



OPEN ACCESS

EDITED BY

Swayam Prakash,
University of California, Irvine, United States

REVIEWED BY

Fares Z. Najjar,
Oklahoma State University, United States
Khin Phyu Pyar,
Ministry of Health Myanmar, Myanmar

*CORRESPONDENCE

José Alexandre Felizola Diniz-Filho
✉ diniz@ufg.br

RECEIVED 01 June 2025

ACCEPTED 26 July 2025

PUBLISHED 13 August 2025

CITATION

Diniz-Filho JAF, Nunes R, Targueta CP, dos Santos Braga Ferreira R, Melo-Ximenes A, de Oliveira Dias R, de Curcio JS, Melo e Silva D, Dias e Souza MBdL, Fiaccadori FS, de Paula Silveira Lacerda E, Toscano CM, Rangel TF and Telles MPdC (2025) Temporal dynamics of SARS-CoV-2 phylogenetic diversity in Central Brazil reveals evolutionary shifts among variants of concern during the pandemic.
Front. Microbiol. 16:1639187.
doi: 10.3389/fmicb.2025.1639187

COPYRIGHT

© 2025 Diniz-Filho, Nunes, Targueta, dos Santos Braga Ferreira, Melo-Ximenes, de Oliveira Dias, de Curcio, Melo e Silva, Dias e Souza, Fiaccadori, de Paula Silveira Lacerda, Toscano, Rangel and Telles. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Temporal dynamics of SARS-CoV-2 phylogenetic diversity in Central Brazil reveals evolutionary shifts among variants of concern during the pandemic

José Alexandre Felizola Diniz-Filho^{1*}, Rhewter Nunes², Cintia Pelegrineti Targueta^{3,4}, Ramilla dos Santos Braga Ferreira⁵, Amanda Melo-Ximenes^{3,6}, Renata de Oliveira Dias³, Juliana Santana de Curcio³, Daniela Melo e Silva³, Menira Borges de L. Dias e Souza⁷, Fabíola Souza Fiaccadori⁷, Elisângela de Paula Silveira Lacerda³, Cristiana Maria Toscano⁸, Thiago Fernando Rangel¹ and Mariana Pires de Campos Telles^{3,9}

¹Departamento de Ecologia, Instituto de Ciências Biológicas, Goiânia, Brazil, ²Universidade Estadual de Goiás, Unidade Universitária de Iporá, Iporá, Brazil, ³Departamento de Genética, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil, ⁴Hospital Veterinário, Escola de Veterinária e Zootecnia, Universidade Federal de Goiás, Goiânia, Brazil, ⁵Instituto de Ciências Exatas e Naturais, Universidade Federal de Rondonópolis, Rondonópolis, Brazil, ⁶Programa de Pós-Graduação em Genética and Biologia Molecular, Instituto de Ciências Biológicas, Universidade de Federal de Goiás, Goiânia, Brazil, ⁷Departamento de Biociências e Tecnologia (Microbiologia), Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiás, Brazil, ⁸Departamento de Saúde Coletiva, Instituto de Patologia Tropical e Saúde Pública, Universidade Federal de Goiás, Goiânia, Brazil, ⁹Escola de Ciências Médicas e da Vida, Pontifícia Universidade Católica de Goiás, Goiânia, Brazil

The COVID-19 pandemic has led to substantial health, economic, and social impacts worldwide, and now, after more than 5 years since the start of the pandemic, it is possible to retrospectively evaluate patterns of SARS-CoV-2 spread and its consequences. Here we investigate the temporal dynamics of SARS-CoV-2 phylogenetic diversity in Goiás State, Central Brazil, using genomic data from 8,937 viral sequences obtained from GISAID between March 2020 and October 2024. Phylogenetic diversity was assessed through median pairwise distances (MedPD) and phylogenetic eigenvector regression (PVR) derived from principal coordinate analysis (PCoA) of pairwise distances among sequences. Results show evolutionary shifts associated with the emergence of new variants of concern (VOCs), particularly Gamma and Omicron, corresponding to distinct peaks in phylogenetic diversity through time. The initial rise in MedPD coincided with the Gamma variant's emergence in early 2021, while a more pronounced peak followed the spread of the Omicron variant in late 2021. Although a third peak appeared in late 2023, it was based on smaller sample sizes and did not correspond to a major VOC. Moreover, the temporal dynamics of MedPD tends to mirror the epidemiological characterization of the epidemic over time, including morbidity and mortality, reflecting the impact of vaccination in the disease burden of subsequent variants. The strong phylogenetic signal over time, reflected in the first PCoA axis, highlights the evolutionary trajectory of the virus. This study illustrates how

genomic surveillance provides critical insights into viral diversification and public health responses during pandemics.

KEYWORDS

SARS-CoV2, phylogenetic comparative analyses, mean pairwise distances, phylogenetic eigenvectors, GISAID, VOCs

1 Introduction

The COVID-19 pandemic, associated with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to substantial health, economic, and social impacts worldwide (Castro et al., 2021; Ashmore and Sherwood, 2023; World Health Organization, 2023). During the initial phases of COVID-19 pandemic, in early 2020, significant emphasis was placed on the temporal dynamics of the disease expansion at different geo-political units (e.g., municipalities, states, or countries), especially in the context of estimating rates of disease progression and evaluating likelihood of severe disease leading to hospitalization and deaths, thus guiding health policy decisions (Verity et al., 2020; Almeida et al., 2023; Ferguson, 2020). In a second moment, understanding the geographical and temporal patterns of spread established the baseline to assess the impacts of non-pharmaceutical interventions, such as social distancing measures and mandatory mask use, as well as to support strategic planning and policy making of healthcare service provision (e.g., Britton et al., 2020) and, after 2021, once COVID-19 vaccines became available, of vaccination strategies (Muller et al., 2023; Ulrichs et al., 2024).

However, as expected, SARS-CoV-2 rapidly diversified into multiple lineages (Hill et al., 2022; González-Vázquez and Arenas, 2023), some of which were designated as “Variants of Concern” (VOCs) due to increased transmissibility, immune escape, or severity (e.g., Harvey et al., 2021; Tao et al., 2021; Zhao et al., 2022; Jung et al., 2022; Flores-Alanis et al., 2022; Hussain and Wu, 2024). The most relevant VOCs detected worldwide and in Brazil, which emerged in different countries as accumulated mutations spread freely within communities, were Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529) (Salehi-Vaziri et al., 2022; Carabelli et al., 2023). Identifying these mutation patterns and lineages requires significant investments in DNA sequencing and genomics surveillance and, in Brazil, the interest in such programs began in early 2020 (Candido et al., 2020; Xavier et al., 2020) but it mainly increased after the identification of P1 (Gamma) lineage in Manaus (Sabino et al., 2021; Faria et al., 2021; Souza et al., 2020; Perico et al., 2022), which coincided with a major outbreak peak nationwide in early 2021. This marked the second COVID-19 wave, with the highest increase in case numbers and severe events including hospitalizations and deaths due to higher infection rate (Coutinho et al., 2021). The virus continued to diversify, resulting in the evolution of new VOCs over time (Salehi-Vaziri et al., 2022). Despite the sharp decline in severe disease and deaths resulting from the start of the vaccination rollout, concerns about spike mutations and immune escape continued (Tao et al., 2021; Chaudhari et al., 2022; Chen et al., 2024; Souza et al., 2025).

More than 5 years after the start of the pandemic, it is possible to evaluate patterns of SARS-CoV-2 spread and its consequences. In the context of VOCs evolution, cumulative DNA sequencing data can be used to access patterns of evolutionary diversification and learn

lessons that could inform more effective VOC identification and evolution mapping strategies, providing a basis for a more complex evaluation of the temporal and geographical components of lineage evolution (Shaddick et al., 2023). While tracking VOCs is essential and an effective way to evaluate epidemiological impact of viral evolution, other approaches that address diversification in a continuous temporal dimension could be helpful in avoiding pitfalls related to VOCs identification. Thus, our goal is to apply multiple approaches from phylogenetic comparative analyses to evaluate the temporal spread of new lineages of SARS-CoV-2 in Goiás State, Central Brazil, illustrating how the availability of continuous genomic data can provide more insights into viral diversification, enhance surveillance, and support public health responses and decision-making.

2 Materials and methods

2.1 Studied region

The state of Goiás covers about 340,000 km² in Central/Midwest Brazil, with about 7 million people. A substantial portion of the population lives in Goiânia, the state capital (with a population of about 1.5 million), and in nearby municipalities (with a population of about 3 million in the metropolitan region). The remainder of its population is distributed among 246 municipalities, with populations ranging from fewer than 10,000 people to more than 360,000 people. Moreover, the country's capital, Brasília, is located in an administrative region (the “Federal District”) within the state of Goiás, which increases the region's total population to over than 10 million people and thus placing the state of Goiás among the main hubs for the early spread of SARS-CoV-2. Thus, the regional patterns of viral diversification over time are quite representative of patterns throughout the country and can be used to demonstrate comparative analyses of genomic data during the pandemic. The first cases of COVID-19 in the region were detected on February 26th, 2020 in Brasília, and a few weeks later, the first cases were confirmed in Goiás State. The state government enforced strict social distancing measures when only a few imported cases were reported to the surveillance system. Slowly but surely, COVID-19 cases were confirmed throughout the state. By late May 2020, more than 4,000 cases had been recorded in Goiás, occurring in approximately 60% of its municipalities. An additional 10,000 + cases were confirmed in Brasília. In early 2023, when the pandemic was declared to be over by the World Health Organization (WHO), more than 25,000 deaths had been officially recorded in the state.

At the beginning of the pandemic, the State Agency for Research Support of Goiás (FAPEG, the “Fundação de Amparo à Pesquisa do Estado de Goiás”) issued an emergency call for projects to address the health crisis. One of our team members, MPCT received initial funding to support genomic analyses of SARS-CoV-2 even though the interest

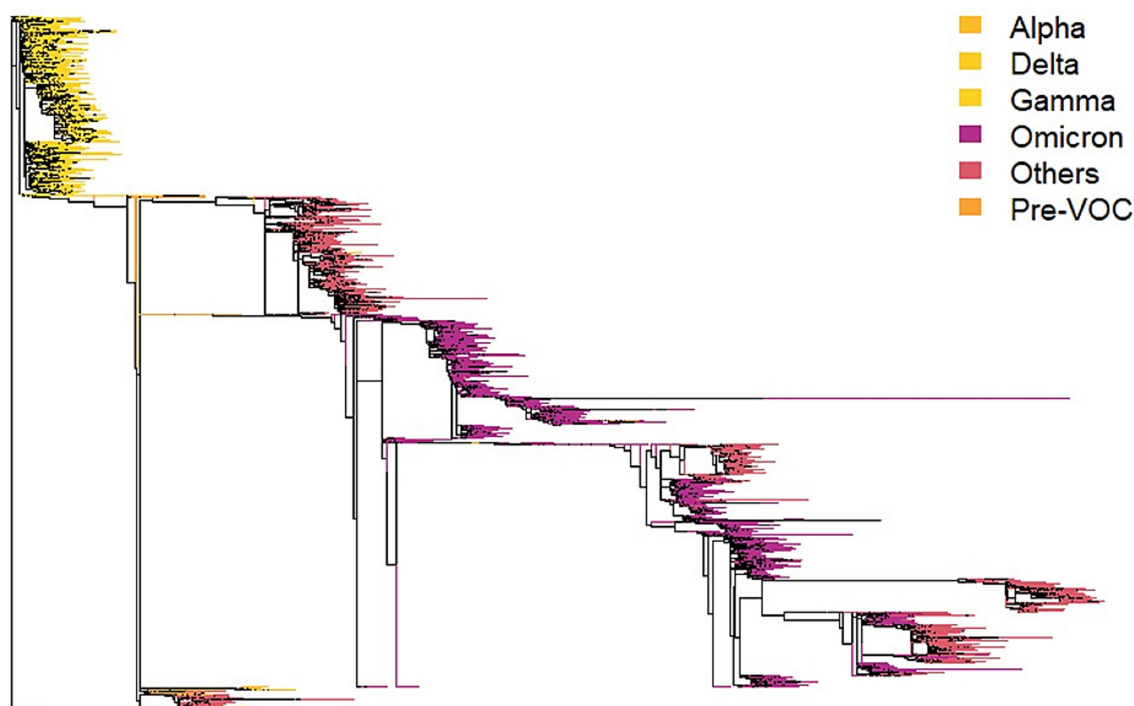


FIGURE 1

Maximum likelihood phylogeny for 8,937 sequences of SARS-CoV-2 from Goiás state, obtained in GISAID. The main categories of VOCs are shown as branch colors, coded as in the insert, along the scale of the first eigenvector extracted from the pairwise distances among sequences (see Figures 5, 6). Pre-VOC in SARS-CoV-2 refers to virus variants that emerged before the Variants of Concern (VOCs). The phylogeny clearly shows the transition from early Alpha, Delta and Gamma lineages (yellow tones) to later Omicron and related lineages, in purple. In our sample, classification of VOCs includes pre-Vocs ($n = 105$), Alpha ($n = 40$), Delta ($n = 5$), Gamma ($n = 2,322$), Omicron ($n = 3,361$) and others (3104). Pre-VOCs refers to Wuhan's original type and first variants like B.1, B.1.1.28, B.1.1.33, whereas "others" include in general more recent variants after Omicron.

in variants of concern (VOCs) intensified following the discovery of the Gamma variant in Manaus in early 2021 (Faria et al., 2021; Sabino et al., 2021; Naveca et al., 2021). From that point onward, the State Secretary of Health of Goiás expanded financial support for genome surveillance efforts, enabling the sequencing of approximately 2,376 genomes using next-generation sequencing (NGS) platforms, including Illumina and Oxford Nanopore technologies. These sequences were subsequently deposited in the GISAID database and combined with additional publicly available sequences for further analysis (see below).

2.2 Data

On April 10, 2025, complete SARS-CoV-2 genome sequences and their associated metadata were downloaded from the GISAID database. Sequences from the state of Goiás, Brazil, were selected and filtered based on the following criteria: (i) complete genomes (defined by GISAID as sequences with >29,000 nucleotides and labeled as high coverage when containing <1% of undefined bases [Ns]); (ii) exclusion of low-coverage entries (>5% Ns); and (iii) inclusion of entries with complete collection dates only. A total of 8,937 complete SARS-CoV-2 genome sequences were aligned using MAFFT v7.503 (Katoh and Standley, 2013). The resulting alignment was then manually curated to identify misaligned regions or problematic sequences. Phylogenetic analysis was performed using the maximum likelihood (ML) method implemented in IQ-TREE 3 (Minh et al., 2020; Wong et al., 2025; see

also Romano et al., 2025). Nucleotide substitution models were evaluated using ModelFinder (Kalyaanamoorthy et al., 2017), and the best-fit model (GTR + F + I + R7) was selected based on the consensus of AIC (Akaike Information Criterion), corrected AIC (cAIC), and BIC (Bayesian Information Criterion). Branch support values were obtained using 1,000 SH-aLRT and ultrafast bootstrap (UFBoot) replicates. The log-likelihood of the consensus tree was -283,306.674.

2.3 Analyses

Pairwise patristic distances between the 8,937 sequences were obtained from the maximum likelihood phylogeny (Figure 1) and identified using the GISAID/Pango classification system.¹ The tips were coded according to the main overall categories of VOCs (i.e., pre-VOC, Alpha, Gamma, Delta, Omicron, "others") to facilitate the understanding of the temporal patterns of lineage diversification. In our sample, classification of VOCs includes pre-Vocs ($n = 105$), Alpha ($n = 40$), Delta ($n = 5$), Gamma ($n = 2,322$), Omicron ($n = 3,361$) and others (3104). Pre-VOCs refers to Wuhan's original type and first variants like B.1, B.1.1.28, B.1.1.33, whereas "others" include in general more recent variants after Omicron.

¹ <https://cov-lineages.org/resources/pangolin.html>

The first goal of our study was to assess phylogenetic diversity over time. To achieve this, we grouped each sequence by month from March 2020 to October 2024. Due to the uneven sampling across months (ranging from 1 to 769 sequences, with a median of 85 sequences) and the skewed distribution of the distances within them, rather than using the more standard Mean Pairwise Distances (MPD) (Swenson, 2014; Tucker et al., 2017), we calculated the median distances and the 95% confidence intervals (MedPD). Even so, the correlation between MPD and median distances is high ($r = 0.821$), and thus the overall temporal patterns from the two metrics are qualitatively similar. For the 3 months without sampling, we performed simple interpolation by means between adjacent months, for graphical purposes only.

We also performed a Principal Coordinate Analysis (PCoA) on the pairwise distance matrix, extracting eigenvectors and eigenvalues to describe the main direction of evolutionary variation among sequences. This served as the basis for a Phylogenetic Eigenvector Regression (PVR; Diniz-Filho et al., 1998, 2012). We then related the first eigenvector to the sampling date (consecutive days from the first sampling), so that the R^2 of this linear correlation estimates the phylogenetic signal of temporal shifts in SARS-CoV-2 diversification. To interpret the temporal patterns in MedPD and PVR, we used the PCoA to ordinate the mean phylogenetic distances among the main VOCs categories.

Finally, we also mapped the mean scores of the first eigenvector at the municipalities of Goiás state, for four different periods of about 1 year, from March 2020 to October 2024, to evaluate whether we could identify geographical patterns in the spread of VOCs along the

pandemic. All phylogenetic comparative analyses were performed in several R packages (R Core Team, 2021) (see [Supplementary material](#) for details of analyses and packages).

3 Results

As expected, the temporal curve of MedPD by month (Figure 2) reveals patterns that match the rise of new VOCs and their impact on the epidemiological characterization of COVID-19. Initially, we observed an increase in median phylogenetic distances up to a peak in January 2021, coinciding with the arrival of the Gamma VOC, which accounted for 60% of our samples. The distances then stabilize and oscillate, with the highest peak arising almost a year later, in November 2021, marked by the emergence of the Omicron VOC, which replaced Delta (with Omicron accounting for 49% of the sample, alongside Delta and other VOCs). There is a third peak in late 2023, but it is based on a relatively small sample size and does not contain any VOCs. Considering the period in which the samples were obtained, Gamma and Omicron are the predominant VOCs, with 26 and 38% of the samples, respectively.

The patterns in Figure 2 and the peaks in MedPD are understandable when considering the PCoA of the mean pairwise distances among VOCs (Figure 3), in which Omicron is indeed the most divergent lineage of SARS-CoV-2, which explains the highest peak in late 2021. On the other hand, Alpha and Gamma cluster closely, as observed in the phylogeny in Figure 1. Thus, although the arrival of Gamma represented a major threat and drove the largest

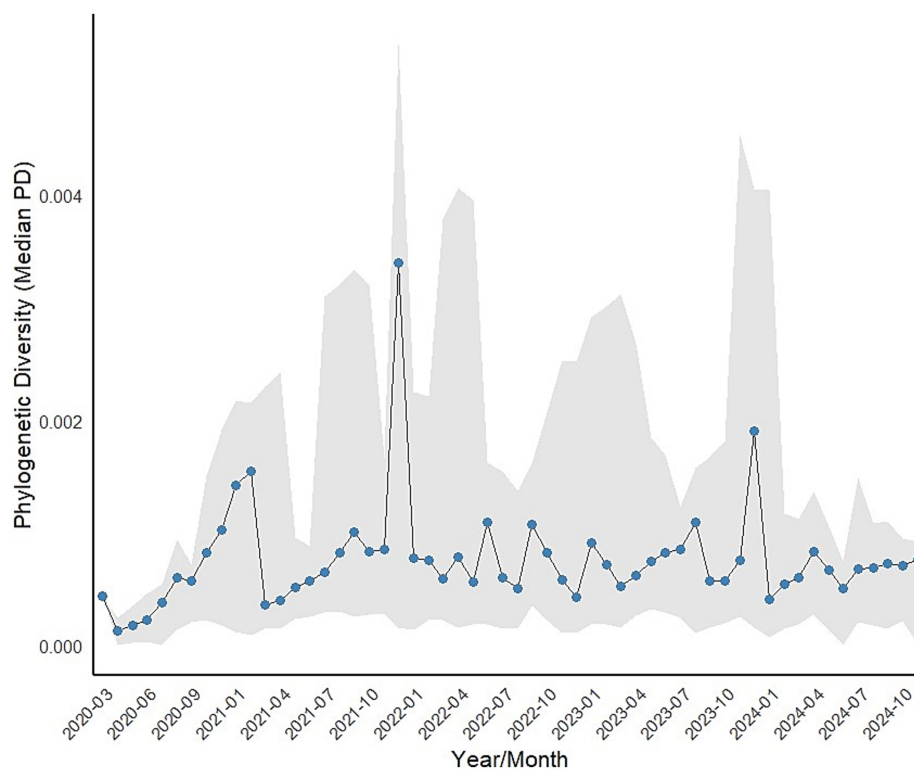


FIGURE 2

Temporal trend in median pairwise phylogenetic distances (MedPD) of SARS-COV-2 from March 2020 to October 2024, with gray shadow showing the 95% confidence interval around each MedPD.

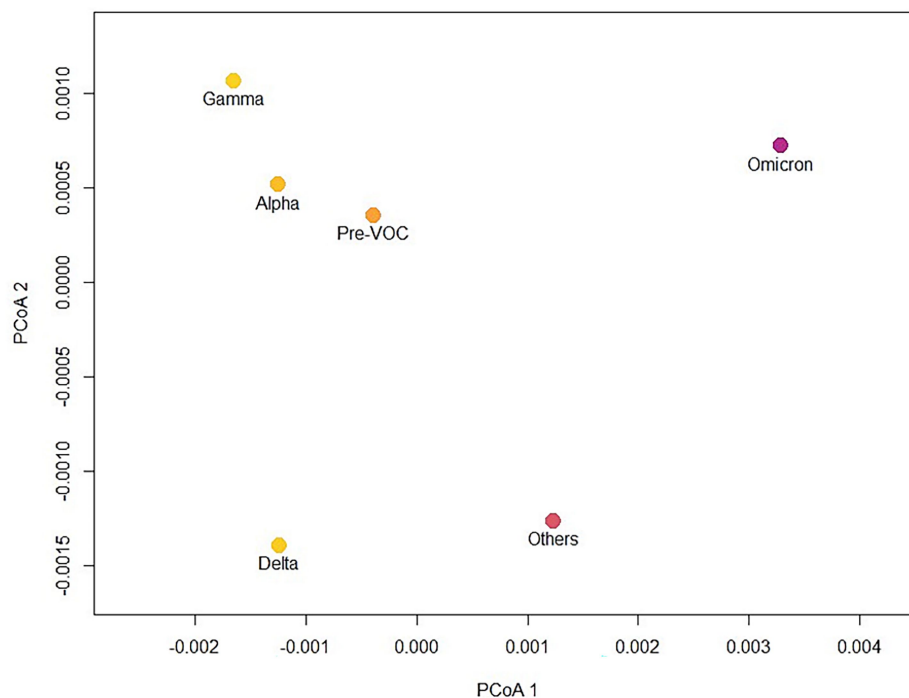


FIGURE 3

Principal coordinates analysis (PCoA) of mean pairwise phylogenetic distances among SARS-CoV-2 variants of concern (VOCs), highlighting the divergence of Omicron (explaining the peak in late 2021 shown in Figure 2) and the similarity of Alpha and Gamma which leads to a smaller peak of MedPD in early 2021.

peak in hospitalizations and deaths in early 2021, it does not appear as a high peak in phylogenetic diversity. Notice, however, that sequences identified as Omicron are also quite variable and have a larger variance along the first principal coordinate of phylogenetic distances (see Figure 5), which may also explain the higher peak in MedPD.

It is also important to evaluate the relationship between MedPD and the epidemiological dynamics of COVID-19, especially the number of deaths. COVID-19 mortality data obtained from publicly available information systems of the Goiás State Health Department and MedPD over time are presented in Figure 4. While the relationship between the two series is not strictly linear or monotonic, and a simple correlation between the series does not yield any significant results, it may provide interesting insights for future surveillance programs regarding delayed peaks in the number of deaths and the indirect effects of vaccination.

The first eigenvector from the pairwise phylogenetic distances among the 8,937 sequences explains 79.9% of the distances variance, making it an adequate synthesis of the phylogenetic patterns in the data. Indeed, it is possible to observe the replacement of the VOCs over time along this axis (Figure 5). The phylogenetic signal in days since the start of the pandemic was equal to $R^2 = 0.64\%$.

Note also that the maps of the mean scores of the first principal coordinate of the pairwise phylogenetic distances show that in the sample used, during the first year of the pandemic (March 2020 up to February 2021) the Alpha, Gamma, and Delta lineages were sampled only in a few municipalities (Figure 6). However, in the second year, when DNA sequencing strategy increased, the Omicron group and its variants were already widespread in Goiás, and in the last 2 years of

the pandemic, they were replaced by other VOCs at the more negative extreme of the first eigenvector. Figure 6 reveals that in 2021, the geographic distribution of SARS-CoV-2 lineages was predominant across all state of Goiás. However, the northern and northeastern regions exhibited a lower impact in terms of the geographic spread of the lineages in all years, which may be associated with lower population density in those areas.

4 Discussion

The COVID-19 pandemic presented several challenges for researchers, epidemiologists, and policymakers, as decisions had to be made despite the significant uncertainty regarding viral transmission patterns and their drivers. Moreover, when VOCs started to emerge, the situation became even more complex. The focus was on evaluating how the evolution of new lineages would allow immune escape and reinfection, and whether the available vaccines would remain effective (Carabelli et al., 2023; Bian et al., 2022). Now, more than five years after the start of the pandemic, with more consolidated data and a better understanding, it is possible to evaluate patterns and determine how alternative methodologies could be helpful for surveillance programs, particularly those with access to high-throughput sequencing technologies.

When considering VOCs, the idea has been to isolate them and check for adaptive shifts, mainly in the spike protein sequence, that could increase the transmission or infection rates of new lineages. Although classifying emerging lineages can be useful and help understand epidemiological curves of cases, hospitalizations, and

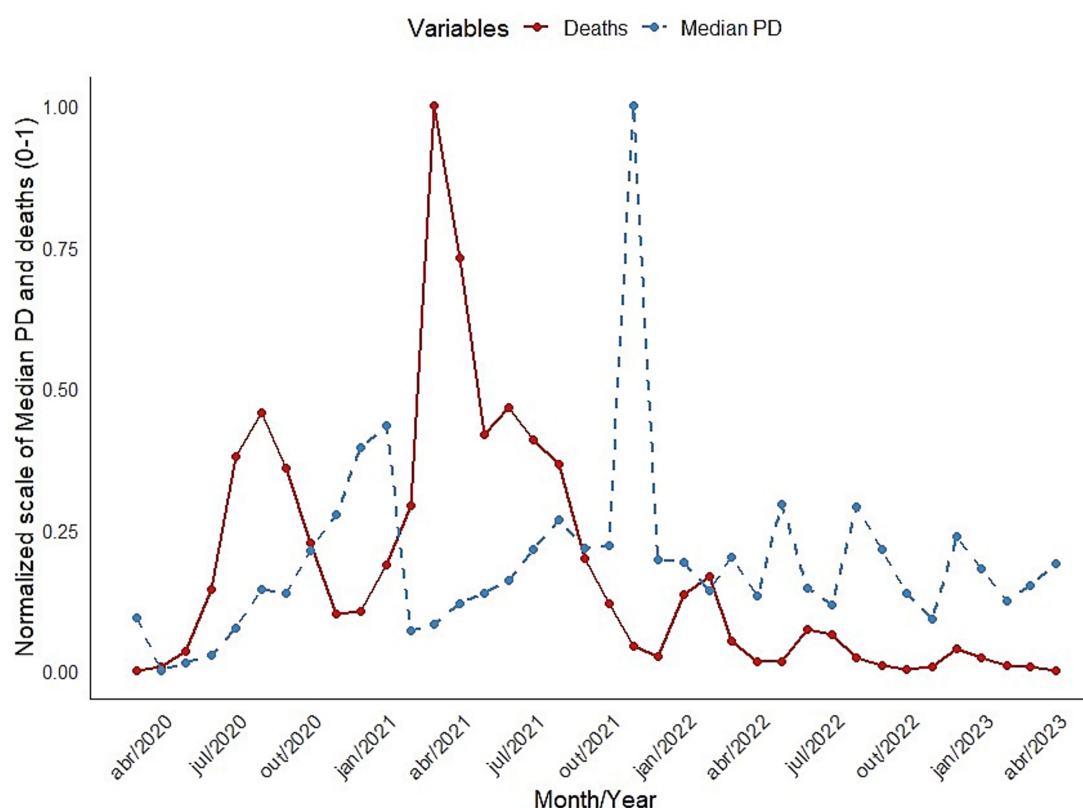


FIGURE 4

Temporal dynamics of MedPD (dashed blue line) and number of deaths (red line) in Goiás State, from March 2020 to April 2023, close to the end of the pandemic. The two variables were normalized to range between 0 and 1 for visualization purposes, but the death peak was equal to 4,175 deaths in March 2021.

deaths, it is an inadequate description of the diversification process in SARS-CoV-2. In fact, SARS-CoV-2 strains carrying point mutations in the spike gene emerged in the Brazilian state of Amazonas, likely due to viral spread from other countries. These mutations suggest the action of evolutionary pressures acting specifically on this region of the viral genome (Naveca et al., 2021).

Here, our focus is to shift from a more “discrete” view of this process, in which lineages are identified and monitored, to a more continuous approach, in which phylogenetic diversity is monitored using an established framework of phylogenetic comparative analyses and community phylogenetics (Tucker et al., 2017; Swenson, 2014). Ideally, these metrics will provide a more explicit pattern for genomic surveillance over time. Rather than monitoring the emergence of a new VOC, which depends on taxonomic issues, a more continuous phylogenetic diversity estimate will be used (measured by pairwise distances among samples within a given time period) to anticipate potential threats due to new VOCs.

The temporal patterns of MedPD in the state of Goiás reveal the usefulness of evaluating the diversification dynamics of SARS-CoV-2 and its VOCs, providing a simple approach to genomic surveillance, allowing it to discuss them in distinct ways. This study provides a retrospective overview of the distribution of SARS-CoV-2 variants in the 5 years after the end of the COVID-19 pandemic, highlighting the importance of accumulating genomic data for robust analytical approaches. Notably, this research group has contributed since the early stages of the pandemic, particularly through the work of

Targueta et al. (2022), who reported the identification of the Delta variant and the optimization of the resequencing methodology. The study demonstrated the predominance of the Alpha and Gamma variants in the state of Goiás between November 2020 and July 2021, based on the real-time analysis of 318 samples.

In general, it is possible to observe from Figure 2 that there are cycles of MedPD, in which the increase in diversity leads and coincides with peaks in epidemiological events and stabilizes when a given VOC becomes dominant, i.e., when diversity is reduced. These peaks in MedPD are also evident when considering the divergence among VOC lineages. For instance, the PCoA analysis of the mean pairwise distances among VOCs (Figure 3) shows that Omicron is indeed the most divergent lineage, explaining the highest peak in late 2021. Some of the sequence variations in Omicron VOC are also related to the epidemiological characteristics of this variant, making it more transmissible (mutations increasing affinity of the ACE2 receptor) (Chatterjee et al., 2023; Planas et al., 2024), but were less lethal. It is important to note that the expansion of vaccination also reduced mortality during the increase in the number of Omicron cases. Indeed, the Omicron appears isolated in the bidimensional ordination space with the first two principal coordinates. The Omicron variant has the highest number of mutations particularly in the spike protein, compared to other VOCs (Parsons and Acharya, 2023). This represents a substantial increase in the number of genetic alterations compared to its predecessors, reflecting a high mutation rate, the emergence of multiple sublineages, immune evasion capacity and increased

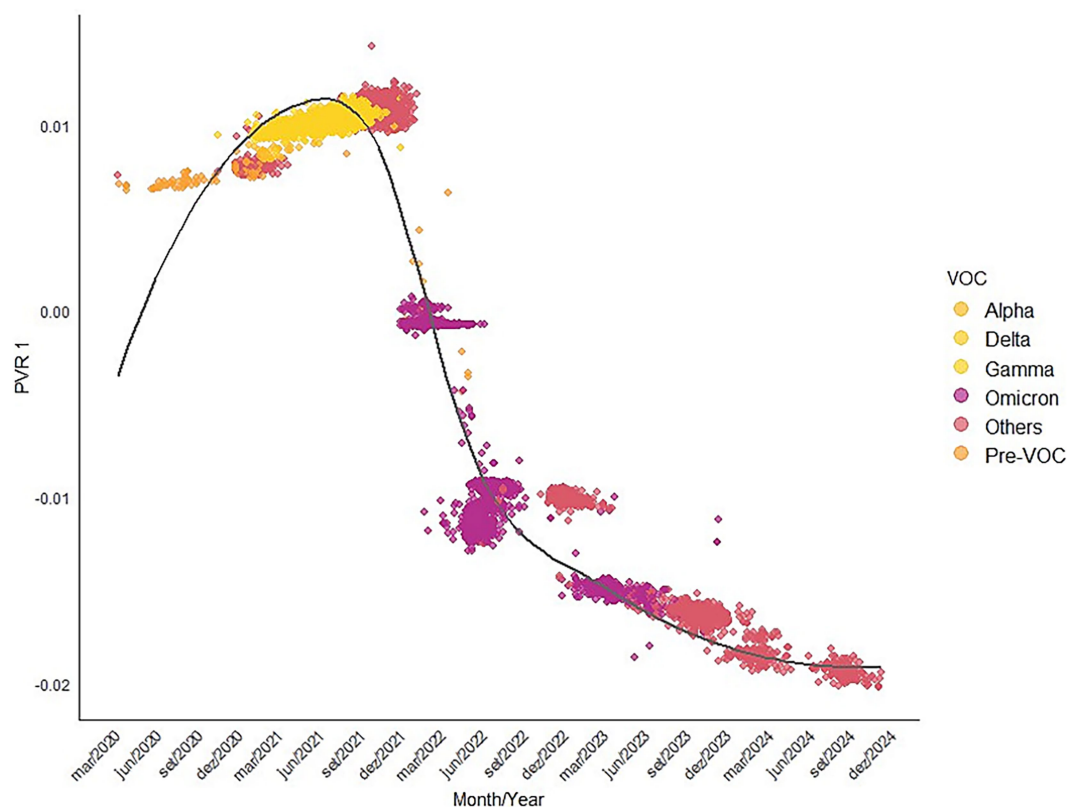


FIGURE 5

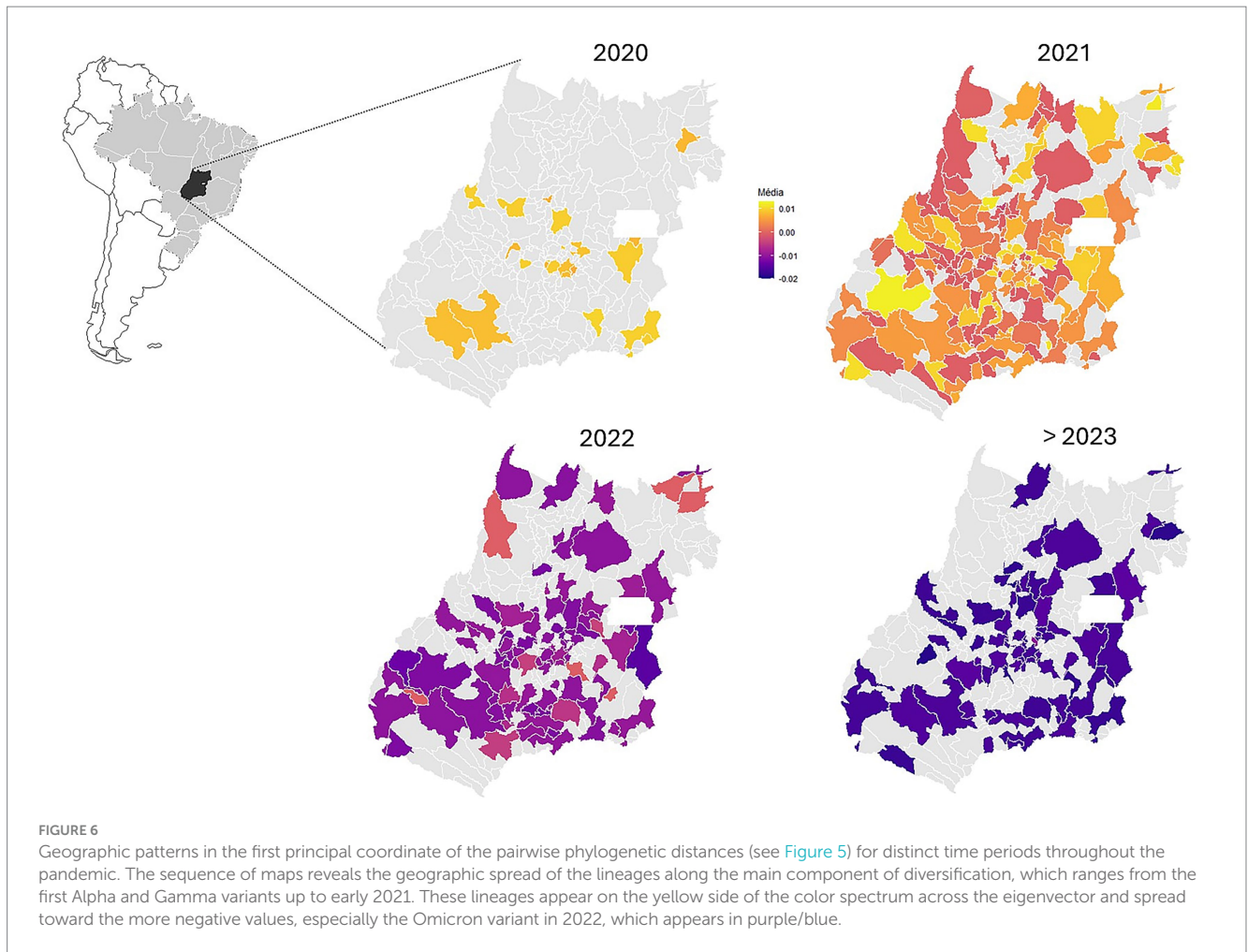
Temporal patterns in the first principal coordinate of the pairwise phylogenetic distances, mapping the VOCs of each sample through time. This first component explains 79.9% of the phylogenetic structure, and the correlation with days since the beginning of the pandemic was equal to 0.8.

transmissibility (Hill et al., 2022). Pre-VOC lineages are centrally located in this space, reflecting their ancestral status, with Alpha and Gamma clustering closely. Thus, although the emergence of Gamma represented a major threat and drove the largest peak in hospitalization and death in early 2021, it did not generate a high peak due to its phylogenetic similarity with Alpha, which dominated the first epidemiological peak in mid-2020.

Another interesting aspect of the temporal patterns in MedPD is their relationship with epidemiological dynamics, especially with the number of COVID-19 deaths, as shown in the Figure 4. Due to the distinct factors involved in the time series of deaths, which create much more complex dynamics, a simple correlation between the series of MedPD and the monthly number of deaths, even with a time lag of 1 or 2 months, does not yield any significant results. However, a visual inspection of the series over time reveals that part of the failure of standard correlation analyses may be due to non-stationary processes underlying the patterns. Indeed, visually inspecting the two series provides useful information for further surveillance programs. For instance, the first peak in deaths in July 2020 does not appear to be related to a new VOC, despite the continuous increase in MedPD that lead to the first peak in January 2021. This peak was followed by another peak approximately 1 month later. The monthly death rate then decreased throughout 2021. However, the highest peak in MedPD due to the Omicron variant was not followed by another mortality peak because, by that time, COVID-19 vaccines had been made available, and mass vaccination campaigns had taken place, reaching in high coverage in Goiás and protection the population

against severe disease and death. Note that even if an effective vaccination program breaks the correlation between the rise of new VOCs and epidemiological parameters, there seems to be a small peak in mortality again about 1 month after the Omicron peak. After that, mortality gradually decreased toward negligible figures at the population level in early 2023 (e.g., Li et al., 2024).

Another new approach proposed here is the PVR (Phylogenetic Eigenvector Regression; Diniz-Filho et al., 1998, 2012) of temporal changes in lineages. The overall idea of PVR is to extract eigenvectors from a phylogenetic distance matrix that can be used as predictors in a regression model. This allows one to describe phylogenetic patterns (signal) in the data and account for an inflated Type I error rate in models due to autocorrelation. In the case of SARS-CoV-2 lineages, for example, the first principal coordinate of the phylogenetic distances explains 79.9% of the phylogenetic structure in the data, thus, it is possible to represent most of the phylogenetic structure along a single axis. When this axis is correlated with the successive days on which sequences appear, its explanatory power is 64%. This indicates a high temporal signal in phylogenetic diversification and, indeed, it is possible to map how emerging VOCs with high prevalence, such as Gamma and Omicron, tend to appear gradually in time. From a more technical perspective within the PVR framework, it is interesting to note that the 64% explanatory power of such temporal patterns is lower than that of the axis itself. This indicates a departure from purely neutral lineage diversification (see Diniz-Filho et al., 2012, 2015). As shown in Figure 5, the relationship between the first axis and time is nonlinear. This departure may be due



to nonstationarity issues and diversification constraints, in which lineages tend to evolve slower than expected by time since divergence. This is indeed expected if non-neutral, adaptive evolution occurs by creating constraints in diversification, for instance, when a given lineage increases its prevalence and, in some way, becomes dominant (which is, by definition, the case of VOCs). This may also be related to sampling bias, as there is a concentration of samples in periods when particular VOCs, particularly Omicron, were more common. Thus, further investigations with larger samples covering more homogeneously the entire period of the pandemic may shed light on how the PVR framework can be used to identify adaptive, non-neutral, evolution of variants that would have the potential to become VOCs.

Our study provides a detailed temporal analysis of the phylogenetic diversity of SARS-CoV-2 in Goiás, revealing how diversification matches key epidemiological milestones during the COVID-19 pandemic, showing peaks in median phylogenetic distances by month and phylogenetic eigenvector aligned with the emergence of major VOCs, particularly before widespread vaccination. Our findings highlight the critical importance of continuous genomic surveillance for understanding the diversification dynamics of SARS-CoV-2 and in guiding evidence-based public health strategies, particularly in the context of epidemic-prone diseases. Further studies would expand the reasoning developed here by evaluating the synchronicity of phylogenetic metrics at multiple spatial scales (among Brazilian States or countries worldwide),

evaluating how phylobetadiversity (i.e., [Graham and Fine, 2008](#)) would track waves of distinct lineages in an explicit geographic context. Although the extensive genomic dataset provides robust insights, sampling biases and variations in sequencing intensity over time should be considered when interpreting the results. Hopefully, the alternative proposed here can complement traditional epidemiological analyses based on classifying and monitoring VOCs in discrete groups, offering critical tools for monitoring ongoing and future viral epidemic threats. Future studies integrating genomic, clinical, and vaccination data could further clarify how to apply these methods, especially the PVR, to assess the complex dynamics between SARS-CoV-2 evolution and COVID-19 severity load.

Data availability statement

The datasets presented in this study can be found in GISAID Platform, and R scripts used for phylogenetic comparative analyses can be found in the [Supplementary material](#).

Author contributions

JD-F: Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing, RN:

Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. CPT: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. RS: Data curation, Methodology, Writing – original draft, Writing – review & editing. AM-X: Data curation, Methodology, Writing – original draft, Writing – review & editing. RO: Formal analysis, Methodology, Writing – original draft, Writing – review & editing. JC: Data curation, Methodology, Writing – original draft, Writing – review & editing. DM: Data curation, Methodology, Writing – original draft, Writing – review & editing. MD: Data curation, Methodology, Writing – original draft, Writing – review & editing. FF: Data curation, Methodology, Writing – original draft, Writing – review & editing. EP: Data curation, Methodology, Writing – original draft, Writing – review & editing. CMT: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. TR: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. MT: Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was developed in the context of the project “Mapeamento das variações genéticas do Coronavírus (SARS-COV-2) em Goiás” supported by FAPEG (proc. 202010267000278), project “Modelagem da dinâmica de transmissão do SARS-CoV-2 no Brasil: Ciência em tempo real para subsidiar gestores na tomada de decisão baseada em evidências” (CNPq proc. 402834/2020) and the National Institute of Science and Technology (INCT) in Ecology, Evolution, and Biodiversity Conservation funded by CNPq (grants 465610/2014-5 and 409197/2024-6) and FAPEG (grant 201810267000023). Work by JAFD-F, RFT, FSC, and CMT is also supported by the “Centro de Excelência em Tecnologia e Inovação em Saúde” (“CETI Saúde”), supported by FAPEG (proc. 202410267000884; agreement 07/2024 FAPEG/UFG/FUNAPE).

References

- Almeida, B. G., Simon, L. M., Bagattini, A. M., Machado, Q., da Rosa, M., Borges, M. E., et al. (2023). Dynamic transmission modeling of COVID-19 to support decision-making in Brazil: a scoping review in the pre-vaccine era. *Front. Glob. Public Health* 3:e0002679. doi: 10.1371/journal.pgph.0002679
- Ashmore, P., and Sherwood, E. (2023). An overview of COVID-19 global epidemiology and discussion of potential drivers of variable global pandemic impacts. *J. Antimicrob. Chemother.* 78, ii2–ii11. doi: 10.1093/jac/dkad311
- Bian, L., Liu, J., Gao, F., Gao, Q., He, Q., Mao, Q., et al. (2022). Research progress on vaccine efficacy against SARS-CoV-2 variants of concern. *Hum. Vaccin. Immunother.* 18:2057161. doi: 10.1080/21645515.2022.2057161
- Britton, T., Ball, F., and Trapman, P. (2020). A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* 369, 846–849. doi: 10.1126/science.abc6810
- Candido, D. S., Claro, I. M., Souza, W. M., Moreira, F. R., Dellicour, S., Mellan, T., et al. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369, 1255–1260. doi: 10.1126/science.abd2161
- Carabelli, A. M., Peacock, T. P., Thorne, L. G., Harvey, W. T., Hughes, J., COVID-19 Genomics UK Consortium et al. (2023). SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat. Rev. Microbiol.* 21, 162–177. doi: 10.1038/s41579-022-00841-7
- Castro, M. C., Kim, S., Barberia, L., Ribeiro, A. F., Gurzenda, S., Ribeiro, K. B., et al. (2021). Spatiotemporal pattern of COVID-19 spread in Brazil. *Science* 372, 821–826. doi: 10.1126/science.abh1558
- Chatterjee, S., Bhattacharya, M., Nag, S., Dhama, K., and Chakraborty, C. (2023). A detailed overview of SARS-CoV-2 omicron: its sub-variants, mutations and pathophysiology, clinical characteristics, immunological landscape, immune escape, and therapies. *Viruses* 15:167. doi: 10.3390/v15010167
- Chaudhari, A. M., Joshi, M., Kumar, D., Patel, A., Lokhande, K. B., Krishnan, A., et al. (2022). Evaluation of immune evasion in SARS-CoV-2 Delta and omicron variants. *Comput. Struct. Biotechnol. J.* 20, 4501–4516. doi: 10.1016/j.csbj.2022.08.010
- Chen, L., He, Y., Liu, H., Shang, Y., and Guo, G. (2024). Potential immune evasion of the severe acute respiratory syndrome coronavirus 2 omicron variants. *Front. Immunol.* 15:1339660. doi: 10.3389/fimmu.2024.1339660
- Coutinho, R. M., Marquitti, F. M. D., Ferreira, L. S., Borges, M. E., da Silva, R. L. P., Canton, O., et al. (2021). Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. *Commun. Med. (Lond.)* 1:48. doi: 10.1038/s43856-021-00048-6
- Diniz-Filho, J. A. F., Alves, D. M. C. C., Villalobos, F., Sakamoto, M., Brusatte, S. L., and Bini, L. M. (2015). Phylogenetic eigenvectors and nonstationarity in the evolution of theropod dinosaur skulls. *J. Evol. Biol.* 28, 1410–1416. doi: 10.1111/jeb.12660
- Diniz-Filho, J. A. F., Rangel, T. F., Santos, T., and Bini, L. M. (2012). Exploring patterns of interspecific variation in quantitative traits using sequential phylogenetic eigenvector regressions. *Evolution* 66, 1079–1090. doi: 10.1111/j.1558-5646.2011.01499.x
- Diniz-Filho, J. A. F., Sant’Ana, C. E. R., and Bini, L. M. (1998). An eigenvector method for estimating phylogenetic inertia. *Evolution* 52, 1247–1262. doi: 10.1111/j.1558-5646.1998.tb02006.x

Acknowledgments

The authors are deeply grateful to the members of the sequencing network established with the support of the “Mulheres do Brasil” group, “Magalu,” and Mendelics Genomic Analysis for support in implementing Nanopore sequencing technology, and to all the researchers and organizations collaborating to maintain and share SARS-CoV-2 genomic data on the GISAID Platform.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2025.1639187/full#supplementary-material>

- Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. S., Mishra, S., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372, 815–821. doi: 10.1126/science.abb2644
- Ferguson, N. M. (2020). Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial Coll. Rep* 16:482. doi: 10.25561/77482
- Flores-Alanis, A., Delgado, G., Espinosa-Camacho, L. F., Rodríguez-Gómez, F., Cruz-Rangel, A., Sandner-Miranda, L., et al. (2022). Two years of evolutionary dynamics of SARS-CoV-2 in Mexico, with emphasis on the variants of concern. *Front. Microbiol.* 13:886585. doi: 10.3389/fmicb.2022.886585
- González-Vázquez, L. D., and Arenas, M. (2023). Molecular evolution of SARS-CoV-2 during the COVID-19 pandemic. *Genes* 14:407. doi: 10.3390/genes14020407
- Graham, C. H., and Fine, P. V. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecol. Lett.* 11, 1265–1277. doi: 10.1111/j.1461-0248.2008.01256.x
- Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., et al. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424. doi: 10.1038/s41579-021-00573-0
- Hill, V., Du Plessis, L., Peacock, T. P., Aggarwal, D., Colquhoun, R., Carabelli, A. M., et al. (2022). The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol.* 8:veac080. doi: 10.1093/ve/veac080
- Hussain, B., and Wu, C. (2024). Evolutionary and phylogenetic dynamics of SARS-CoV-2 variants. *Viruses* 16:907. doi: 10.3390/v16060907
- Jung, C., Kmiec, D., Koepke, L., Zech, F., Jacob, T., Sparrer, K. M., et al. (2022). Omicron: what makes the latest SARS-CoV-2 variant of concern so concerning? *J. Virol.* 96:e02077-21. doi: 10.1128/jvi.02077-21
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Li, S. L., Prete, C. A., Zarebski, A. E., de Souza Santos, A. A., Sabino, E. C., Nascimento, V. H., et al. (2024). The Brazilian COVID-19 vaccination campaign: a modelling analysis of sociodemographic factors on uptake. *BMJ Open* 14:e076354. doi: 10.1136/bmjopen-2023-076354
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Muller, G. C., Ferreira, L. S., Mesias Campos, F. E., Borges, M. E., Almeida, G. B., Poloni, S., et al. (2023). Modeling the impact of child vaccination (5–11 y) on overall COVID-19 related hospitalizations and mortality in a context of omicron variant predominance and different vaccination coverage paces in Brazil. *Lancet Reg. Health Am.* 17:100396. doi: 10.1016/j.lana.2022.100396
- Naveca, F. G., Nascimento, V., De Souza, V. C., Corado, A. D., Nascimento, F., Silva, G., et al. (2021). 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
- Parsons, R. J., and Acharya, P. (2023). Evolution of the SARS-CoV-2 omicron spike. *Cell Rep.* 42:113444. doi: 10.1016/j.celrep.2023.113444
- Perico, C. P., De Pierri, C. R., Neto, G. P., Fernandes, D. R., Pedrosa, F. O., de Souza, E. M., et al. (2022). Genomic landscape of the SARS-CoV-2 pandemic in Brazil suggests an external P.1 variant origin. *Front. Microbiol.* 13:1037455. doi: 10.3389/fmicb.2022.1037455
- Planas, D., Staropoli, I., Michel, V., Lemoine, F., Donati, F., Prot, M., et al. (2024). Distinct evolution of SARS-CoV-2 omicron XBB and BA.2.86/JN.1 lineages combining increased fitness and antibody evasion. *Nat. Commun.* 15, 2254–2217. doi: 10.1038/s41467-024-46490-7
- R Core Team (2021). R: A language and environment for statistical computing. Vienna, Austria: R Found. Stat. Comput.
- Romano, G., Ferrari, A., and Baldanti, F. (2025). Phylogenetic and epidemiological insights into centenarians' resilience to COVID-19: exploring the role of past coronavirus pandemics. *Front. Microbiol.* 16:1572763. doi: 10.3389/fmicb.2025.1572763
- Sabino, E. C., Buss, L. F., Carvalho, M. P. S., Prete, C. A. Jr., Crispim, M. A. E., Fraiji, N. A., et al. (2021). Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* 397, 452–455. doi: 10.1016/S0140-6736(21)00183-5
- Salehi-Vaziri, M., Fazlalipour, M., Seyed Khorrami, S. M., Azadmanesh, K., Pouriaeyali, M. H., Jalali, T., et al. (2022). The ins and outs of SARS-CoV-2 variants of concern (VOCs). *Arch. Virol.* 167, 327–344. doi: 10.1007/s00705-022-05365-2
- Shaddick, G., Zidek, J. V., and Schmidt, A. M. (2023). Spatio-temporal methods in environmental epidemiology with R. 2nd Edn. London: Chapman & Hall/CRC Press.
- Souza, W. M., Buss, L. F., Candido, D. D. S., Carrera, J. P., Li, S., Zarebski, A. E., et al. (2020). Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat. Hum. Behav.* 4, 856–865. doi: 10.1038/s41562-020-0928-4
- Souza, U. J. B., Spilki, F. R., Tanuri, A., Rohe, P. M., and Campos, F. S. (2025). Two years of SARS-CoV-2 omicron genomic evolution in Brazil (2022–2024): subvariant tracking and assessment of regional sequencing efforts. *Viruses* 17:64. doi: 10.3390/v17010064
- Swenson, N. G. (2014). Functional and phylogenetic ecology in R. New York, NY: Springer.
- Tao, K., Tzou, P. L., Nouhin, J., Gupta, R. K., de Oliveira, T., Kosakovsky Pond, S. L., et al. (2021). The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* 22, 757–773. doi: 10.1038/s41576-021-00408-x
- Targueta, C. P., Braga-Ferreira, R. S., Melo, A. A., Curcio, J. S., Nunes, R., Dias, R. O., et al. (2022). Optimization of Illumina AmpliSeq protocol for SARS-CoV-2 and detection of circulating variants in Goiás state, Brazil from November 2020 to July 2021. *Genet. Mol. Res.* 21, 1–15. doi: 10.4238/gmr19041
- Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., et al. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev.* 92, 698–715. doi: 10.1111/brv.12252
- Ulrichs, T., Rolland, M., Wu, J., Nunes, M. C., el Guerche-Séblain, C., and Chit, A. (2024). Changing epidemiology of COVID-19: potential future impact on vaccines and vaccination strategies. *Expert Rev. Vaccines* 23, 510–522. doi: 10.1080/14760584.2024.2346589
- Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., et al. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* 20, 669–677. doi: 10.1016/S1473-3099(20)30243-7
- Wong, T. K. F., Ly-Trong, N., Ren, H., Baños, H., Roger, A. J., Susko, E., et al. (2025). IQ-TREE 3: phylogenomic inference software using complex evolutionary models. *EcoEvoRxiv*. doi: 10.32942/X2P62N
- World Health Organization (2023). Coronavirus dashboard. Geneva: WHO.
- Xavier, J., Giovanetti, M., Adelino, T., Fonseca, V., Barbosa da Costa, A. V., Ribeiro, A. A., et al. (2020). The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *Emerg. Microbes Infect.* 9, 1824–1834. doi: 10.1080/22221751.2020.1803146
- Zhao, W., Liu, L., Ma, Z., and Yang, J. (2022). covSampler: a subsampling method with balanced genetic diversity for large-scale SARS-CoV-2 genome data sets. *GigaScience* 11:giac093. doi: 10.1093/gigascience/giac093