Higher frequency of interstate over international transmission chains of SARS-CoV-2 virus at the Rio Grande do Sul - Brazil state borders

Filipe Zimmer Dezordi [a,b,1], José Valter Joaquim Silva Júnior [c,d,e,f,g,h,1], Terimar Facin Ruoso [i,1], Angela Giovana Batista [i,j], Pedro Mesquita Fonseca [i], Larissa Paim Bernardo [k], Richard Steiner Salvato [l], Tatiana Schäffer Gregianini [l], Thaísa Regina Rocha Lopes [c,g], Eduardo Furtado Flores [c], Rudi Weiblen [c], Patrícia Chaves Brites [m], Mônica de Medeiros Silva [n], João Batista Teixeira da Rocha [n], Gustavo de Lima Barbosa [o], Lais Ceschini Machado [a], Alexandre Freitas da Silva [a,b], Marcelo Henrique Santos Paiva [a,p], Matheus Filgueira Bezerra [q], Tulio de Lima Campos [b], Tiago Gräf [r], Daniel Angelo Sganzerla Graichen [s,2], Elgion Lucio da Silva Loreto [m,2], Gabriel da Luz Wallau [a,b,t,2,*], Fiocruz Genomic Network

[a] Departamento de Entomologia, Instituto Aggeu Magalhães (IAM)-Fundação Oswaldo Cruz-FIOCRUZ, Recife Pernambuco, 50670-420, Brazil
[b] Núcleo de Bioinformática (NBI), Instituto Aggeu Magalhães (IAM), FIOCRUZ-Pernambuco, Recife, Pernambuco, 50670-420, Brazil
[c] Setor de Virologia, Departamento de Medicina Veterinária Preventiva, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, 97105-900, Brazil
[d] Setor de Virologia, Instituto Keizo Asami, Universidade Federal de Pernambuco, Pernambuco, 50670-901, Brazil
[e] Laboratório NB3 de Neuroimunologia, Universidade Federal de Santa Maria, Rio Grande do Sul, 97105-900, Brazil
[f] Departamento de Microbiologia e Parasitologia, Universidade Federal de Santa Maria, Rio Grande do Sul, 97105-900, Brazil
[g] Programa de Pós-graduação em Medicina Veterinária, Universidade Federal de Santa Maria, Rio Grande do Sul, 97105-900, Brazil
[h] Programa de Pós-graduação em Farmacologia, Universidade Federal de Santa Maria, Rio Grande do Sul, Brazil
[i] Campus Palmeira das Missões, Universidade Federal de Santa Maria, Rio Grande do Sul, 98300-000, Brazil
[j] Life Sciences Institute, Universidade Federal de Juiz de Fora. Governador Valadares, Minas Gerais, 35010-180, Brazil
[k] Departamento de Ciências da Vida - DCVIDA, Universidade Regional do Noroeste do Estado do Rio Grande do Sul - UNIJUÍ, Ijuí, Rio Grande do Sul, 98700-000, Brazil
[l] Centro Estadual de Vigilância em Saúde. Secretaria Estadual da Saúde do Rio Grande do Sul. Porto Alegre, Rio Grande do Sul, 90610-000, Brazil
[m] Hospital Universitário de Santa Maria (HUSM), Universidade Federal de Santa Maria (UFSM), Av. Roraima, 1000, Santa Maria, Rio Grande do Sul, 97105-900, Brazil
[n] Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Santa Maria (UFSM), Av. Roraima, 1000, Santa Maria, Rio Grande do Sul, 97105-900, Brazil
[o] Núcleo de Plataformas Tecnológicas (NPT), Instituto Aggeu Magalhães (IAM), FIOCRUZ-Pernambuco, Recife, Pernambuco, 50670-420, Brazil
[p] Núcleo de Ciências da Vida, Universidade Federal de Pernambuco (UFPE), Centro Acadêmico do Agreste-Rodovia BR-104, Caruaru, Pernambuco, 55002-970, Brazil
[q] Departamento de Microbiologia, Instituto Aggeu Magalhães (IAM), FIOCRUZ-Pernambuco, Recife, Pernambuco, 50670-420, Brazil
[r] Laboratório de Virologia Molecular, Instituto Carlos Chagas, Fundação Oswaldo Cruz, Curitiba, Paraná, Brazil
[s] Departamento de Zootecnia e Ciências Biológicas, Universidade Federal de Santa Maria, Palmera das Missões, Rio Grande do Sul 98300-000, Brazil
[t] Department of Arbovirology and Entomology, Bernhard Nocht Institute for Tropical Medicine, WHO Collaborating Center for Arbovirus and Hemorrhagic Fever Reference and Research. National Reference Center for Tropical Infectious Diseases. Bernhard-Nocht-Straße 74 20359 Hamburg, Germany

ARTICLE INFO

ABSTRACT

Brazil's COVID-19 response has faced challenges due to the continuous emergence of variants of concern (VOCs), emphasizing the need for ongoing genomic surveillance and retrospective analyses of past epidemic waves to reassess and fine tune containment protocols. Rio Grande do Sul (RS), Brazil's southernmost state, has international borders and trades with Argentina and Uruguay, along with significant domestic connections within Brazil. The identification of source and sink transmission chains at national and international scales can identify main hubs and pathways to target future interventions. In this study we investigated the RS state role in the

* Corresponding author.
E-mail address: gabriel.wallau@fiocruz.br (G.L. Wallau).
[1] Equal contribution
[2] Equal contribution

national and international SARS-CoV-2 transmission chains, which has not been fully explored. Nasopharyngeal samples from various municipalities in RS were collected between June 2020 and July 2022. SARS-CoV-2 whole genome amplification and sequencing were performed using high-throughput Illumina sequencing. Bioinformatics analysis encompassed the development of scripts and tools to perform subsampling taking into account epidemiological information to reduce sequencing disparities bias among the regions/countries, genome assembly, and large-scale alignment and phylogenetic reconstruction. We sequenced a total of 1,480 SARS-CoV-2 genomes from RS, covering all major regions. Sequences predominantly represented Gamma (April-June 2021) and Omicron (January-July 2022) lineages. Phylogenetic analysis revealed a regional pattern for transmission dynamics, particularly with Southeast Brazil for Gamma, and a range of inter-regional connections for Delta and Omicron within the country. On the other hand, international and cross-border transmission with Argentina and Uruguay was rather limited. We evaluated the three VOCs circulation over two years in RS using a new subsampling strategy based on the number of cases in each state during the circulation of each VOC. In summary, the retrospective analysis of genomic surveillance data demonstrated that virus transmission was less intense between country borders than within the country. These findings suggest that while non-pharmacological interventions were effective to mitigate transmission across international RS land borders, they were insufficient to contain transmission at the domestic level.

1. Introduction

The first cases of effective human-to-human transmission of SARS-CoV-2 were identified in Wuhan, China, in December 2019 (WHO Director, 2020). Since then, this virus has spread globally culminating in a pandemic declared on 11th March 2020 (WHO, 2020). As of February 2024, there have been 676 million global cases and more than 6.8 million deaths (COVID-19 Map) and Brazil registered more than 37 million cases and almost 700 thousand deaths in the same period (COVID-19 Map). These figures are a consequence of the high transmissibility of air-born pathogens such as the SARS-CoV-2 through symptomatic and asymptomatic patients and continuous adaptations to the new human host (Harrison et al., 2020; Hu et al., 2020; Yanes-Lane et al., 2020). The rapid evolution of the SARS-CoV-2 under changing evolutionary pressures associated with millions of infected patients (COVID-19 Map) has resulted in the emergence and spread of more transmissible and/or immunological evasive SARS-CoV-2 lineages around the globe that are dynamically classified as Variants of Interest (VOIs) and Variants of Concern (VOCs) (CDC, 2020).

Besides the development of different vaccine technologies (COVID19 Vaccine Tracker) and the success of vaccine campaigns across the world, which lead to a decrease in severe disease and hospitalizations (Anand and Stahel, 2021; Mohammed et al., 2022; Thompson Mark G. et al., 2021), the SARS-CoV-2 is still circulating and evolving into new lineages. These new lineages carry specific sets of immune escape mutations allowing the reinfection of previously infected or vaccinated hosts (Ai et al., 2022; Dejnirattisai et al., 2021; Hoffmann et al., 2021). Since the start of the pandemic, five major VOCs encompassing several lineages have originated and spread across the globe. The first VOC identified was the Alpha lineage in September 2020 (Hill et al., 2022) followed by Beta in October 2020 (Tegally et al., 2021), Gamma in November 2020 (Naveca et al., 2021), Delta in October but spread globally in May 2021 (McCrone et al., 2022), and Omicron in November 2021 (Viana et al., 2022), which is the single VOC circulating nowadays, and now considered as a VOI (Tracking SARS-CoV-2 variants), with several sublineages co-circulating (Tracking SARS-CoV-2 variants).

Knowledge about the spreading of different SARS-CoV-2 lineages is only possible due to the efforts of genomic surveillance, an effort carried out by thousands of research groups across the globe (Brito et al., 2022; Robishaw et al., 2021). In Brazil, due to its large geographic territory, several groups have been characterizing the emergence and spreading of different SARS-CoV-2 lineages across the country (Arantes et al., 2022; Faria et al., 2021; Giovanetti et al., 2022a, 2022b; Naveca et al., 2021; Nonaka et al., 2021). Besides the efforts to perform genomic surveillance studies in Brazil, the majority of Brazilian studies focus on the Southeast region - the most populous region, and with the major number of cases - (Giovanetti et al., 2022b). Studies focusing on Rio Grande do Sul state were carried out at the beginning of the pandemic (Francisco Jr et al.,

2021; Martins et al., 2021; Mayer et al., 2023) or focused on specific cities (Franceschi et al., 2021b, 2021a). Herein, we aimed to characterize the circulation of SARS-CoV-2 lineages during the first two and a half years of the pandemic (up to July 2022) across Rio Grande do Sul state and comparatively evaluate the rate of within (Rio Grande do Sul and other Brazilian states) and between countries (Argentina and Uruguay) transmission.

In order to address these questions we sequenced 1480 SARS-CoV-2 genomes from samples collected from June 2020 to July 2022. The genomes represent several lineages from Gamma, Delta, and Omicron VOCs as well as non-VOI and non-VOC lineages. Moreover, the samples cover 185 cities across the Rio Grande do Sul being the most spatiotemporal comprehensive dataset analyzed from the state so far. Our results highlight that non-pharmacological interventions were effective in controlling inter country transmission by land borders but not sufficient to restrict within country transmission. These findings are of particular importance to prioritize non-pharmacological interventions at the interstate and national level during the following SARS-CoV-2 waves.

2. Material and methods

Nasal and oropharyngeal specimens, previously identified as positive for SARS-CoV-2, were obtained from the Universidade Federal de Santa Maria main campus, the Palmeira das Missões Campus and the samples forwarded to LACEN-RS from across the state representing 185 Rio Grande do Sul municipalities mainly covering the Central and Northwest regions. The data was accessed for research purposes from 23 June 2020 to 7 July 2022. Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR) assays were conducted utilizing the Biomol One-Step Covid-19 Kit (IBMP, Paraná, BR) and the Molecular SARS-CoV-2 Kit (Bio-manguinhos, Rio de Janeiro, RJ, BR), strictly adhering to the specifications provided by the manufacturers. Specimens exhibiting a Cycle Threshold (Ct) value below 25 were subsequently subjected to amplification and sequencing, as delineated in (Itokawa et al., 2020; Quick, 2020).

2.1. Viral genetic material sequencing

For this study, two distinct methodologies were employed to synthesize complementary DNA (cDNA) and to amplify the SARS-CoV-2 genome. These included: the short amplicon approach based on the ARTIC protocol (accessible at https://github.com/artic-network/artic-ncov2019) (Quick, 2020); and the incorporation of three sets of primers within the COVIDSeq protocol, following Naveca, et al. (2021). After the generation of genome-wide amplicons, samples were processed for sequencing employing either the Illumina DNA Prep (Illumina, San Diego, CA, USA) or the COVIDSeq (Illumina, San Diego, CA, USA)

library preparation protocols, following the guidelines provided by the manufacturer. The sequencing process was executed using the Illumina MiSeq system, specifically employing the MiSeq Reagent Kit V3 for a paired-end 150 cycles flow cell. Ethical clearance for this study was granted by CEP UFSM under the authorization number 52,939, 821.5.0000.5346 and 47,588,621.7.1001.5346 and by CEP from Escola de Saúde Pública/SES-RS under the authorization number CAAE: 67, 181,123.1.0000.5312.

2.2. Genome assembly

In this study, we used the ViralFlow v0.0.6 workflow (Dezordi et al., 2022). This workflow was designed for efficient and accurate assembly of viral genomes. Briefly, raw sequence data were quality filtered to remove reads (–cut_front –cut_tail –qualified_quality_phred 20 -l 35 -f 35 -t 35 -F 35 -T 35) with fastp tool (Chen et al., 2018). Subsequently, the cleaned reads were mapped with BWA (Li and Durbin, 2009) against the Wuhan SARS-CoV-2 genome (NC_045512.2). The consensus genomes were generated with the iVar tool (Grubaugh et al., 2019), where a base is called to the consensus when it reaches at least 5x of coverage depth; on multi-allele loci, alleles with higher frequency (VAF $\geq$ 0.51) are considered for the consensus genome. The assembly metrics and SARS-CoV-2 lineages are defined using bamdst (shiquan, 2024) and pangolin (O'Toole et al., 2021), respectively, both present in the Viral-Flow workflow. All samples with 70 % or more of coverage breadth are included in the initial phylogenetic analysis with MAPLE v0.3.1 (De Maio et al., 2023) to detect if the lack of sequencing regions results in long branches on phylogenetic trees.

2.3. Recombinant screening

To analyze recombination, samples we screened in the web platform RIVET (Smith et al., 2023) looking for recombinant or descendent of recombinant samples related to the data produced in this work.

2.4. Subsampling strategy

In light of the extensive dataset of SARS-CoV-2 genomes available in the GISAID-EpiCoV database (Khare et al., 2021), which encompasses millions of genomes globally (15,271,031 genomes in the collection date of this study), including hundreds of thousands from Brazil (214,213 genomes), it became imperative to utilize a systematic subsampling strategy informed by epidemiological data. To facilitate this, we developed an in-house R (https://www.r-project.org/) script, named explore. R (see **Data Availability** for details), to determine the appropriate number of genomes to subsample across various Brazilian regions for Gamma, Delta, and Omicron. This script harnesses data accessed for research purpose from both the EpiCoV database and covid19.org.br, collected on February 6, 2023, to assess and use the ratio of COVID-19 cases to estimate the proportional subsampling of the available genomes for each aforementioned VOCs in specific periods.

To delineate the temporal and spatial distribution of the Gamma, Delta, and Omicron variants within Brazil, we referred to existing literature to establish the onset of their circulation in the country: Gamma in November 2020 (Naveca et al., 2021), Delta in April 2021 (McCrone et al., 2022), and Omicron in November 2021 (Viana et al., 2022). Our study's scope was extended until August 2022, based on the date of the most recent sample in our analysis.

To perform the subsampling process, we developed an in-house tool: the Gisaid Subsampling Toolkit, or GIST. GIST processes a JSON file containing the count of genomes from various Brazilian regions, derived from the output of the 'explore.R' script. It employs a combination of Augur (Huddleston et al., 2021), BLAST (Altschul et al., 1990), and MAFFT (Katoh and Standley, 2013) for three distinct analyses (**Code Availability** section for more details, **SupplementaryData1**, all supplementary data can be found at https://doi.org/10.6084/m9.figsh are.25818910.v1). Briefly, in the first analysis, we utilize Augur to sample genomes. This step involves the exclusion of genomes from non-human hosts, those lacking a collection date, and sequences shorter than 28,400 bases, approximately 95 % of the SARS-CoV-2 Wuhan genome's coverage. The sampling is stratified based on the number of genomes by state, further grouped by pango lineage, year, and month. Following this initial sampling, the second phase of analysis enriches the dataset with genetically similar genomes. This is achieved through a BLAST analysis filtering results between 99 % and 99.98 % of identity with query coverage high scoring pais (HSP) greater or equal to 99.9 %. In the final stage of analysis, we conduct a multiple sequence alignment using MAFFT, employing parameters such as –keeplength, –kimura 1, and –6merpair, while specifically mask the UTR regions.

2.5. Phylogenetic analysis

Three initial trees - one for Gamma, Delta, and Omicron datasets - were reconstructed using MAPLE v0.3.1 (De Maio et al., 2023) to detect long branches. Genomes showing long branches (irrespective of genomic coverage) were removed from the original alignment ($n = 97$, details in **SupplementaryData2**). The edited alignments were used in a maximum likelihood phylogenetic analysis with IQ-TREE2 v.2.2.0 (Minh et al., 2020), the best substitution model was defined using ModelFinder (Kalyaanamoorthy et al., 2017) implemented on IQ-TREE2 and the SH-aLRT test (Guindon et al., 2010) estimated the branch support. The phylogenetic trees were annotated in iTOL v.6 (Letunic and Bork, 2007) to identify clades of specific lineages with branch support higher than 0.8 (aLRT) with at least 4 genomes from Rio Grande do Sul. The identified clades were split into different datasets, one per clade.

2.6. Temporal estimation

To estimate the tMRCA of clades identified in phylogenetic analysis, each lineage dataset was used in an analysis flow which encompassed: A first phylogenetic tree reconstruction with IQ-TREE2 v.2.2.0; A TempEst v.1.5.3 (Rambaut et al., 2016) analysis to identify the age root correlation; removal of outliers with in-house R scripts, that receives the output of TempEst and identify outliers sequences based on the thresholds of 1.5x of Interquartile Range IQR of all samples in the dataset; A second tree reconstruction with IQ-TREE2 v.2.2.0 using the dataset alignments without outliers; And finally a temporal analysis with TreeTime v.0.11 (Sagulenko et al., 2018).

3. Results

3.1. Sequencing results

During the period and for the purpose of the present study, 1,480 genomes were sequenced and deposited on EpiCoV - GISAID database (**SupplementaryData3**). The samples cover municipalities from the North, Center, and West of the Rio Grande do Sul state (https://microre act.org/project/2utyqQKoA7zDkYHEuzvyTG-sarsufsmlacen), and encompass approximately 30% of all genomes submitted for the state (1,480 from 5,065 genomes) and deposited on GISAID in the period of the study (**SupplementaryData3**). The submitted consensus genomes from our group represent samples that meet one of the following criteria: 90% or more of genome coverage breadth (horizontal coverage) or 70% or more genome coverage breadth and that do not belong to clades with long branches on initial MAPLE phylogenetic analysis. The samples are majorly composed of Gamma lineages and sublineages during the period from April to June 2021 (Fig. 1), and Omicron lineages and sublineages during the period of January to July 2022, which is in line with the general numbers presented on EpiCoV - GISAID (Fig. 2). The recombinant analysis showed a putative recombinant sample (EPI_ISL_16878846) with recombinant signal between P.1 and P.1.15 lineages. Besides, the focus of this study is on Rio Grande do Sul, we
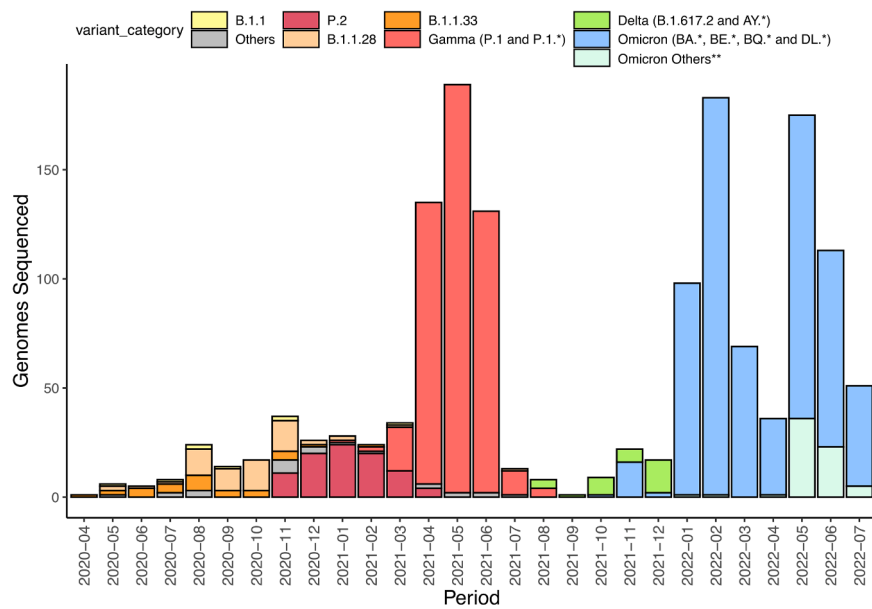
Fig. 1. Number of samples sequenced in this study per month representing the 1480 samples. **Omicron genomes not related to BA, BE, BQ, and DL lineages.

noticed similar patterns of lineage frequencies sequenced in other regions of Brazil (Fig. 2). The most prevalent lineages for each analyzed VOC on Rio Grande do Sul were P.1, P.1.14, P.1.2, P.1.4 and P1.7 for Gamma; AY.43, AY.99.2, AY.99.1, AY.101, AY.34.1 and AY.46.3 for Delta; and the most prevalent lineages BA.1, BA.1.1, BA.2, BA.5, BA.5.2, BA.4 and BQ for Omicron (Fig. 2).

3.2. SARS-Cov-2 clades circulating in Rio Grande do Sul

The subsampling strategy with explorer.R and GIST resulted in three major datasets: One for each VOC Gamma, Delta, and Omicron, encompassing the most frequent lineages on Rio Grande do Sul during the period of the study as well as outgroups with ancestors of respective lineages (**SupplementaryData4**). Each dataset corresponds to Brazilian genomes plus genomes from Latin America (particularly Paraguay, Argentina and Uruguay) and other regions of the globe, recovered from the GIST BLAST analysis (**SupplementaryData5**).

The initial phylogeny of Gamma corresponds to 4545 genomes (Fig. 3A), where 6 clades were identified (Fig. 3B). One clade (P.1.14-I) represents sequences from P.1.14, two clades (P.1-I~II) represent sequences of P.1 lineage, and three (P.1.2-I~III) of P.1.2 lineage. The Gamma analysis shows samples from Rio Grande do Sul clustering with sequences from the Southeast region of Brazil: The P.1.14-I and P.1-II clades encompass basal sequences from Minas Gerais (MG) and São Paulo (SP) states, the P.1-I clade presents basal Rio Grande do Sul samples clustered with South America samples from Argentina and Paraguay, and the clades P.1-II and P.1.2-III have well-defined sister clades from Uruguay and São Paulo, respectively.

The initial phylogeny of Delta corresponds to 4358 genomes (Fig. 4A) where 6 clades were identified (Fig. 4B). Four clades (AY.99.2-I~IV) represent sequences from AY.99.2 lineage, and 2 clades (AY.101-I~II) represent sequences from AY.101 lineage. The Delta analysis shows different clustering of Rio Grande do Sul (RS) and sampled from other Brazilian regions. The AY.99.2-I clade shows a possible transmission from Rio Grande do Sul to Minas Gerais, and the AY.99–2-IV from Minas Gerais to Rio Grande do Sul. The clade AY.101-I shows a route of transmission from Rio Grande do Sul to the Central-West region of Brazil, as well as from Rio Grande do Sul to Rio de Janeiro (RJ), and the AY.101-II clade from Rio Grande do Sul to São Paulo state.

The omicron phylogenetic analysis was performed with 4639 genomes (Fig. 5A) representing 5 distinct clades (Fig. 5B) with Rio Grande

do Sul samples from BA.1.1, BA.5.2, and BA.2 lineages. Different from Delta and Gamma analysis, where several clades clustered with other clades from the South-East region - mainly from São Paulo, Rio de Janeiro, and Minas Gerais states - the Omicron clades are mostly composed of Rio Grande do Sul samples, except for BA.5.2-I clade, that clustered with samples from the Goias state (GO) - Central-West region.

### 3.3. Temporal estimation of 7 lineages circulating on Rio Grande do Sul

To identify the possible emergence and spreading timing of the lineages associated with the identified clades within the state, we performed temporal analysis using treetime for 7 datasets (Table 1) which represents the clades of lineages identified with Rio Grande do Sul samples in the evolutionary analysis and which have not been analyzed in other studies.

Considering the tRMCA from treetime as the start time of lineages circulation, and the last collection date of samples (**Supplementary Data3**) as the end time of circulation, the Gamma P.1.2 and P.1.14 lineages were estimated to have occurred on Rio Grande do Sul between mid-December 2020 to late June 2021 for P.1.14 and mid-January 2020 to late June 2021 for P.1.2.

For Delta lineages, the AY.101 was estimated to have circulated between mid-March 2021 to late November 2021 and the AY.99.2 from late March 2021 to early December 2021, and for Omicron lineages, the BA.1.1 started circulating in the state in mid-August 2021, the BA.2 in mid-December 2021 and the BA.5.2 in mid-February 2022 and both of them are still circulating during the time of the data collection of this study: early July 2022.

4. Discussion

The SARS-CoV-2 virus is one of the latest example of a zoonotic pathogen that has made its way to effective human-to-human transmission resulting in the COVID-19 pandemic. Despite numerous efforts to implement large-scale genomic surveillance around the world, there remain underrepresented populations and areas where laboratory and genomic surveillance is scarce I. (Brito et al., 2022). In Brazil, SARS-CoV-2 genomic surveillance was established across the country, but as expected, most of the data comes from the most populous cities on the Atlantic coast and large capital cities from states without sea borders where there were more available diagnostic and research laboratories
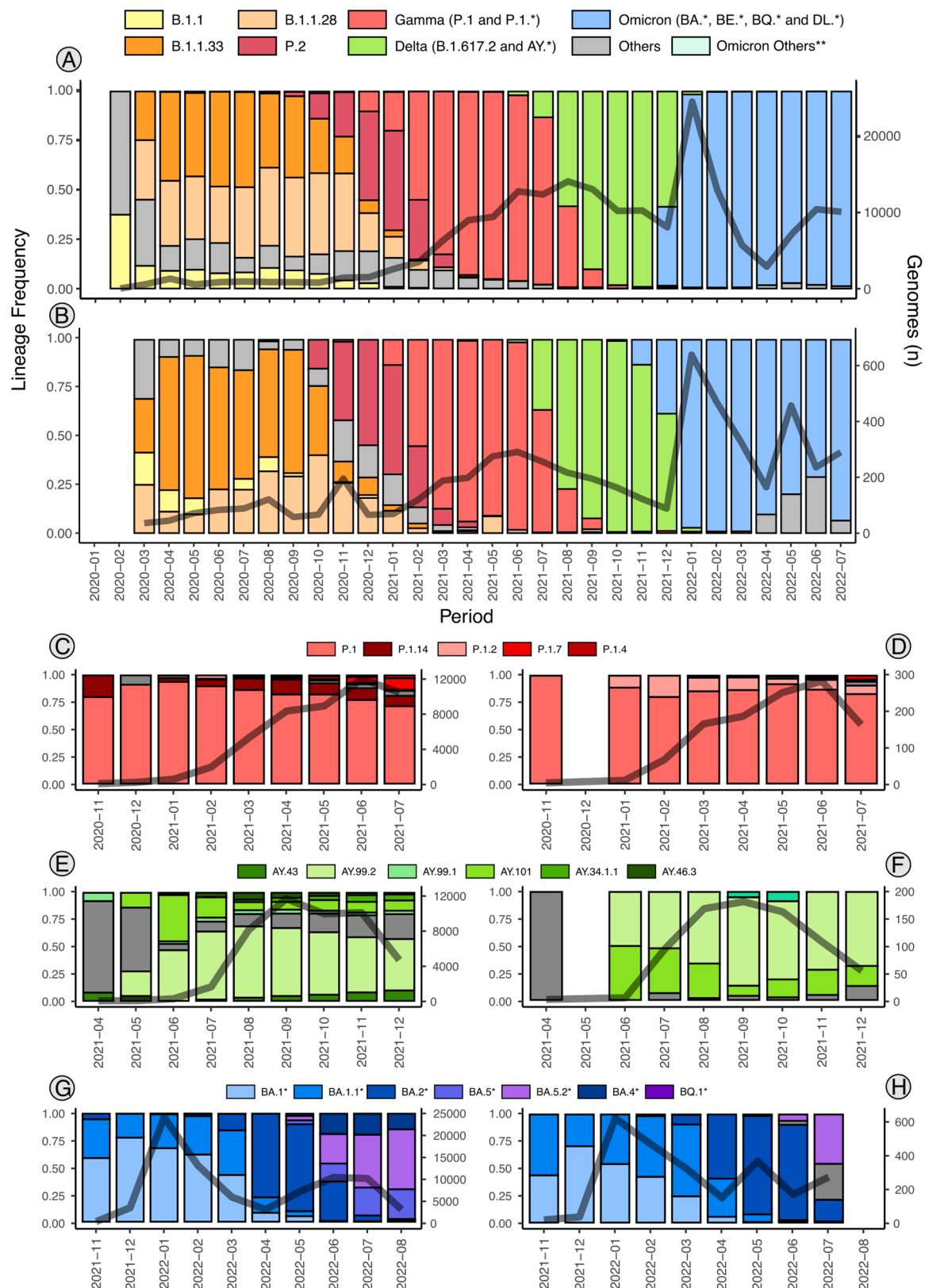
Fig. 2. Prevalence of SARS-CoV-2 lineages and number of genomes per period, based on EpiCoV GISAID data. **A.** Specific lineages and variants in Brazil. **B.** Specific lineages and variants in Rio Grande do Sul. **C.** Gama lineages in Brazil. **D.** Gama lineages in Rio Grande do Sul. **E.** Delta lineages in Brazil. **F.** Delta lineages in Rio Grande do Sul. **G.** Omicron lineages in Brazil. **H.** Omicron lineages in Rio Grande do Sul. **Omicron genomes not related to BA, BE, BQ, and DL lineage.
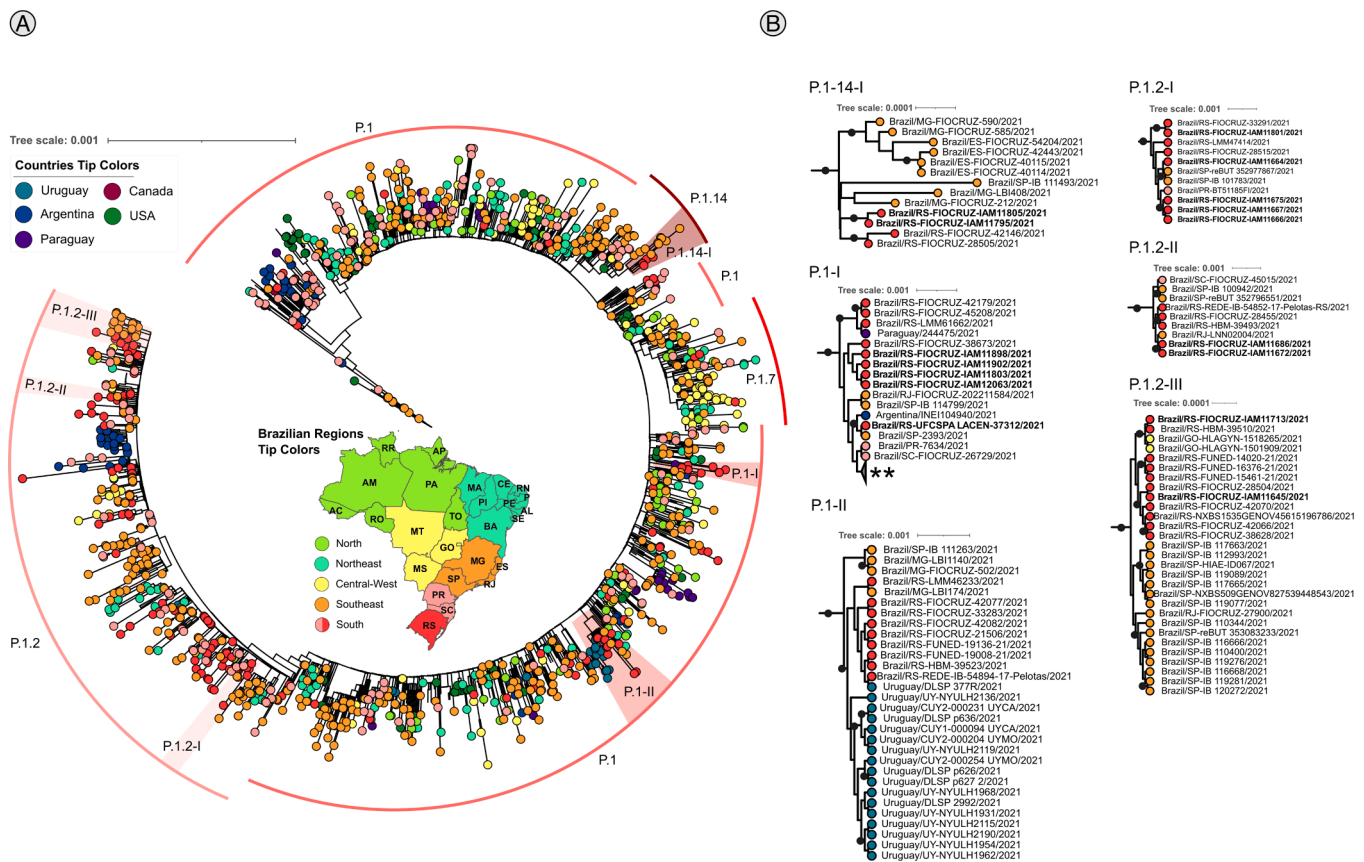
Fig. 3. Maximum likelihood analysis of Gamma variant lineages with 4545 samples, Branch support was accessed by the aLRT branch support. Black dots represent support ≥ 0.8. Tip colors are based on Brazilian regions and the 5 countries that presented the highest number of sequences recovered via blast. **A.** Complete tree. **B.** Well-supported subclades. Bold sample lanes represent genomes sequenced in this study. ** Collapsed clade that represents all sequences after P.1-I clade. Complete figure with branch support and uncollapsed clade is available on **SupplementaryData6**.

with sequencing capacity. That is the case of the Rio Grande do Sul state where there is a higher concentration of diagnostic and genomic information from the capital city Porto Alegre and surrounding municipalities. In this study, we generated 1480 genomes mainly from the central and Northeast regions of the state to augment the genomic information from the Rio Grande do Sul state. Then we subsampled the main lineages circulating in the state to test hypotheses regarding the sink and source role of the state to other Brazilian states and border countries Argentina and Uruguay. We found that inter-state transmission was much more important in SARS-CoV-2 lineages transmission spread from and to the Rio Grande do Sul state than inter-country transmission.

The extremely large number of SARS-CoV-2 genomes currently available (the global dataset used in this study included 15,271,032 genomes) makes phylogenetic inferences timing and computationally consuming (Morel et al., 2021). Current approaches to handle such a massive amount of data are focused on lineages subsampling taking into consideration lineages presence in different locations and prevalence (Hill et al., 2021). The Augur software (Huddleston et al., 2021), implements some sub-sampling algorithms but does not consider the epidemiological estimates such as the number of cases per region. Considering the global disparities of genome sequencing throughout different geographic regions (A.F. Brito et al., 2022) we implemented a sampling strategy based on the number of cases and genomes by region per time using the in-house script explore.R which allowed us to more precisely sample genomes following the increase and decrease in cases and mitigate sampling biases present in the global and local datasets.

Our global subsampled phylogenetic reconstruction of Delta and Omicron clades showed that the United States of America and the Central-West region of Brazil were placed in a basal position near the root suggesting that these regions were the source point to spread these lineages to other regions of Brazil. Previous evidence has shown the origins of each analyzed VOC, such as Gamma with P.1 lineage in the North region of Brazil (Naveca et al., 2021), Delta with B.1.617.2 lineage in India (McCrone et al., 2022), and Omicron with B.1.1.529 lineage in South Africa (Viana et al., 2022). But this is the first large-scale analysis showing that the Delta lineage was probably introduced from the USA to Brazil and that the Central-West region of Brazil acted as an important hub for the spreading of the Omicron lineage to other major regions of the country.

Based on the augmented dataset of SARS-CoV-2 genomes from Rio Grande do Sul including genomes of previously undersampled regions (Center and Northwestern) we next evaluated if the main transmission chains with Rio Grande do Sul occurred with other Brazilian states or with border countries Argentina and Uruguay. Previous studies characterized the spreading of B.1.1.28 and B.1.1.33 lineages from Rio Grande do Sul to Uruguay (Mir et al., 2021) and the spreading of the P.1 lineage from Brazil to Uruguay (Rego e al, 2021). Besides the large terrestrial frontier between Rio Grande do Sul, Uruguay and Argentina, our results showed only one probable migration event of the P.1 lineage from Rio Grande do Sul to Uruguay (Fig. 3B, clade P.1-II). The clades of P.1.2 lineage contains samples from Rio Grande do Sul as ancestors from samples of the Southeast, in line with the evidence showing the likely origin of this sub-lineage in the state (Franceschi et al., 2021b). The P.1.14-I clade, showed an ancestral origin of P.1.14 samples from Southeast towards Rio Grande do Sul in line with the findings shown by Ferrareze, et al. (2024).

There are few studies about Delta lineages on Rio Grande do Sul, one of the reasons may be related to the low number of samples sequenced in
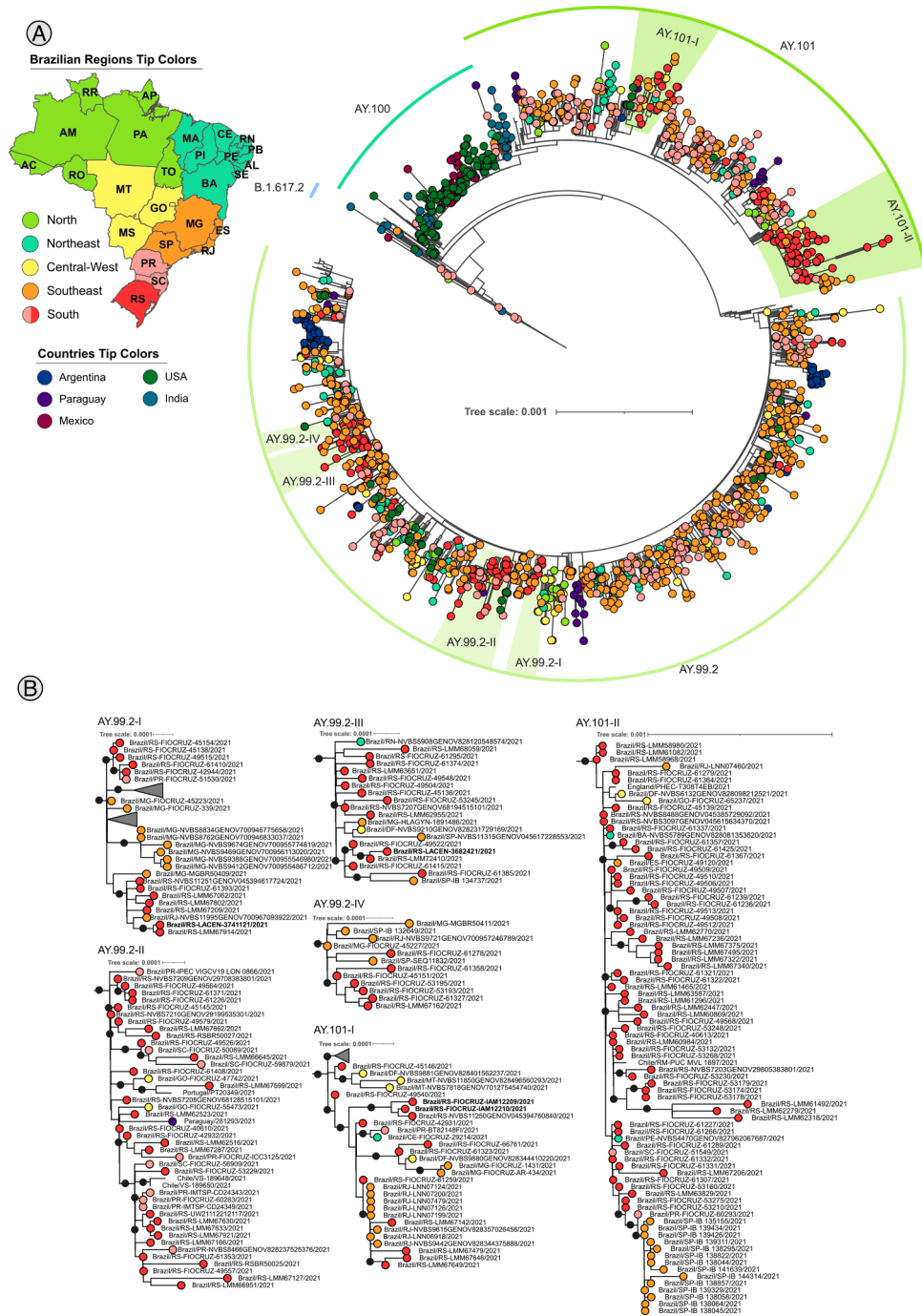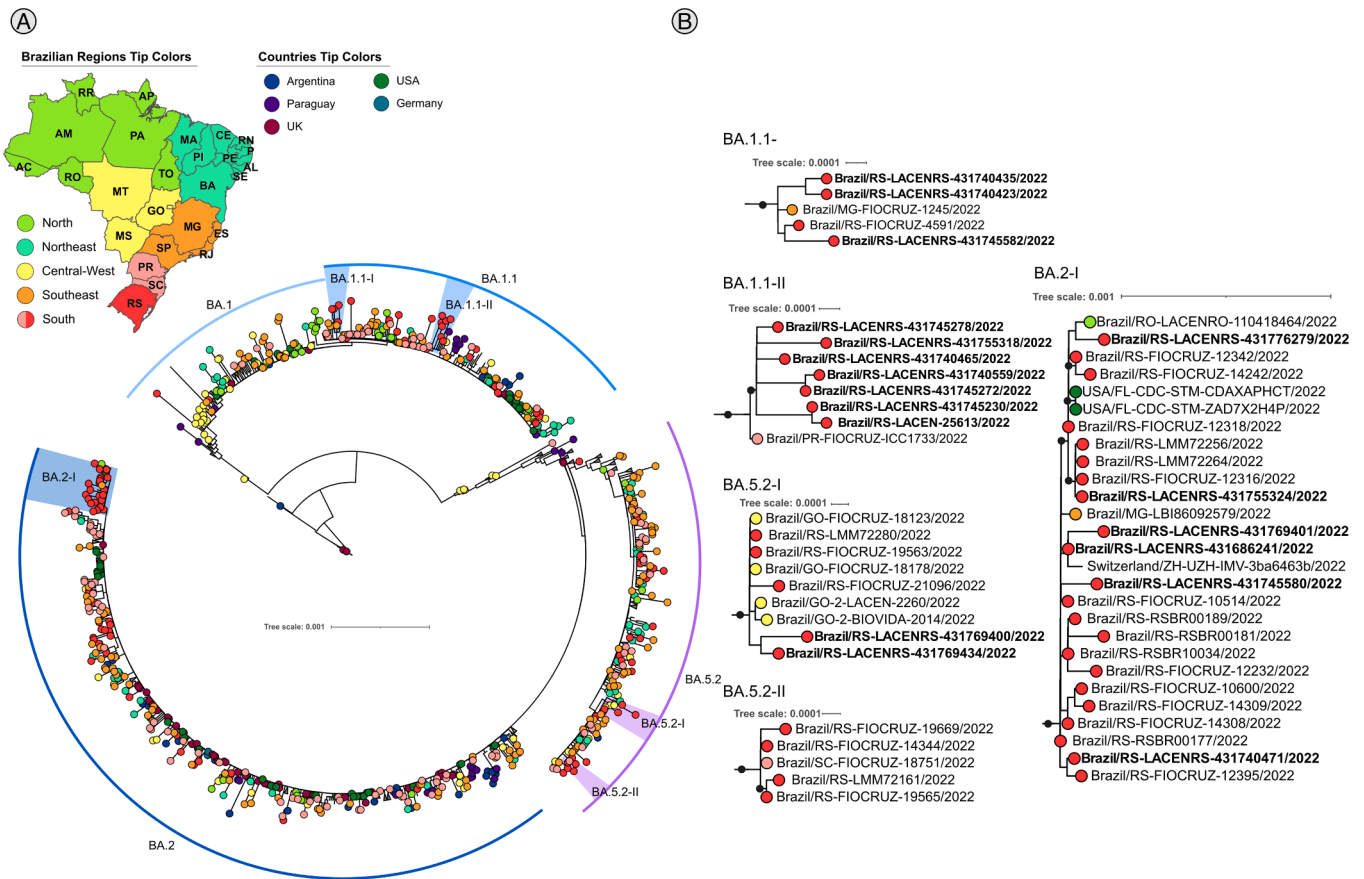
Fig. 4. Maximum likelihood analysis of Delta variant lineages with 4338 samples, numbers within with squares are aLRT branch support of respective clades. Black dots represent support ≥ 0.8. Colorful tips based on Brazilian regions and the 5 countries that presented the highest number of sequences recovered via blast. **A.** Complete tree. **B.** Weel supported subclades including Rio Grande do Sul samples. Bold sample lanes represent genomes sequenced in this study. Complete figure with branch support and uncollapsed clade is available on **SupplementaryData7**.

this period (Fig. 2**B**). Our results showed that two main Delta lineages circulated on Rio Grande do Sul: the AY.99.2 and the AY.101, corroborating the results from Castro et al. (2023) and Arantes et al. (2022) that addressed the early emergence of Delta in Brazil. The circulation of those lineages appears to have occurred from July 2021 to January 2022, with the tMRCA of those lineages estimated from late March 2021, which is in line with the estimations of Arantes et al., 2023 for AY.101 (Arantes et al., 2022). The clades of AY.99.2 showed samples from Rio Grande do Sul as ancestors (Fig. 4**B** Clade AY.99.2-I) with spreading to Minas Gerais (Southeast region), while the clade AY.99.2-IV showed the

opposite, which indicates a recurrent spreading pattern of AY.99.2 between those 2 states. The analysis of AY.101 clades also revealed Rio Grande do Sul samples as ancestors and a putative migration to Southeast states: Rio de Janeiro (AY.101-I clade) and São Paulo (AY.101-II clade).

Our results showed that ancestral samples of AY.99.2 from the Rio Grande do Sul state and posterior spreading to the Southeast were in line with other studies (Ferrareze et al., 2024). On the other hand, the same study identified the introduction of AY.101 lineage in the Southeast region in early January 2021, and the posterior introduction in Rio

Fig. 5. Maximum likelihood analysis of lineages of the Omicron variant with 4639 samples, aLRT branch support. Black dots represent support ≥ 0.8. Colorful tips based on Brazilian regions and the 5 countries that presented the highest number of sequences recovered via blast. **A**. Complete tree. **B**. Subclades. Bold sample lanes represent genomes sequenced in this study. Complete figure with branch support and uncollapsed clade is available on **SupplementaryData8**.

Table 1
Datasets and tMRCA inferred by treetime tool.

lineage	number of samples	tMRCA
P.1.2	482	2021–01–12 ~ 2021.03
P.1.14	540	2020–12–15 ~ 2020.95
AY.101	1030	2021–03–22 ~ 2021.22
AY.99.2	1953	2021–03–26 ~ 2021.23
BA.1.1	1467	2021–10–09 ~ 2021.77
BA.5.2	384	2022–02–11 ~ 2022.11
BA.2	1530	2021–12–09 ~ 2021.93

Grande do Sul in July 2021 which diverged from our findings of an introduction from Rio Grande do Sul to the Southeast region. The divergence about the AY.101 phylogeographic inferences may be the result of different subsampling strategies, where Ferrareze et al. (2024) used the Augur subsampling strategy based on Pango lineage, and we used the Augur inside GIST performing a subsampling based on normalized genomes by state cases and then the sampling by Pango Lineage and collection date.

The Omicron analysis showed three main lineages circulating in Rio Grande do Sul during the period of the study: BA.1.1, BA.5.2, and BA.2 which is in line with other Brazilian states (Fig. 2**G, H**). During the first quarter of 2022, the BA.1 and BA.1.1 lineages were the most prevalent, as described previously (da Silva et al., 2022) but were further replaced by BA.2, BA.5, and BQ from June to July 2022 (Rhoden et al., 2024). There is no phylogenetic information in the literature about the relation of Omicron lineages from Rio Grande do Sul and other locations, our results showed that excluding the BA.5.2-I clade (Fig. 5**B**), the omicron

clades are majorly formed by samples from Rio Grande do Sul. The BA.5.2-I clade showed a close relation between Rio Grande do Sul samples and Goias samples (Central West Brazilian region).

Despite the standardization of sampling methods, sequencing disparities (Brito et al., 2022) may still generate sampling biases that can affect genomic surveillance studies, here we identified both local sampling disparities on the Rio Grande do Sul (by state regions and periods, see https://microreact.org/project/2utyqQKoA7zDkYHEuzvyTG-sarsu fsmlacen) and in countries that have terrestrial borders with the Southern region of Brazil (Argentina (Gamma = 251, Delta = 196, and Omicron = 206 sequences included in each analysis), and Uruguay (Gamma = 117, Delta = 0, and Omicron = 69 sequences included in each analysis), details see **SupplementaryData 9**). These results coupled to the absence of traveling information as well as no diagnostic test focused on Rio Grande do Sul state land borders may have biased our inferences. On the other hand, this was the most comprehensive study posed to augment the Rio Grande do Sul dataset in underrepresented areas as well as the first to employ a case incidence aware subsampling approach to mitigate the impact of these biases.

Overall, our findings suggest that despite the large land border and intense international transportation of goods between Rio Grande do Sul, Argentina, and Uruguay few examples of cross-border transmission occurred, likely due to more effective non-pharmacological interventions applied at countries' borders and that the major sink and source transmission pathways of SARS-CoV-2 with the Rio Grande do Sul state occurred between Brazilian states from the Southeast and Central West part of the country.

5. Conclusion

Overall, our study underscores the importance of robust sampling strategies and comprehensive genomic surveillance efforts to accurately track the spatio-temporal spread of SARS-CoV-2 lineages. By applying a sampling strategy based on the number of genomes and epidemiological information, we described the most prevalent clades of Gamma, Delta, and Omicron lineages that circulated in Rio Grande do Sul Brazil, during the period from June 2020 to July 2022 and showed that the large majority of transmission events took place between Brazilian states of the South-East and Central-West regions than between countries sharing land borders with Rio Grande do Sul.

CRediT authorship contribution statement

Filipe Zimmer Dezordi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **José Valter Joaquim Silva Júnior:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Terimar Facin Ruoso:** Writing – review & editing, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Angela Giovana Batista:** Writing – review & editing, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Pedro Mesquita Fonseca:** Methodology, Investigation, Formal analysis, Data curation. **Larissa Paim Bernardo:** Resources, Methodology, Investigation, Data curation. **Richard Steiner Salvato:** Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Data curation. **Tatiana Schäffer Gregianini:** Methodology, Investigation, Data curation. **Thaísa Regina Rocha Lopes:** Resources, Methodology, Investigation, Data curation. **Eduardo Furtado Flores:** Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Rudi Weiblen:** Project administration, Investigation, Funding acquisition. **Patrícia Chaves Brites:** Resources, Methodology, Investigation, Funding acquisition, Conceptualization. **Mònica de Medeiros Silva:** Methodology, Investigation, Formal analysis, Conceptualization. **João Batista Teixeira da Rocha:** Project administration, Funding acquisition. **Gustavo de Lima Barbosa:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Lais Ceschini Machado:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Alexandre Freitas da Silva:** Methodology, Investigation, Formal analysis, Data curation. **Marcelo Henrique Santos Paiva:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Matheus Filgueira Bezerra:** Writing – review & editing, Supervision, Resources, Methodology, Investigation. **Tulio de Lima Campos:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Tiago Gräf:** Writing – review & editing, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Daniel Angelo Sganzerla Graichen:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Elgion Lucio da Silva Loreto:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Gabriel da Luz Wallau:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2024.199500.

Data availability

All the data is available in the supplementary material of the manuscript and accession numbers have been provided.

References

Ai, J., Zhang, H., Zhang, Yi, Lin, K., Zhang, Yanliang, Wu, J., Wan, Y., Huang, Y., Song, J., Fu, Z., Wang, H., Guo, J., Jiang, N., Fan, M., Zhou, Y., Zhao, Y., Zhang, Q., Liu, Q., Lv, J., Li, P., Qiu, C., Zhang, W., 2022. Omicron variant showed lower neutralizing sensitivity than other SARS-CoV-2 variants to immune sera elicited by vaccines after boost. Emerg. Microbes Infect. 11, 337–343. https://doi.org/10.1080/22221751.2021.2022440.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anand, P., Stahel, V.P., 2021. The safety of Covid-19 mRNA vaccines: a review. Patient Saf. Surg. 15, 20. https://doi.org/10.1186/s13037-021-00291-9.

Arantes, I, Gomes Naveca, F., Gräf, T., COVID-19 Fiocruz Genomic Surveillance Network, Miyajima, F., Faoro, H., Luz Wallau, G., Delatorre, E., Reis Appolinario, L., Cavalcante Pereira, A., Venas, T.M.M., Sampaio Rocha, A., Serrano Lopes, R., Mendonça Siqueira, M., Bello, G., Cristina Resende, P., 2022. Emergence and Spread of the SARS-CoV-2 Variant of Concern Delta across Different Brazilian Regions. Microbiol. Spectr. 10. https://doi.org/10.1128/spectrum.02641-21 e02641-21.

Brito, A.F., Semenova, E., Dudas, G., Hassler, G.W., Kalinich, C.C., Kraemer, M.U.G., Ho, J., Tegally, H., Githinji, G., Agoti, C.N., Matkin, L.E., Whittaker, C., Howden, B. P., Sintchenko, V., Zuckerman, N.S., Mor, O., Blankenship, H.M., de Oliveira, T., Lin, R.T.P., Siqueira, M.M., Resende, P.C., Vasconcelos, A.T.R., Spilki, F.R., Aguiar, R.S., Alexiev, I., Ivanov, I.N., Philipova, I., Carrington, C.V.F., Sahadeo, N.S. D., Branda, B., Gurry, C., Maurer-Stroh, S., Naidoo, D., von Eije, K.J., Perkins, M.D., van Kerkhove, M., Hill, S.C., Sabino, E.C., Pybus, O.G., Dye, C., Bhatt, S., Flaxman, S., Suchard, M.A., Grubaugh, N.D., Baele, G., Faria, N.R., 2022. Global disparities in SARS-CoV-2 genomic surveillance. Nat. Commun. 13, 7003. https://doi.org/10.1038/s41467-022-33713-y.

CDC, 2020. Coronavirus disease 2019 (COVID-19). Cent. Dis. Control Prev. URL. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (accessed 4.22.24).

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

COVID-19 Map. Johns Hopkins Coronavirus Resour. Cent. URL https://coronavirus.jhu.edu/map.html (accessed 4.22.24).

da Silva, M.S., Gularte, J.S., Filippi, M., Demoliner, M., Girardi, V., Mosena, A.C.S., Pereira, V.M., de, A.G., Hansen, A.W., Weber, M.N., de Almeida, P.R., Fleck, J.D., Bó, A.G.B.L.D., Jones, M.H., Friedrich, F., Filho, L.A., Klamt, F., Spilki, F.R., 2022. Genomic and epidemiologic surveillance of SARS-CoV-2 in Southern Brazil and identification of a new Omicron-L452R sublineage. Virus. Res. 321, 198907. https://doi.org/10.1016/j.virusres.2022.198907.

De Maio, N., Kalaghatgi, P., Turakhia, Y., Corbett-Detig, R., Minh, B.Q., Goldman, N., 2023. Maximum likelihood pandemic-scale phylogenetics. Nat. Genet. 55, 746–752. https://doi.org/10.1038/s41588-023-01368-0.

Dejnirattisai, W., Zhou, D., Supasa, P., Liu, C., Mentzer, A.J., Ginn, H.M., Zhao, Y., Duyvesteyn, H.M.E., Tuekprakhon, A., Nutalai, R., Wang, B., López-Camacho, C., Slon-Campos, J., Walter, T.S., Skelly, D., Clemens, S.A.C., Naveca, F.G., Nascimento, V., Nascimento, F., Costa, C.F.da, Resende, P.C., Pauvolid-Correa, A., Siqueira, M.M., Dold, C., Levin, R., Dong, T., Pollard, A.J., Knight, J.C., Crook, D., Lambe, T., Clutterbuck, E., Bibi, S., Flaxman, A., Bittaye, M., Belij-Rammerstorfer, S., Gilbert, S.C., Carroll, M.W., Klenerman, P., Barnes, E., Dunachie, S.J., Paterson, N.G., Williams, M.A., Hall, D.R., Hulswit, R.J.G., Bowden, T.A., Fry, E.E., Mongkolsapaya, J., Ren, J., Stuart, D.I., Screaton, G.R., 2021. Antibody evasion by the P.1 strain of SARS-CoV-2. Cell 184, 2939–2954. https://doi.org/10.1016/j.cell.2021.03.055 e9.

Dezordi, F.Z., Neto, A.M., da, S., Campos, T., de, L., Jeronimo, P.M.C., Aksenen, C.F., Almeida, S.P., Wallau, G.L., 2022. ViralFlow: a versatile automated workflow for SARS-CoV-2 genome assembly, lineage assignment, mutations and intrahost variant detection. Viruses. 14, 217. https://doi.org/10.3390/v14020217.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D., da, S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulswit, R.J.G., Franco, L.A.M., Ramundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Gaburo, N., Camilo, C., da, C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., de Souza, L.M., de Pinho, M.C., Araujo, L.J.T., Malta, F.S.V., de Lima, A. B., Silva, J., do, P., Zauli, D.A.G., Ferreira, A.C., de, S., Schnekenberg, R.P., Laydon, D.J., Walker, P.G.T., Schlüter, H.M., dos Santos, A.L.P., Vidal, M.S., Del Caro, V.S., Filho, R.M.F., dos Santos, H.M., Aguiar, R.S., Proença-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Prete, C.A., Nascimento, V.H., Suchard, M.A., Bowden, T.A., Pond, S.L.K., Wu, C.H., Ratmann, O., Ferguson, N.M., Dye, C., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M., do, P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science 372, 815–821. https://doi.org/10.1126/science.abh2644.

Ferrareze, P.A.G., Cybis, G.B., de Oliveira, L.F.V., Zimerman, R.A., Schiavon, D.E.B., Peter, C., Thompson, C.E., 2024. Intense P.1 (Gamma) diversification followed by rapid Delta substitution in Southern Brazil: a SARS-CoV-2 genomic epidemiology study. Microbes. Infect. 26, 105216. https://doi.org/10.1016/j.micinf.2023.105216.

Franceschi, V.B., Caldana, G.D., de Menezes Mayer, A., Cybis, G.B., Neves, C.A.M., Ferrareze, P.A.G., Demoliner, M., de Almeida, P.R., Gularte, J.S., Hansen, A.W., Weber, M.N., Fleck, J.D., Zimerman, R.A., Kmetzsch, L., Spilki, F.R., Thompson, C.E., 2021. Genomic epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil. BMC. Genomics. 22, 371. https://doi.org/10.1186/s12864-021-07708-w.

Franceschi, V.B., Caldana, G.D., Perin, C., Horn, A., Peter, C., Cybis, G.B., Ferrareze, P.A. G., Rotta, L.N., Cadegiani, F.A., Zimerman, R.A., Thompson, C.E., 2021. Predominance of the SARS-CoV-2 Lineage P.1 and Its Sublineage P.1.2 in patients from the metropolitan region of porto alegre, Southern Brazil in March 2021. Pathogens. 10, 988. https://doi.org/10.3390/pathogens10080988.

Francisco Jr, R., da, S., Benites, L.F., Lamarca, A.P., de Almeida, L.G.P., Hansen, A.W., Gularte, J.S., Demoliner, M., Gerber, A.L., de C Guimarães, A.P., Antunes, A.K.E., Heldt, F.H., Mallmann, L., Hermann, B., Ziulkoski, A.L., Goes, V., Schallenberger, K., Fillipi, M., Pereira, F., Weber, M.N., de Almeida, P.R., Fleck, J.D., Vasconcelos, A.T. R., Spilki, F.R., 2021. Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection by two different lineages in Rio Grande do Sul, Brazil. Virus. Res. 296, 198345. https://doi.org/10.1016/j.virusres.2021.198345.

Giovanetti, M., Fonseca, V., Wilkinson, E., Tegally, H., San, E.J., Althaus, C.L., Xavier, J., Nanev Slavov, S., Viala, V.L., Ranieri Jerônimo Lima, A., Ribeiro, G., Souza-Neto, J. A., Fukumasu, H., Lehmann Coutinho, L., Venancio da Cunha, R., Freitas, C., Campelo de A e Melo, C.F., Navegantes de Araújo, W., Do Carmo Said, R.F., Almiron, M., de Oliveira, T., Coccuzzo Sampaio, S., Elias, M.C., Covas, D.T., Holmes, E.C., Lourenço, J., Kashima, S., de Alcantara, L.C.J., 2022a. Replacement of the Gamma by the Delta variant in Brazil: impact of lineage displacement on the ongoing pandemic. Virus. Evol. 8, veac024. https://doi.org/10.1093/ve/veac024.

Giovanetti, M., Slavov, S.N., Fonseca, V., Wilkinson, E., Tegally, H., Patané, J.S.L., Viala, V.L., San, E.J., Rodrigues, E.S., Santos, E.V., Aburjaile, F., Xavier, J., Fritsch, H., Adelino, T.E.R., Pereira, F., Leal, A., Iani, F.C.de M., de Carvalho Pereira, G., Vazquez, C., Sanabria, G.M.E., Oliveira, E.C.de, Demarchi, L., Croda, J., dos Santos Bezerra, R., Paola Oliveira de Lima, L., Martins, A.J., Renata dos Santos Barros, C., Marqueze, E.C., de Souza Todao Bernardino, J., Moretti, D.B., Brassaloti, R.A., de Lello Rocha Campos Cassano, R., Mariani, P.D.S.C., Kitajima, J. P., Santos, B., Proto-Siqueira, R., Cantarelli, V.V., Tosta, S., Nardy, V.B., Reboredo de Oliveira da Silva, L., Gómez, M.K.A., Lima, J.G., Ribeiro, A.A., Guimarães, N.R., Watanabe, L.T., Barbosa Da Silva, L., da Silva Ferreira, R., da Penha, M.P.F., Ortega, M.J., de la Fuente, A.G., Villalba, S., Torales, J., Gamarra, M.L., Aquino, C., Figueredo, G.P.M., Fava, W.S., Motta-Castro, A.R.C., Venturini, J., do Vale Leone de Oliveira, S.M., Gonçalves, C.C.M., do Carmo Debur Rossa, M., Becker, G.N., Giacomini, M.P., Marques, N.Q., Riediger, I.N., Raboni, S., Mattoso, G., Cataneo, A. D., Zanluca, C., Duarte dos Santos, C.N., Assato, P.A., Allan da Silva da Costa, F., Poleti, M.D., Lesbon, J.C.C., Mattos, E.C., Banho, C.A., Sacchetto, L., Moraes, M.M., Grotto, R.M.T., Souza-Neto, J.A., Nogueira, M.L., Fukumasu, H., Coutinho, L.L., Calado, R.T., Neto, R.M., Bispo de Filippis, A.M., Venancio da Cunha, R., Freitas, C., Peterka, C.R.L., de Fátima Rangel Fernandes, C., Navegantes, W., do Carmo Said, R. F., Campelo de A e Melo, C.F., Almiron, M., Lourenço, J., de Oliveira, T., Holmes, E. C., Haddad, R., Sampaio, S.C., Elias, M.C., Kashima, S., Junior de Alcantara, L.C., Covas, D.T., 2022b. Genomic epidemiology of the SARS-CoV-2 epidemic in Brazil. Nat. Microbiol. 7, 1490–1500. https://doi.org/10.1038/s41564-022-01191-z.

Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Van Rompay, K.K.A., Isern, S., Michael, S.F., Coffey, L.L., Loman, N.J., Andersen, K.G., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 20, 8. https://doi.org/10.1186/s13059-018-1618-7.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010.

Harrison, A.G., Lin, T., Wang, P., 2020. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends. Immunol. 41, 1100–1115. https://doi.org/10.1016/j.it.2020.10.004.

Hill, V., Du Plessis, L., Peacock, T.P., Aggarwal, D., Colquhoun, R., Carabelli, A.M., Ellaby, N., Gallagher, E., Groves, N., Jackson, B., McCrone, J.T., O'Toole, Á., Price, A., Sanderson, T., Scher, E., Southgate, J., Volz, E., Barclay, W.S., Barrett, J.C., Chand, M., Connor, T., Goodfellow, I., Gupta, R.K., Harrison, E.M., Loman, N., Myers, R., Robertson, D.L., Pybus, O.G., Rambaut, A., 2022. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. Virus. Evol. 8, veac080. https://doi.org/10.1093/ve/veac080.

Hill, V., Ruis, C., Bajaj, S., Pybus, O.G., Kraemer, M.U.G., 2021. Progress and challenges in virus genomic epidemiology. Trends. Parasitol. 37, 1038–1049. https://doi.org/10.1016/j.pt.2021.08.007.

Hoffmann, M., Arora, P., Groß, R., Seidel, A., Hörnich, B.F., Hahn, A.S., Krüger, N., Graichen, L., Hofmann-Winkler, H., Kempf, A., Winkler, M.S., Schulz, S., Jäck, H.M., Jahrsdörfer, B., Schrezenmeier, H., Müller, M., Kleger, A., Münch, J., Pöhlmann, S., 2021. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. Cell 184, 2384–2393. https://doi.org/10.1016/j.cell.2021.03.036 e12.

Hu, Zhiliang, Song, C., Xu, C., Jin, G., Chen, Y., Xu, X., Ma, H., Chen, W., Lin, Y., Zheng, Y., Wang, J., Hu, Zhibin, Yi, Y., Shen, H., 2020. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing. China. Sci. China Life Sci. 63, 706–711. https://doi.org/10.1007/s11427-020-1661-4.

Huddleston, J., Hadfield, J., Sibley, T.R., Lee, J., Fay, K., Ilcisin, M., Harkins, E., Bedford, T., Neher, R.A., Hodcroft, E.B., 2021. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. J. Open Source Softw. 6, 2906. https://doi.org/10.21105/joss.02906.

Itokawa, K., Sekizuka, T., Hashino, M., Tanaka, R., Kuroda, M., 2020. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. PLoS. One 15, e0239403. https://doi.org/10.1371/journal.pone.0239403.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

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

Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., Akite, N., Ho, J., Lee, R.T., Yeo, W., Team, G.C.C., Maurer-Stroh, S., 2021. GISAID's role in pandemic response. China CDC. Wkly. 3, 1049–1051. https://doi.org/10.46234/ccdcw2021.255.

Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinforma. Oxf. Engl. 23, 127–128. https://doi.org/10.1093/bioinformatics/btl529.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Martins, A.F., Zavascki, A.P., Wink, P.L., Volpato, F.C.Z., Monteiro, F.L., Rosset, C., De-Paris, F., Ramos, Á.K., Barth, A.L., 2021. Detection of SARS-CoV-2 lineage P.1 in patients from a region with exponentially increasing hospitalisation rate, February 2021, Rio Grande do Sul, Southern Brazil. Eurosurveillance 26, 2100276. https://doi.org/10.2807/1560-7917.ES.2021.26.12.2100276.

Mayer, A., de, M., Gröhs Ferrareze, P.A., de Oliveira, L.F.V., Gregianini, T.S., Neves, C.L. A.M., Caldana, G.D., Kmetzsch, L., Thompson, C.E., 2023. Genomic characterization and molecular evolution of SARS-CoV-2 in Rio Grande do Sul State. Brazil. Virology 582, 1–11. https://doi.org/10.1016/j.virol.2023.03.005.

McCrone, J.T., Hill, V., Bajaj, S., Pena, R.E., Lambert, B.C., Inward, R., Bhatt, S., Volz, E., Ruis, C., Dellicour, S., Baele, G., Zarebski, A.E., Sadilek, A., Wu, N., Schneider, A., Ji, X., Raghwani, J., Jackson, B., Colquhoun, R., O'Toole, Á., Peacock, T.P., Twohig, K., Thelwall, S., Dabrera, G., Myers, R., Faria, N.R., Huber, C., Bogoch, I.I., Khan, K., du Plessis, L., Barrett, J.C., Aanensen, D.M., Barclay, W.S., Chand, M., Connor, T., Loman, N.J., Suchard, M.A., Pybus, O.G., Rambaut, A., Kraemer, M.U.G., 2022. Context-specific emergence and growth of the SARS-CoV-2 Delta variant. Nature 610, 154–160. https://doi.org/10.1038/s41586-022-05200-3.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mir, D., Rego, N., Resende, P.C., Tort, F., Fernández-Calero, T., Noya, V., Brandes, M., Possi, T., Arleo, M., Reyes, N., Victoria, M., Lizasoain, A., Castells, M., Maya, L., Salvo, M., Schäffer Gregianini, T., Mar da Rosa, M.T., Garay Martins, L., Alonso, C., Vega, Y., Salazar, C., Ferrés, I., Smircich, P., Sotelo Silveira, J., Fort, R.S., Mathó, C., Arantes, I., Appolinario, L., Mendonça, A.C., Benítez-Galeano, M.J., Simoes, C., Graña, M., Motta, F., Siqueira, M.M., Bello, G., Colina, R., Spangenberg, L., 2021. Recurrent dissemination of SARS-CoV-2 Through the Uruguayan–Brazilian border. Front. Microbiol. 12. https://doi.org/10.3389/fmicb.2021.653986.

Mohammed, I., Nauman, A., Paul, P., Ganesan, S., Chen, K.H., Jalil, S.M.S., Jaouni, S.H., Kawas, H., Khan, W.A., Vattoth, A.L., Al-Hashimi, Y.A., Fares, A., Zeghlache, R., Zakaria, D., 2022. The efficacy and effectiveness of the COVID-19 vaccines in reducing infection, severity, hospitalization, and mortality: a systematic review. Hum. Vaccines Immunother.

Morel, B., Barbera, P., Czech, L., Bettisworth, B., Hübner, L., Lutteropp, S., Serdari, D., Kostaki, E.G., Mamais, I., Kozlov, A.M., Pavlidis, P., Paraskevis, D., Stamatakis, A., 2021. Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult. Mol. Biol. Evol. 38, 1777–1791. https://doi.org/10.1093/molbev/msaa314.

Naveca, F.G., Nascimento, V., de Souza, V.C., Corado, A., de, L., Nascimento, F., Silva, G., Costa, Á., Duarte, D., Pessoa, K., Mejía, M., Brandão, M.J., Jesus, M., Gonçalves, L., da Costa, C.F., Sampaio, V., Barros, D., Silva, M., Mattos, T., Pontes, G., Abdalla, L., Santos, J.H., Arantes, I., Dezordi, F.Z., Siqueira, M.M., Wallau, G.L., Resende, P.C., Delatorre, E., Gräf, T., Bello, G., 2021. COVID-19 in Amazonas, Brazil, was driven by

the persistence of endemic lineages and P.1 emergence. Nat. Med. 27, 1230–1238. https://doi.org/10.1038/s41591-021-01378-7.

Nonaka, C.K.V., Franco, M.M., Gräf, T., Barcia, C.A. de L., Mendonça, R.N. de Á., Sousa, K.A.F. de, Neiva, L.M.C., Fosenca, V., Mendes, A.V.A., Aguiar, R.S. de, Giovanetti, M., Souza, B.S. de F., 2021. Genomic Evidence of SARS-CoV-2 Reinfection Involving E484K Spike Mutation, Brazil - Volume 27, Number 5—May 2021 - Emerging Infectious Diseases journal - CDC. https://doi.org/10.3201/eid2705.210191.

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J.T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S.W., Aanensen, D.M., Holmes, E.C., Pybus, O.G., Rambaut, A., 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus. Evol. 7, veab064. https://doi.org/10.1093/ve/veab064.

Quick, J., 2020. nCoV-2019 sequencing protocol.

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

Rego, N., Costábile, A., Paz, M., Salazar, C., Perbolianachis, P., Spangenberg, L., et al., 2021. Real-time genomic surveillance for SARS-CoV-2 variants of concern. Uruguay. Emerg Infect Dis. 27 (11), 2957–2960. https://doi.org/10.3201/eid2711.211198.

Rhoden, J., Hoffmann, A.T., Stein, J.F., da Silva, M.S., Gularte, J.S., Filippi, M., Demoliner, M., Girardi, V., Spilki, F.R., Fleck, J.D., Rigotto, C., 2024. Diversity of Omicron sublineages and clinical characteristics of hospitalized patients in the southernmost state of Brazil. BMC Infect. Dis. 24, 193. https://doi.org/10.1186/s12879-024-09089-3.

Robishaw, J.D., Alter, S.M., Solano, J.J., Shih, R.D., DeMets, D.L., Maki, D.G., Hennekens, C.H., 2021. Genomic surveillance to combat COVID-19: challenges and opportunities. Lancet Microbe 2, e481–e484. https://doi.org/10.1016/S2666-5247(21)00121-X.

Sagulenko, P., Puller, V., Neher, R.A., 2018. TreeTime: maximum-likelihood phylodynamic analysis. Virus Evol. 4, vex042. https://doi.org/10.1093/ve/vex042.

shiquan, 2024. shiquan/bamdst.

Smith, Kyle, Ye, Cheng, Turakhia, Yatish, 2023. Tracking and curating putative SARS-CoV-2 recombinants with RIVET. Bioinformatics. https://doi.org/10.1093/bioinformatics/btad538.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, N., Korsman, S., Davies, M.A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature 592, 438–443. https://doi.org/10.1038/s41586-021-03402-9.

Thompson Mark, G., Edward, Stenehjem, Shaun, Grannis, Ball Sarah W, Naleway Allison, L., Ong Toan, C., DeSilva Malini, B., Karthik, Natarajan, Bozio Catherine, H., Ned, Lewis, Kristin, Dascomb, Dixon Brian, E., Birch Rebecca, J., Irving Stephanie, A., Suchitra, Rao, Elyse, Kharbanda, Jungmi, Han, Sue, Reynolds, Kristin, Goddard, Nancy, Grisel, F, Fadel William, E, Levy Matthew, Jill, Ferdinands, Bruce, Fireman, Julie, Arndorfer, Valvi, Nimish R., Rowley, Elizabeth A., Patel, Palak, Zerbo, Ousseny, Griggs, Eric P., Porter, Rachael M., Maria, Demarco, Lenee, Blanton, Andrea, Steffens, Yan, Zhuang, Natalie, Olson, Michelle, Barron, Patricia, Shifflett, Schrag Stephanie, J., Verani Jennifer, R., Alicia, Fry, Manjusha, Gaglani, Eduardo, Azziz-Baumgartner, Klein Nicola, P., 2021. Effectiveness of Covid-19 vaccines in ambulatory and inpatient care settings. N. Engl. J. Med. 385, 1355–1371. https://doi.org/10.1056/NEJMoa2110362.

Tracking SARS-CoV-2 variants. https://www.who.int/activities/tracking-SARS-CoV-2-variants (accessed 4.22.24).

Vaccines – COVID19 Vaccine Tracker. https://covid19.trackvaccines.org/vaccines/#approved (accessed 4.22.24).

Viana, R., Moyo, S., Amoako, D.G., Tegally, H., Scheepers, C., Althaus, C.L., Anyaneji, U. J., Bester, P.A., Boni, M.F., Chand, M., Choga, W.T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., Giovanetti, M., Hardie, D., Hill, V., Hsiao, N.Y., Iranzadeh, A., Ismail, A., Joseph, C., Joseph, R., Koopile, L., Kosakovsky Pond, S.L., Kraemer, M.U.G., Kuate-Lere, L., Laguda-Akingba, O., Lesetedi-Mafoko, O., Lessells, R.J., Lockman, S., Lucaci, A.G., Maharaj, A., Mahlangu, B., Maponga, T., Mahlakwane, K., Makatini, Z., Marais, G., Maruapula, D., Masupu, K., Matshaba, M., Mayaphi, S., Mbhele, N., Mbulawa, M.B., Mendes, A., Mlisana, K., Mnguni, A., Mohale, T., Moir, M., Moruisi, K., Mosepele, M., Motsatsi, G., Motswaledi, M.S., Mphoyakgosi, T., Msomi, N., Mwangi, P.N., Naidoo, Y., Ntuli, N., Nyaga, M., Olubayo, L., Pillay, S., Radibe, B., Ramphal, Y., Ramphal, U., San, J.E., Scott, L., Shapiro, R., Singh, L., Smith-Lawrence, P., Stevens, W., Strydom, A., Subramoney, K., Tebeila, N., Tshiabuila, D., Tsui, J., van Wyk, S., Weaver, S., Wibmer, C.K., Wilkinson, E., Wolter, N., Zarebski, A.E., Zuze, B., Goedhals, D., Preiser, W., Treurnicht, F., Venter, M., Williamson, C., Pybus, O.G., Bhiman, J., Glass, A., Martin, D.P., Rambaut, A., Gaseitsiwe, S., von Gottberg, A., de Oliveira, T., 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature 603, 679–686. https://doi.org/10.1038/s41586-022-04411-y.

WHO Director, General's opening remarks at the media briefing on COVID-19 - 11 March 2020 https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020 (accessed 4.22.24).

Y Castro, T.R., Piccoli, B.C., Vieira, A.A., Casarin, B.C., Tessele, L.F., Salvato, R.S., Gregianini, T.S., Martins, L.G., Resende, P.C., Pereira, E.C., Moreira, F.R.R., de Jesus, J.G., Seerig, A.P., Lobato, M.A.O., de Campos, M.M.A., Goularte, J.S., da Silva, M.S., Demoliner, M., Filippi, M., Pereira, V.M.A.G., Schwarzbold, A.V., Spilki, F.R., Trindade, P.A., 2023. Introduction, Dispersal, and Predominance of SARS-CoV-2 Delta Variant in Rio Grande do Sul. A Retrosp. Analy.. Microorgan. 11, 2938. https://doi.org/10.3390/microorganisms11122938.

Yanes-Lane, M., Winters, N., Fregonese, F., Bastos, M., Perlman-Arrow, S., Campbell, J. R., Menzies, D., 2020. Proportion of asymptomatic infection among COVID-19 positive persons and their transmission potential: a systematic review and meta-analysis. PLoS. One 15, e0241536. https://doi.org/10.1371/journal.pone.0241536.