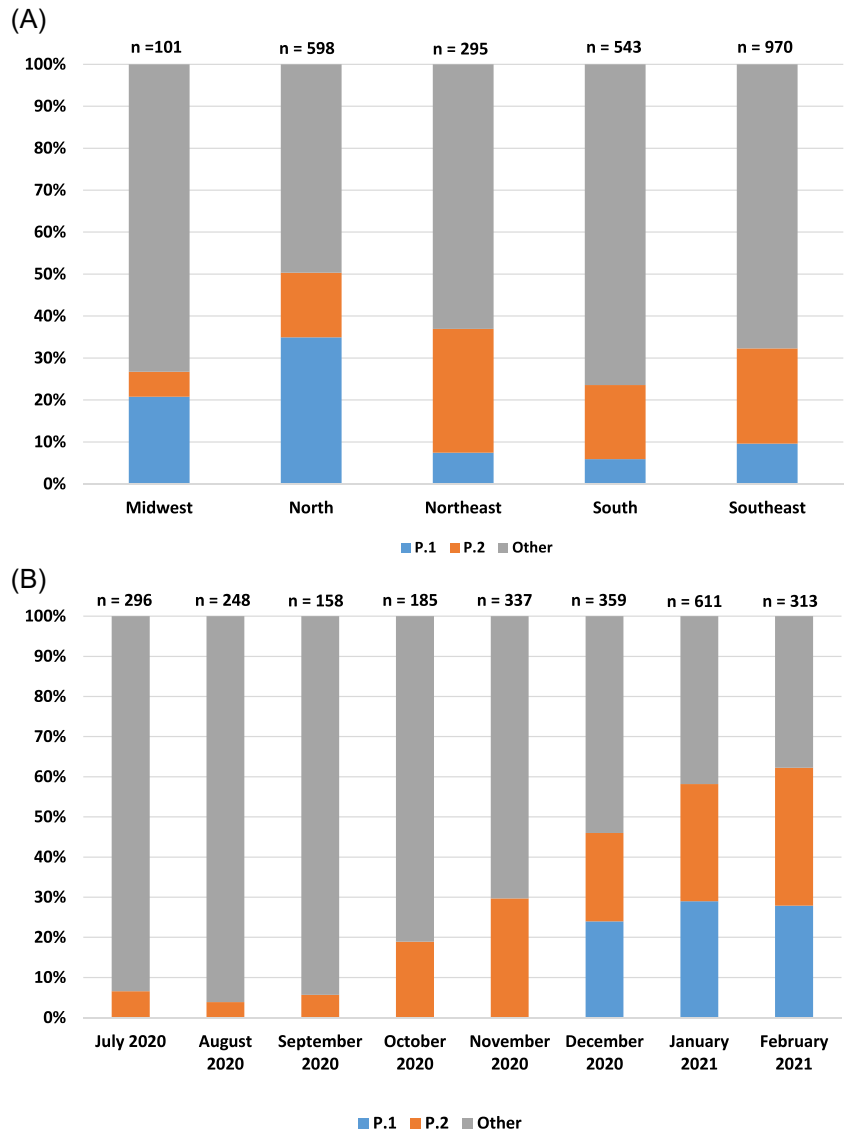
### 3.1 | SARS-CoV-2 P.1 and P.2 lineages in the second pandemic wave in Brazil

A total of 2507 SARS-CoV-2 WGSs of all Brazilian regions were obtained from GISAID between July 2020 and February 2021. The total number of WGSs from the P.1 and P.2 lineages were 377 (15.0%) and 501 (20.0%), respectively. P.1 was more frequent in the North ( $n = 209$ ; 34.9%) and in the Midwest ( $n = 21$ ; 20.8%), and less frequent in the Southeast ( $n = 93$ ; 9.6%), in the Northeast ( $n = 22$ ; 7.5%), and in the South ( $n = 32$ ; 5.9%). On opposite, P.2 was more frequent in the Northeast ( $n = 87$ ; 29.5%) and in the Southeast ( $n = 220$ ; 22.7%), and less frequent in the South ( $n = 96$ ; 17.7%), in the North ( $n = 92$ , 15.4%), and in the Midwest ( $n = 6$ ; 5.9%) (Figure 1A).

P.2 WGS lineage was firstly sequenced in July 2020, with 21 (7.1%) sequences reported in this month. After, this lineage was reported in August ( $n = 10$ ; 4.4%), September ( $n = 9$ ; 6.3%), October ( $n = 35$ ; 19.6%), November ( $n = 101$ ; 30.5%), December ( $n = 79$ ; 22.8%) of 2020, and after in January ( $n = 196$ ; 32.4%), and February ( $n = 119$ ; 38.3%) of 2021. However, P.1 was firstly demonstrated in December 2020 ( $n = 86$ ; 24.5%), and after in January ( $n = 196$ ; 32.4%), and February 2021 ( $n = 97$ ; 31.1%) (Figure 1B).

Two P.2 WGSs were first reported in the Midwest ( $n = 1$ ) and the Northeast ( $n = 1$ ) in July 2020. There were only two more published P.2 WGSs in August and September 2020, but this number increased to 35 in October 2020, most of them in the Southeast ( $n = 21$ ), but also in the North ( $n = 7$ ), in the Midwest ( $n = 3$ ), in the Northeast ( $n = 2$ ) and in the South ( $n = 2$ ) Brazil. In November 2020, the number of P.2 WGSs increased mainly in the South ( $n = 62$ ) and Southeast ( $n = 30$ ), but it was also detected in the North ( $n = 5$ ) and Northeast ( $n = 3$ ). In December 2020, P.2 WGSs were reported mainly in three regions: North ( $n = 34$ ), Southeast ( $n = 22$ ), and South ( $n = 14$ ). In January 2021, P.2 WGSs were more widely identified in the Southeast ( $n = 109$ ) and less in the North ( $n = 31$ ) and South ( $n = 17$ ), but it was also detected in the Northeast ( $n = 51$ ). Finally, in February 2021, P.2 WGSs were reported in the Southeast ( $n = 53$ ), Northeast ( $n = 23$ ), North ( $n = 8$ ), and South ( $n = 3$ ). P.1 lineage WGSs were reported only in the North region in December 2020 ( $n = 86$ ). In January 2021, there was an increase in the number of the P.1 WGSs in the North ( $n = 116$ ) and this lineage was also reported in the other four Brazilian regions: Southeast ( $n = 37$ ), Midwest ( $n = 21$ ), Northeast ( $n = 18$ ), and South ( $n = 3$ ) in January 2021. In February 2020, P.1 WGSs were reported in the Southeast ( $n = 56$ ), in the South ( $n = 29$ ), in the North ( $n = 7$ ), and in the Northeast ( $n = 4$ ) Brazil. The comparison among the frequencies of P.1, P.2, and other lineages in different Brazilian regions according to the growing second pandemic wave along the months is presented in Figure 2.

**FIGURE 1** (A) Distribution of P.1 (Gamma), P.2, and other lineages in Brazilian regions. (B) Distribution of P.1, P.2, and other lineages between July 2020 and February 2021 in Brazil. Data obtained from GISAID (<https://www.gisaid.org>). GISAID, Global Initiative on Sharing All Influenza Data



### 3.2 | P.1 and P.2 lineages nucleotide and amino acids substitutions

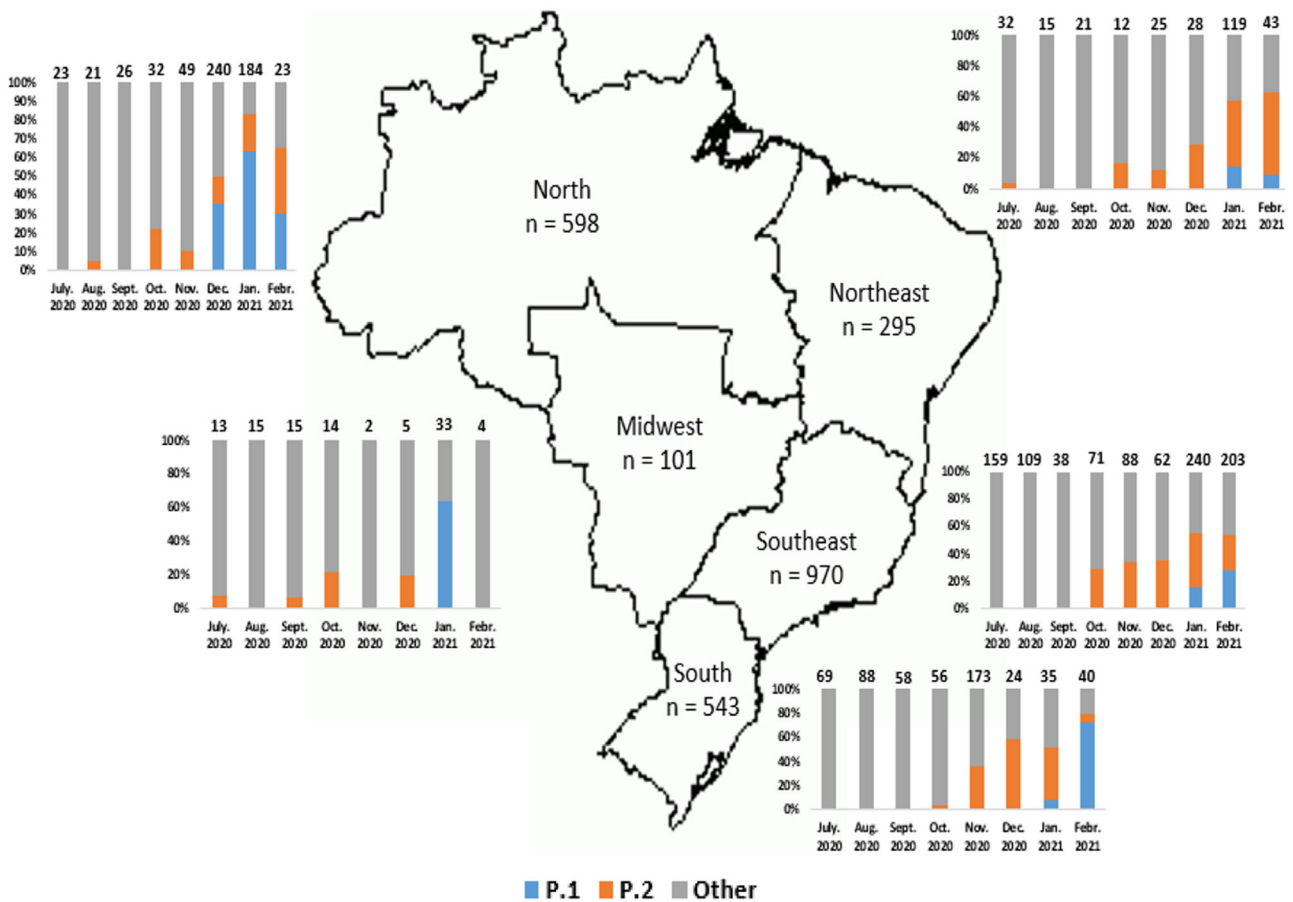
Nucleotides and amino acids substitutions were evaluated in the data sets of the alignments of the P.1 and P.2 lineages (Tables S4 and S5, respectively). The rates of nucleotide mutations were  $33.7 \pm 3.2$  (median = 34.0; minimum = 24.0; maximum = 54.0) for P.1 and  $23.4 \pm 4$  (median = 23.0; minimum = 15.0; maximum = 43.0) for P.2. In the evaluation of amino acids, the substitution rates were  $22.4 \pm 1.9$  (median = 22.0; minimum = 16.0; maximum = 35.0) for P.1 and  $12.7 \pm 2.0$  (median = 12.0; minimum = 8.0; maximum = 21.0) for P.2. The rates of nucleotide and amino acid substitutions were statistically higher in the P.1 lineage than P.2 ( $p < 0.01$ ) (Figure 3).

### 3.3 | P.1 lineage phylodynamic in Brazil

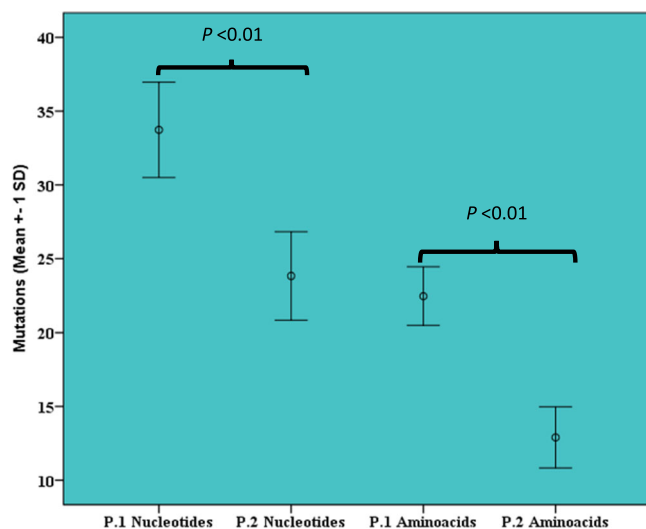
SARS-CoV-2 P.1 lineage phylogenetic tree showed one sequence from North Brazil (hCoV-19/Brazil/RR-1089/2021|

EPI\_ISL\_943972|2021-01-25) in the basal root tree. Two main clades were observed in the phylogenetic tree (I and II), both of them disseminated to all Brazilian regions. In Clade I ( $n = 86$ ), 46 WGSs were from the North (53.5%), 17 from the Southeast (19.8%), 11 from the Northeast (12.8%), seven from the South (8.1%), and five from the Midwest (5.8%). In Clade II ( $n = 290$ ), 162 WGSs were from the North (55.9%), 76 from the Southeast (26.2%), 25 from the South (8.6%), 16 from the Midwest (5.5%), and 11 from the Northeast (3.8%) (Figure 4A).

P.1 lineage root of the phylogenetic tree dated back to April 2020 and one sequence from North clustered in the more basal branches from the phylogenetic tree. P.1 lineage substitution rates was  $6.02E10^{-3}$  (HPD 95%:  $8.53E10^{-4}$  to  $1.23E10^{-3}$ ) nucleotides per site per year (s/s/y). Clades I and II dated back to the first fifteen days of January 2021. P.1 lineage strains seem to be disseminated from North to Southeast, Northeast, Midwest, and South (Figure 4A). The BSP analysis of P.1 lineage genomes showed that the dissemination accelerated mainly between 3 and 15 January 2021 (Figure 4B).



**FIGURE 2** Severe acute respiratory syndrome coronavirus 2 P.1 (Gamma), P.2, and other lineages in Brazilian regions between July 2020 and February 2021



**FIGURE 3** Mean bar chart of nucleotides and amino acids mutations of P.1 (Gamma) and P.2 lineages

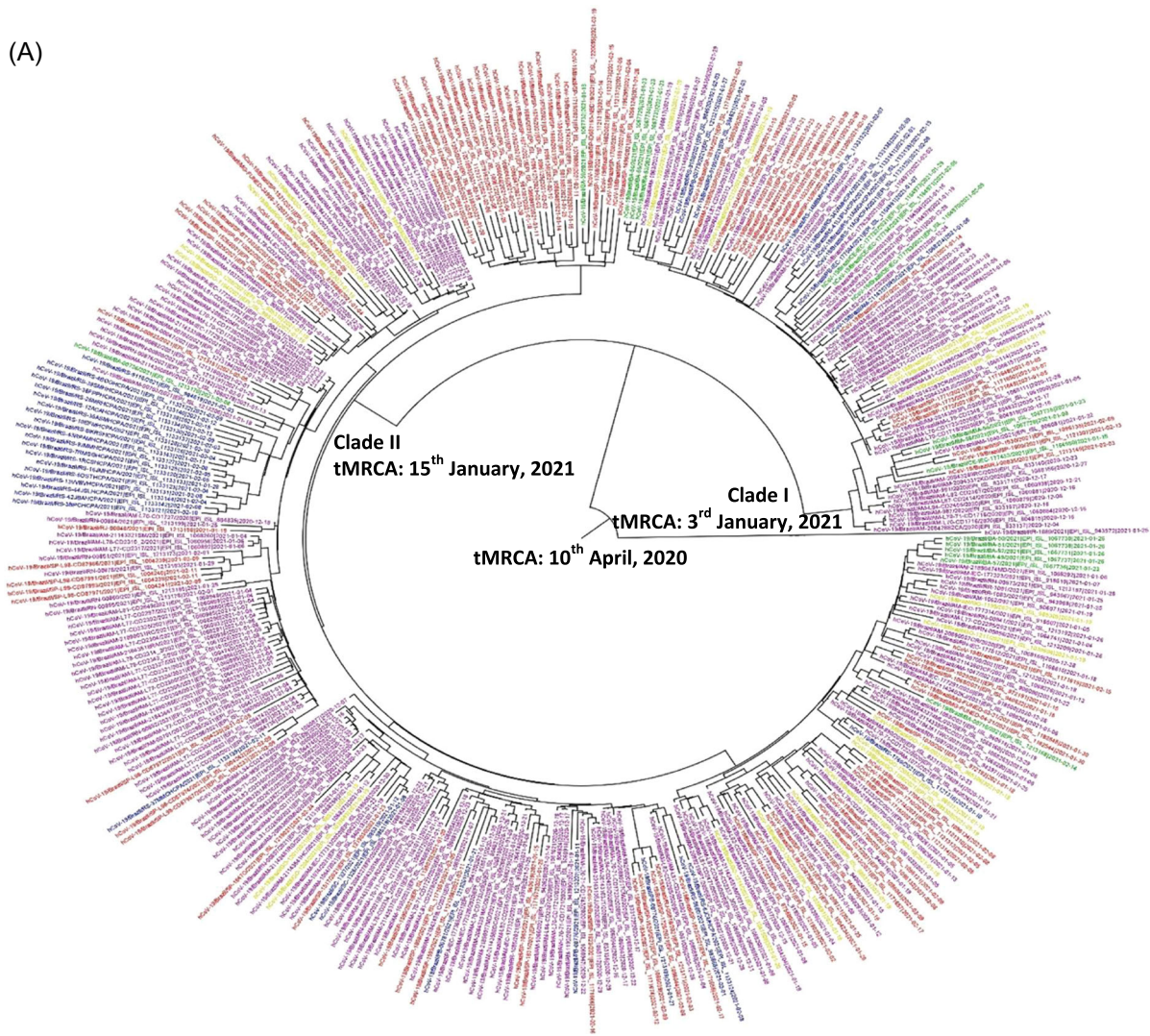
### 3.4 | P.2 lineage phylodynamic in Brazil

SARS-CoV-2 P.2 lineage phylogenetic tree has also shown one sequence from North Brazil (hCoV-19/Brazil/TO-1185/2020)

EPI\_ISL\_943984|2020-10-26) in the basal root tree. Six main clades were observed in the phylogenetic tree (I–VI). In Clade I ( $n = 8$ ), seven WGSs were obtained in the Southeast (85.7%) and one in the Midwest (14.3%). In Clade II ( $n = 72$ ), 36 WGSs were from the Northeast (50.0%), 17 from the North (23.6%), 13 from the Southeast (18.1%), three from the South (4.2%), and three from the Midwest (4.2%). In Clade III ( $n = 24$ ), 20 WGSs were from the Southeast (83.3%) and four from the Northeast (16.7%). In Clade IV ( $n = 14$ ), 10 WGSs were from the Southeast (71.4%), one from the Northeast (7.1%), one from the North (7.1%), one from the South (7.1%), and one from the Midwest (7.1%). In Clade V ( $n = 139$ ), 79 WGSs were from the South (56.8%), 47 from the Southeast (33.8%), seven from the Northeast (5.0%), and six from the North (4.3%). Finally, in Clade VI ( $n = 244$ ), 123 WGSs were from the Southeast (50.4%), 70 from the North (28.3%), 36 from the Northeast (15.2%), 13 from the South (5.3%), and two from the Midwest (0.8%) (Figure 5A).

P.2 lineage root of the phylogenetic tree dated back to April 2020 and one sequence from North clustered in the more basal branches from the phylogenetic tree. P.2 lineage substitution rates was  $3.20E10^{-3}$  (HPD 95%:  $7.32E10^{-4}$  to  $1.54E10^{-3}$ ) s/s/y. Clade I dated back to August 2020, Clades II and III to September 2020, Clade IV to December 2020, Clade V, and VI to January 2021. P.2

(A)



(B)

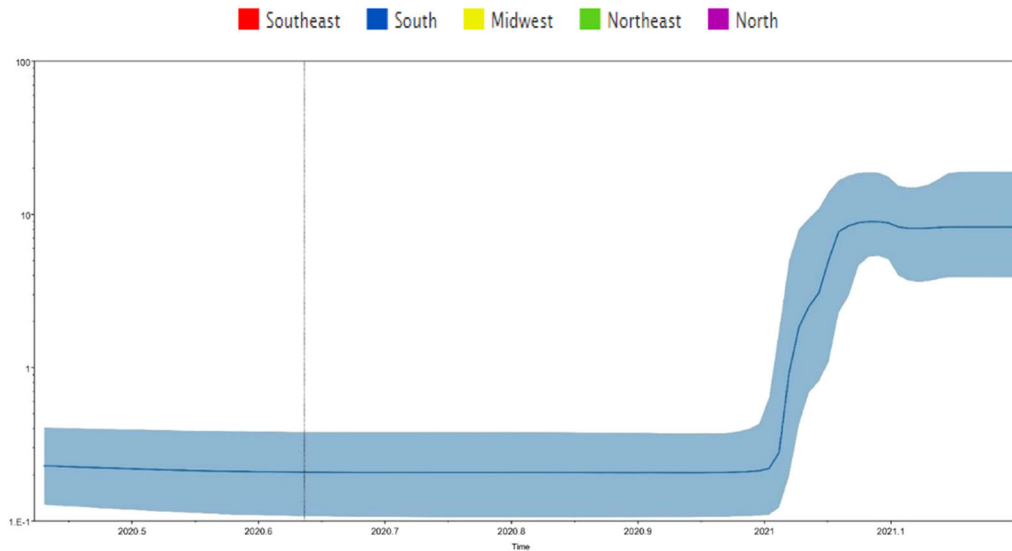


FIGURE 4 (See caption on next page)

lineage strains seem to be disseminated from North to Midwest, Southeast, Northeast, and South (Figure 5A). The BSP analysis of P.2 lineage genomes showed that the dissemination grew mainly between 15 September 2020 and 23 January 2021 (Figure 5B).

## 4 | DISCUSSION

The phylogenetic characterization of an emerging viral infection can help in monitoring the epidemic progression. Currently, there is a very large amount of WGSs data to understand the recent SARS-CoV-2 pandemic spreading. Phylogenetic analyses have been largely used to determine the SARS-CoV-2 genetic and antigenic diversity aiming to define essential issues related to the current pandemic. Brazil is one of the countries with the most concerning spreading of novel lineages and variants as a result of a massive SARS-CoV-2 growing dissemination wave from September 2020 to the present data. Two main SARS-CoV-2 VOIs, P.1 and P.2, were discovered through active genomic surveillance performed in all regions of the country, both characterized by specific sets of significant amino acid substitutions: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, and T1027I.<sup>7,18,20,24,39–41</sup> In addition, both VOIs rapidly spread in the last 6 months.<sup>7,8,19,20,24,41–43</sup>

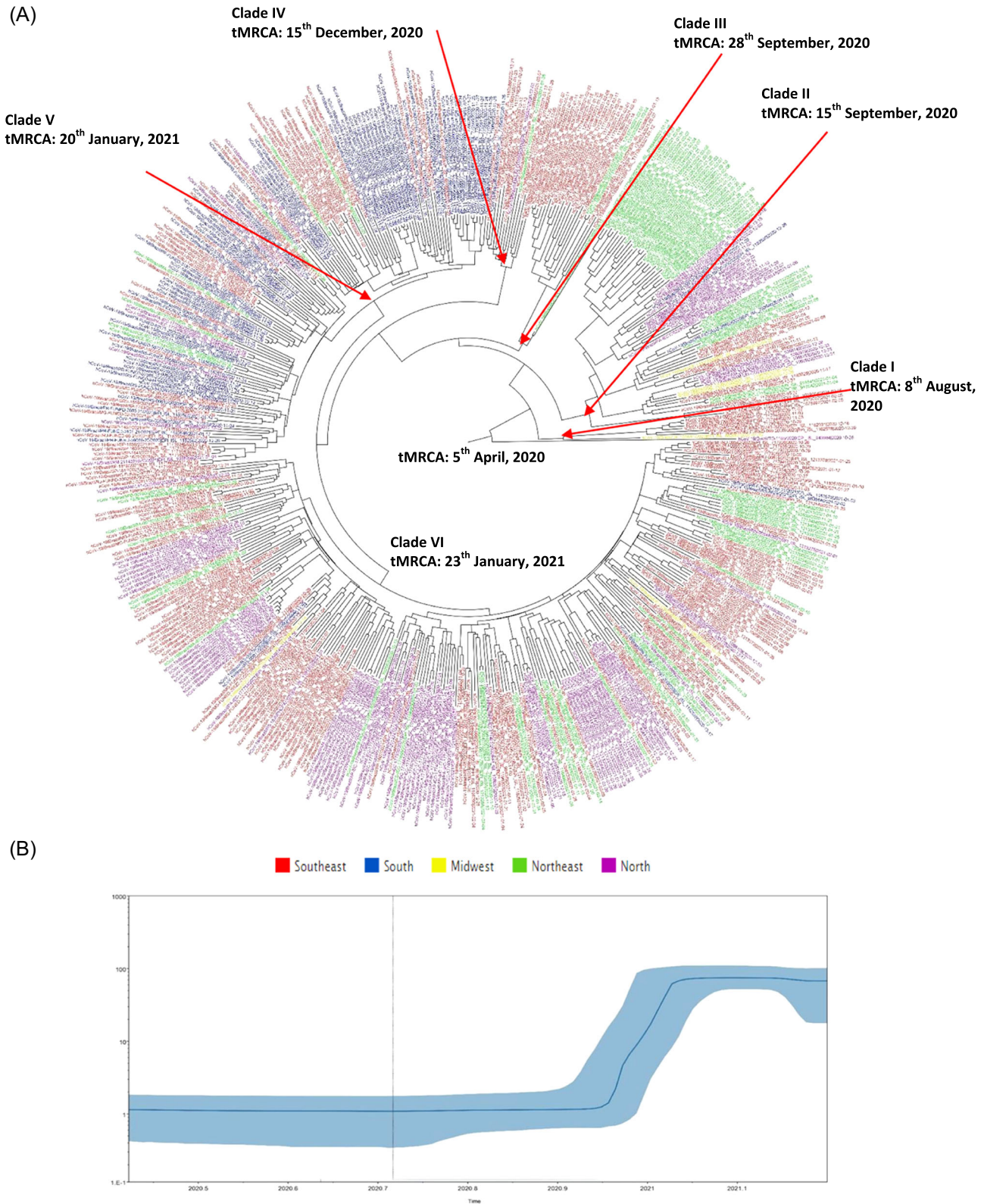
P.2 was first detected in July 2020 and represented 20% of all sequenced SARS-CoV-2 from this month until February 2021. Importantly, it increased from less than 10% in September 2020 to more than 30% of all lineages from October 2020 to February 2021. Further, it was first disseminated in the South and Southeast until December 2020 and after in the Northeast in February 2021. Other studies have also demonstrated the high frequency of P.2 first in the South and the Southeast and after in other Brazilian regions.<sup>24</sup>

In oppose, P.1 (Gamma) was detected for the first time only in December of 2020. This lineage is a descendent of B.1.1.28, which has been largely detected in several places in Brazil and abroad in the first pandemic wave.<sup>19,20,29,42</sup> P.1 reached a very high frequency in Manaus, a city with fast viral dissemination and a very serious pandemic crisis in January 2021.<sup>5</sup> Also, it carries a set of mutations with important biological significance, mainly in the spike protein (E484K, K417T, and N501Y).<sup>18</sup> In the present study, it was observed that this lineage is responsible for 15% of all Brazilian SARS-CoV-2 WGSs from July 2020 to February 2021. In a separate analysis of 2021, this frequency value increase for more than 30% of all Brazilian WGSs. Furthermore, P.1 is still disseminating in all Brazilian regions as well as in other countries.<sup>5,7,30,39,40</sup>

P.1 and P.2 lineages have become predominant because of the high dissemination in the last months as clearly demonstrated in both skyline plots in this study. Also, the rates of nucleotide and amino acid substitutions were statistically higher in the P.1 lineage than P.2, suggesting a potential for greater transmissibility for the P.1 lineage. It is hypothesized that the several mutations in the RBD of the spike protein helped these new viruses to attach to the receptor and to invade the host cell.<sup>7,8,16,41</sup> Specifically, some amino acid substitutions in the RBD of the spike protein (such as E484K, K417T, and N501Y) raise concerns about their potential to shift the dynamics and public health impact of the current pandemic in the World. These mutations change the intensely glycosylated viral spike (S) and improve the membrane fusion capabilities between the SARS-CoV-2 and the host cell, increasing viral transmissibility, and pathogenicity.<sup>8,16,44</sup> Recent studies have reported that these and other mutations in the SARS-CoV-2 spike gene seem to be improving, even more, the viral fitness. Novel strains and lineages with very high transmissibility and pathogenicity have also emerged in other continents of the World, such as B.1.1.7 (Alpha) in the UK (Europe), B.1.351 (Beta) in South Africa (Africa), and more recently B.1.617 (Delta) in India (Asia).<sup>12,45,46</sup>

In addition, these two Brazilian variants seem to present a clear geographic distribution and route of dissemination in Brazil. P.2 was initially more frequent in the Southeast in October 2020, while P.1 first predominated in the North in December 2020. P.2 lineage showed a consistent increase of WGSs between September and January 2021 in a similar way in different regions (Southeast, South, and Northeast) until the introduction of P.1. On oppose, P.1 appears to have been disseminated from the North to the Midwest and South, before reaching the Southeast and Northeast Brazil in early January 2021. The very fast speed of P.1 dissemination, as demonstrated in the high mutation rates and the exponential growth in the Skyline plot, is strong evidence this lineage is the main responsible for the dramatic scenario of COVID-19 in all countries in the last months. This is further evidenced in some recent reports in different Brazilian regions.<sup>5,7,8,19,20,22,24,41,43</sup> Additionally, a recent epidemiological report from the São Paulo state demonstrated that more than 90% of the SARS-CoV-2 genomes sequenced in this state were P.1.<sup>47</sup> All these findings indicate that these lineages have been extremely important in the establishment of the second wave of the SARS-CoV-2 pandemic in Brazil. Furthermore, North and Southeast regions were fundamental for the emergence and fast spreading of these both SARS-CoV-2 VOIs.<sup>5,7,8,18–20,22,24,41,43</sup>

**FIGURE 4** (A) Time-scaled maximum clade credibility tree from the evolutionary reconstruction by Bayesian analysis of SARS-CoV-2 P.1 (Gamma) lineage whole-genome sequences from Brazil available in GISAID (from December 2020 to February 2021). The time of the most recent common ancestor (tMRCA) is demonstrated in the nodes with significant posterior probabilities ( $\geq 0.95$ ). (B) Bayesian skyline plot (BSP) of SARS-CoV-2 P.1 (Gamma) lineage whole-genome sequences obtained from GISAID. The effective number of infections is reported on the Y-axis. The timeline is reported on the X-axis. The colored area corresponds to the 95% credibility intervals of the highest probability density (95% HPD). The vertical line indicates the 95% lower HPD (dotted) of the tree root. GISAID, Global Initiative on Sharing All Influenza Data; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2



**FIGURE 5** (A) Time-scaled maximum clade credibility tree from the evolutionary reconstruction by Bayesian analysis of SARS-CoV-2 P.2 lineage whole-genome sequences from Brazil available in GISAID (from July 2020 to February 2021). The time of the most recent common ancestor (tMRCA) is demonstrated in the nodes with significant posterior probabilities ( $\geq 0.95$ ). (B) Bayesian skyline plot (BSP) of SARS-CoV-2 P.2 lineage whole-genome sequences obtained from GISAID. The effective number of infections is reported on the Y-axis. The timeline is reported on the X-axis. The colored area corresponds to the 95% credibility intervals of the highest probability density (95% HPD). The vertical line indicates the 95% lower HPD (dotted) of the tree root. GISAID, Global Initiative on Sharing All Influenza Data; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2



Brazil announced COVID-19 as a national public health crisis on February 3, 2020.<sup>48,49</sup> After the development of a national crisis plan and the early establishment of molecular diagnostic offices over Brazil's network of public health laboratories, the country detailed its first confirmed COVID-19 case on February 25, 2020, in a traveler returning to São Paulo from northern Italy.<sup>49,50</sup> After this period, there was an increase in the evolutionary diversity of strains of SARS-CoV-2 circulating in Brazil until July, due to the different processes of introduction of this virus in the country.<sup>50</sup> Since then, an evolutionary stationary phase was observed and it was directly associated with the drastic reduction of a social movement from April to August 2020.<sup>28</sup> However, the gradual reduction in social distance levels and the carelessness with the necessary minimum sanitary practices (use of masks, wash the hands, etc.) associated to some national important events (such as the electoral Brazilian process in October and November 2020, Christmas in December 2020, New-Year celebrations in January 2021, and Carnival in 2021) probably resulted in a new high increase in the SARS-CoV-2 spreading in this period.

Finally, SARS-CoV-2 lineages with different virulence and pathogenicity features are continuously emerging in the COVID-19 pandemic. The continuous monitoring of the most frequent viral clades, lineages, VOIs, and VOCs as well as the study of the specific dynamic evolution processes are really necessary to define public health measures. All this information will be also necessary to develop more appropriate diagnostic tests (as the molecular biology methods) and vaccines for the circulating SARS-CoV-2 lineages. Deeply surveillance of viral transmission at local and global scales and the evaluation of the effect of the different control measures on COVID-19 transmission will offer assistance to decide an ideal mitigation procedure to minimize infections and decrease public health-care demand.

## 5 | CONCLUSION

SARS-CoV-2 P.1 (Gamma) and P.2 lineages were disseminated in the second wave pandemic phase in Brazil: P.1 (from North to Southeast, Northeast, Midwest, and South) and P.2 (from North to Midwest, Southeast, Northeast, and South). The statistical results suggested that strains from P.1 lineage spread mainly between January 3 and January 15, 2021, while P.2 between September 15, 2020 and January 23, 2021. The phylodynamic analyses of these lineages showed a slight rise in the spreading for P.2 and an almost vertical exponential growth for P.1. The continuous monitoring of the most frequent SARS-CoV-2 lineages as well as their specific dynamic evolution processes is now imperative for epidemiologists to define public health measures, diagnostic tests, and vaccines.

## ACKNOWLEDGMENTS

This study was funded by Simbios Biotecnologia and Brazilian Innovation Agencies FINEP and FAPERGS (COVID - 07/2020 Program FINEP - TECNOVA/RS II - 2nd Edition - Economic Subsidy to Innovation, process numbers: 48080.599.26791.12062020). V.R.

Lunge was also financially supported by the National Council for Scientific and Technological Development from Brazil (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico; process numbers 311010/2017-2). J.M. Wolf was further supported by the Coordination for the Improvement of Higher Education Personnel from Brazil (CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Finance Code 001).

## CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

## AUTHOR CONTRIBUTIONS

Jonas M. Wolf and Vagner R. Lunge designed the study. Jonas M. Wolf performed the bioinformatics analyses. Jonas M. Wolf, Gabriela R. Borges, Diéssy Kipper, André F. Streck, and Vagner R. Lunge wrote the first draft of the manuscript and contributed to the literature review and discussion of results. All authors contributed to and have approved the final manuscript.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## REFERENCES

- Hallal PC, Hartwig FP, Horta BL, et al. SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Glob Health*. 2020;8(11):e1390-e1398. doi:10.1016/S2214-109X(20)30387-9
- Castells M, Lopez-Tort F, Colina R, Cristina J. Evidence of increasing diversification of emerging severe acute respiratory syndrome coronavirus 2 strains. *J Med Virol*. 2020;92(10):2165-2172. doi:10.1002/jmv.26018
- Dowd JB, Andriano L, Brazel DM, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc Natl Acad Sci U S A*. 2020;117(18):9696-9698. doi:10.1073/pnas.2004911117
- Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *J Med Virol*. 2020;92(6):675-679. doi:10.1002/jmv.25723
- Faria NR, Claro IM, Candido D, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological*. Accessed June 6, 2021. <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>
- Fujino T, Nomoto H, Kutsuna S, et al. Novel SARS-CoV-2 variant in travelers from Brazil to Japan. *Emerg Infect Dis*. 2021;27(4):1243-1245. doi:10.3201/eid2704.210138
- Martins AF, Zavascki AP, Wink PL, et al. Detection of SARS-CoV-2 lineage P.1 in patients from a region with exponentially increasing hospitalisation rate, February 2021, Rio Grande do Sul, Southern Brazil. *Euro Surveill*. 2021;26(12):2100276. doi:10.2807/1560-7917.ES.2021.26.12.2100276
- Sabino EC, Buss LF, Carvalho MPS, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet*. 2021;397(10273):452-455. doi:10.1016/S0140-6736(21)00183-5
- Laamarti M, Alouane T, Kartti S. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. *PLoS One*. 2020;15(11):e0240345. doi:10.1371/journal.pone.0240345

10. van Dorp L, Acman M, Richard D, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol.* 2020;83:104351. doi:10.1016/j.meegid.2020.104351
11. Volz E, Hill V, McCrone JT, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell.* 2021;184(1):64-75.e11. doi:10.1016/j.cell.2020.11.020
12. Volz E, Mishra S, Chand M, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv.* Published online January 4, 2021. doi:10.1101/2020.12.30.20249034
13. Dearlove B, Lewitus E, Bai H, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc Natl Acad Sci U S A.* 2020;117(38):23652-23662. doi:10.1073/pnas.2008281117
14. Nowak MD, Sordillo EM, Gitman MR, Mondolfi AEP. Coinfection in SARS-CoV-2 infected patients: where are influenza virus and rhinovirus/enterovirus? *J Med Virol.* 2020. 2020 Oct;92:1699-1700. doi:10.1002/jmv.25953
15. Peddu V, Shean RC, Xie H, et al. Metagenomic analysis reveals clinical SARS-CoV-2 infection and bacterial or viral superinfection and colonization. *Clin Chem.* 2020;66(7):966-972. doi:10.1093/clinchem/hvaa106
16. Pango Lineages. Accessed June 6, 2021. [https://cov-lineages.org/global\\_report\\_P.1.html](https://cov-lineages.org/global_report_P.1.html)
17. Altmann DM, Boynton RJ, Beale R. Immunity to SARS-CoV-2 variants of concern. *Science.* 2021;371(6534):1103-1104. doi:10.1126/science.abg7404
18. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. *medRxiv.* Published online March 3, 2021. doi:10.1101/2021.02.26.21252554
19. Naveca F, Nascimento V, Souza V, et al. Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. *Virological.* Accessed June 6, 2021. <https://virological.org/t/phylogenetic-relationship-of-sarscov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585>
20. Nonaka CKV, Franco MM, Gräf T, et al. Genomic evidence of SARS-CoV-2 reinfection involving E484K spike mutation, Brazil. *Emerg Infect Dis.* 2021;27(5):1522-1524. doi:10.3201/eid2705.210191
21. Rambaut A, Loman N, Pybus O, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020. Accessed June 6, 2021. <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>
22. Resende PC, Bezerra JF, Teixeira Vasconcelos RH, et al. Severe acute respiratory syndrome coronavirus 2 P.2 lineage associated with reinfection case, Brazil, June-October 2020. *Emerg Infect Dis.* 2021;27(7):1789-1794. doi:10.3201/eid2707.210401
23. Tegally H, Wilkinson E, Giovanetti M, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv.* 2020. doi:10.1101/2020.12.21.20248640
24. Lamarca AP, de Almeida LGP, Francisco RS, et al. Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil. *medRxiv.* 2021. doi:10.1101/2021.03.21.21253418
25. Weisblum Y, Schmidt F, Zhang F, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife.* 2020;9:e61312. doi:10.7554/eLife.61312
26. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci U S A.* 2020; 117(17):9241-9243. doi:10.1073/pnas.2004999117
27. Mercatelli D, Triboli L, Fornasari E, Ray F, Giorgi FM. Coronapp: a web application to annotate and monitor SARS-CoV-2 mutations. *J Med Virol.* 2021;93(5):3238-3245. doi:10.1002/jmv.26678
28. Wolf JM, Streck AF, Fonseca A, Ikuta N, Simon D, Lunge VR. Dissemination and evolution of SARS-CoV-2 in the early pandemic phase in South America. *J Med Virol.* 2021;93(7):4496-4507. doi:10.1002/jmv.26967
29. Voloch CM, da Silva Francisco R Jr, de Almeida LGP, et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J Virol.* 2021; JVI.00119-21. doi:10.1128/JVI.00119-21
30. Hirotsu Y, Omata M. Discovery of a SARS-CoV-2 variant 1 from the P.1 lineage harboring K417T/E484K/N501Y mutations in Kofu, Japan. *J Infect.* 2021;82(16):276-316. doi:10.1016/j.jinf.2021.03.013
31. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015;1(1):vev003. doi:10.1093/ve/vev003
32. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772-780. doi:10.1093/molbev/mst010
33. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268-274. doi:10.1093/molbev/msu300
34. Minh BQ, Nguyen MA, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 2013;30(5):1188-1195. doi:10.1093/molbev/mst024
35. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016;2(1):vew007. doi:10.1093/ve/vew007
36. Bouckaert R, Vaughan TG, Barido-Sottani J, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2019;15(4):e1006650. doi:10.1371/journal.pcbi.1006650
37. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc.* 1995;90(430):773-795. doi:10.1080/01621459.1995.10476572
38. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67(5):901-904. doi:10.1093/sysbio/syy032
39. Delli Compagni E, Jurisic L, Caporale M, et al. Genome sequences of three SARS-CoV-2 P.1 strains identified from patients returning from Brazil to Italy. *Microbiol Resour Announc.* 2021;10(12):e00177-21. doi:10.1128/MRA.00177-21
40. Maggi F, Novazzi F, Genoni A, et al. Imported SARS-CoV-2 variant P.1 in traveler returning from Brazil to Italy. *Emerging Infect Dis.* 2021;27(4):1249-1251. doi:10.3201/eid2704.210183
41. Tosta S, Giovanetti M, Brandão Nardy V, et al. Active genomic surveillance of SARS-1 CoV-2 suspected cases from recent travelers reveals the circulation of the P1 variant of concern in Bahia state Northeast Brazil. *medRxiv.* Published online March 1, 2021. doi:10.1101/2021.02.25.21252490
42. Francisco RDS Jr, Benites LF, Lamarca AP, et al. Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. *Virus Res.* 2021;296:198345. doi:10.1016/j.virusres.2021.198345
43. Salvato PS, Gregianini TS, Campos AAS, et al. Epidemiological investigation reveals local transmission of SARS-CoV-2 lineage P.1 in Southern Brazil. *doi.* Research Square. Published online March 2, 2021. doi:10.17058/reci.v1i1.16335
44. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 2020; 367(6483):1260-1263. doi:10.1126/science.abb2507

45. Ranjan R, Sharma A, Verma MK. Characterization of the second wave of COVID-19 in India. medRxiv. Accessed 6, 2021. <https://www.medrxiv.org/content/10.1101/2021.04.17.21255665v1>
46. Wibmer CK, Ayres F, Hermanus T, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *bioRxiv*. Published online January 19, 2021. doi:10.1101/2021.01.18.427166
47. Paulo Estadode São. Governo de SP conclui estudo sobre variantes do novo coronavírus no Estado. Accessed June 6, 2021. <https://www.saopaulo.sp.gov.br/noticias-coronavirus/governo-de-sp-conclui-estudo-sobre-variantes-do-novo-coronavirus-no-estado/>
48. Croda J, Oliveira WK, Frutuoso RL, et al. COVID-19 in Brazil: advantages of a socialized unified health system and preparation to contain cases. *Rev Soc Bras Med Trop*. 2020;53:e20200167. doi:10.1590/0037-8682-0167-2020
49. Poterico JA, Mestanza O. Genetic variants and source of introduction of SARS-CoV-2 in South America. *J Med Virol*. 2020;92(10):2139-2145. doi:10.1002/jmv.26001
50. Jesus JG, Sacchi C, Candido DDS, et al. Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev Inst Med Trop Sao Paulo*. 2020;62:e30. doi:10.1590/s1678-9946202062030

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Wolf JM, Kipper D, Borges GR, Streck AF, Lunge VR. Temporal spread and evolution of SARS-CoV-2 in the second pandemic wave in Brazil. *J Med Virol*. 2022;94: 926-936. doi:10.1002/jmv.27371