UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE – UFCSPA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

Gabriel Dickin Caldana

A TRAJETÓRIA DO SARS-COV-2: RETROSPECTIVA EPIDEMIOLÓGICA NO BRASIL E ANÁLISE GENÔMICA NO RIO GRANDE DO SUL

Gabriel Dickin Caldana

A trajetória do SARS-CoV-2: retrospectiva epidemiológica no Brasil e análise genômica no Rio Grande do Sul

Dissertação submetida ao Programa de Pós-Graduação em Ciências da Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre como requisito para a obtenção do grau de Mestre

Orientadora: Profa. Dra. Claudia Elizabeth Thompson

Coorientador: Prof. Dr. Vlademir Vicente Cantarelli

Catalogação na Publicação

Caldana, Gabriel Dickin
A trajetória do SARS-CoV-2: retrospectiva epidemiológica no Brasil e análise genômica no Rio Grande do Sul / Gabriel Dickin Caldana. -- 2022.
145 p. : il., graf., tab. ; 30 cm.

Dissertação (mestrado) -- Universidade Federal de Ciências da Saúde de Porto Alegre, Programa de Pós-Graduação em Ciências da Saúde, 2022.

Orientador(a): Profa. Dra. Claudia Elizabeth Thompson ; coorientador(a): Prof. Dr. Vlademir Vicente Cantarelli.

1. SARS-CoV-2. 2. Linhagens. 3. Genômica. 4. Sequenciamento. 5. Epidemiologia. I. Título.

AGRADECIMENTOS

Gostaria de expressar minha profunda e genuína gratidão pelas pessoas que, direta ou indiretamente, fizeram parte deste trabalho.

À minha esposa Marília, brilhante, exemplar e exitosa pesquisadora, que me inspirou na decisão de realizar o Mestrado, alumiou meus caminhos me dando condições psicológicas de seguir em frente em tempos de angústia e tormenta. Entre tantas coisas, me ensinou a escrever melhor, a melhorar meu pensamento crítico e a ver o mundo por diferentes perspectivas. Obrigado pelo teu carinho e apoio incondicionais.

À minha família, Susi, Marcelo e Silas, fonte inesgotável de amor, que desde criança me ensinou valores e o valor da educação. Apoio e amparo desde o jardim de infância me ajudaram em todo percurso profissional e acadêmico.

À minha orientadora professora Claudia, próspera pesquisadora e grande entusiasta da Ciência. Agradeço por ter me acolhido, ensinado, incluído em diversos projetos - por meio dos quais pude conhecer e colaborar com colegas e outros professores - e me ajudar a crescer como cientista.

Ao me coorientador professor Vlademir, pesquisador experiente que me acolheu e me proporcionou excepcionais experiências de bancada com diagnóstico molecular de COVID-19.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da minha bolsa de Mestrado. Ao Programa de Pós-graduação em Ciências da Saúde (PPGCS) e seu coordenador professor Pedro Romão pelos seus incansáveis esforços para esclarecer dúvidas, planejar disciplinas e atividades, honrando a excelência.

Aos membros da banca examinadora desta dissertação, Prof. Dr. Rafael Andrade Caceres, Profa. Dra. Lívia Kmetzsch e Prof. Dr. Luís Fernando Saraiva Macedo Timmers, pelo tempo aplicado na leitura deste trabalho e pelos apontamentos e sugestões para aprimoramento da dissertação.

Aos meus colegas do grupo de pesquisa, especialmente Patrícia Ferrareze, Vinícius Francheschi - pela enorme ajuda com o aprendizado da bioinformática, com a escrita de artigos e imensa paciência -, Amanda Mayer e Janini Paiz pelo coleguismo no projeto GPS COVID de Esteio, auxílio na coleta de amostras, dados clínicos, nas análises e na escrita de manuscritos.

À prefeitura de Esteio e entidades financiadoras. Aos hospitais e laboratórios que contribuíram com a coleta e sequenciamento das amostras utilizadas neste trabalho. Aos pesquisadores voluntários do Projeto Brasil.IO, aos pesquisadores ao redor do mundo e a rede de colaboração para compartilhamento de dados (*Global Initiative on Sharing Avian Influenza Data*) que proporcionou dados preciosos analisados nessa dissertação.

RESUMO

A pandemia de COVID-19, causada pelo SARS-CoV-2, se espalhou rapidamente pelo mundo, após sua descoberta no final de 2019. Desde então, esforços internacionais de pesquisa multidisciplinar têm contribuído com o combate do avanço do vírus. Diante disso, essa dissertação apresenta dois estudos que têm como objetivo a análise de dados genômicos do Rio Grande do Sul (RS) a fim de identificar novas linhagens e mutações importantes e de dados epidemiológicos para pormenorizar o histórico da pandemia no país. No primeiro, sequenciamos 56 amostras de 13 municípios do Rio Grande do Sul coletadas durante a segunda onda pandêmica em março de 2021. O trabalho confirmou a predominância da variante Gamma no estado, sua difusão a partir de múltiplas introduções virais e o aparecimento da sublinhagem P1.2. No segundo, analisamos as sequências brasileiras disponíveis na plataforma GISAID - até 7 de dezembro de 2021, 76.413 genomas e seus metadados - juntamente com as informações demográficas e epidemiológicas extraídas da base de dados do IBGE do projeto Brasil.IO. Assim, identificamos as proporções de casos e mortes em cada região brasileira, a ascensão e declínio das principais linhagens e variantes e sua incidência sobre a população. Até o momento, a variante Gamma se trata da mais letal versão do SARS-CoV-2 na história da pandemia no país, visto que em todas as regiões foi a aparente causa do recorde no número de casos e mortes por COVID-19. Utilizando genomas completos, metadados e dados epidemiológicos e demográficos, contribuimos para o entendimento da evolução e do comportamento do SARS-CoV-2. Promissoramente, este trabalho tem o potencial de guiar o desenvolvimento de intervenções não medicamentosas e possíveis soluções para prevenção da doença e proteção da população.

Palavras-Chave: SARS-CoV-2; COVID-19; Variantes; Linhagens; Genômica; Epidemiologia; Sequenciamento; Biologia molecular

ABSTRACT

The COVID-19 pandemic, caused by SARS-CoV-2, spread rapidly around the world after its discovery in late 2019. International multidisciplinary research efforts have contributed to combating the spread of the virus. Therefore, this dissertation presents two studies aiming to analyze genomic data from Rio Grande do Sul (RS) in order to identify new lineages and important mutations and epidemiological data to detail the history of the pandemic in the country. In the first study, we sequenced 56 samples from 13 municipalities in Rio Grande do Sul collected during the second pandemic wave in March 2021. The work confirmed the predominance of Gamma in the state, its diffusion from multiple viral introductions and the appearance of the P1.2 sublineage. In the second one, we analyzed all the Brazilian sequences available on the GISAID platform - as of December 7th, 2021, 76,413 genomes and their metadata - together with the demographic and epidemiological information extracted from the IBGE database and the Brasil.IO project. We identified the proportions of cases and deaths in each Brazilian region, the rise and decline of the main lineages and variants and their incidence on the population. The Gamma variant seems to be the most lethal version of SARS-CoV-2 in the history of the pandemic in the country so far, since in all regions it was the apparent cause of the record number of cases and deaths by COVID-19. Using complete genomes, metadata, and epidemiological and demographic data, we aim to contribute to the understanding of the evolution and behavior of SARS-CoV-2. Promisingly, this work can help guide the development of non-pharmaceutical interventions and possible solutions to prevent the disease and protect population.

Keywords: SARS-CoV-2; COVID-19; Variants; Lineages; Genomics; Epidemiology; Sequencing; Molecular biology

LISTA DE FIGURAS E TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ATP - Adenosina Trifosfato

cDNA - DNA complementar

CoV - Coronavírus

COVID-19 - *Coronavirus Disease 2019*

Ct - *Cycle threshold*

DNA - *Desoxiribonucleic acid*

dsDNA - *double-stranded DNA*

E - Proteína do envelope viral

GISAID - *Global Initiative on Sharing Avian Influenza Data*

hACE2 - *human Angiotensin-Converting Enzyme 2*

IBV - *Infectious Bronchitis Virus*

ICTV - *International Committee on Taxonomy of Viruses*

IgG - Imunoglobulina G

IgM - Imunoglobulina M

ILP - Imunocromatografia Lateral em Papel

LAMP - *Loop-mediated isothermal amplification*

M - Proteína da membrana viral

MHV - Murine Hepatitis Virus

mRNA - RNA mensageiro

N - Proteína do nucleocapsídeo

NGS - *Next Generation Sequencing*

NIAID - *National Institute of Allergy and Infectious Diseases*

Nsp - non-structural protein

NTD - *N-Terminal Domain*

OMS - Organização Mundial da Saúde

ORF - *Open Reading Frame*

RBD - *Receptor Binding Domain*

RdRp - *RNA-dependent RNA polymerase*

RNA - *Ribonucleic acid*

RT - *Reverse Transcriptase*

RTC - *Replication-Transcription Complex*

RT-qPCR - *Reverse Transcriptase quantitative Polymerase Chain Reaction*

S - Proteína Spike

SARS-CoV-2 - *Severe Acute Respiratory Syndrome Coronavirus 2*

TC - Tomografia Computadorizada

TMPRSS2 - *Transmembrane Serine Protease 2*

UTR - *Untranslated Region*

VOC - *Variant of Concern*

VOI - *Variant of Interest*

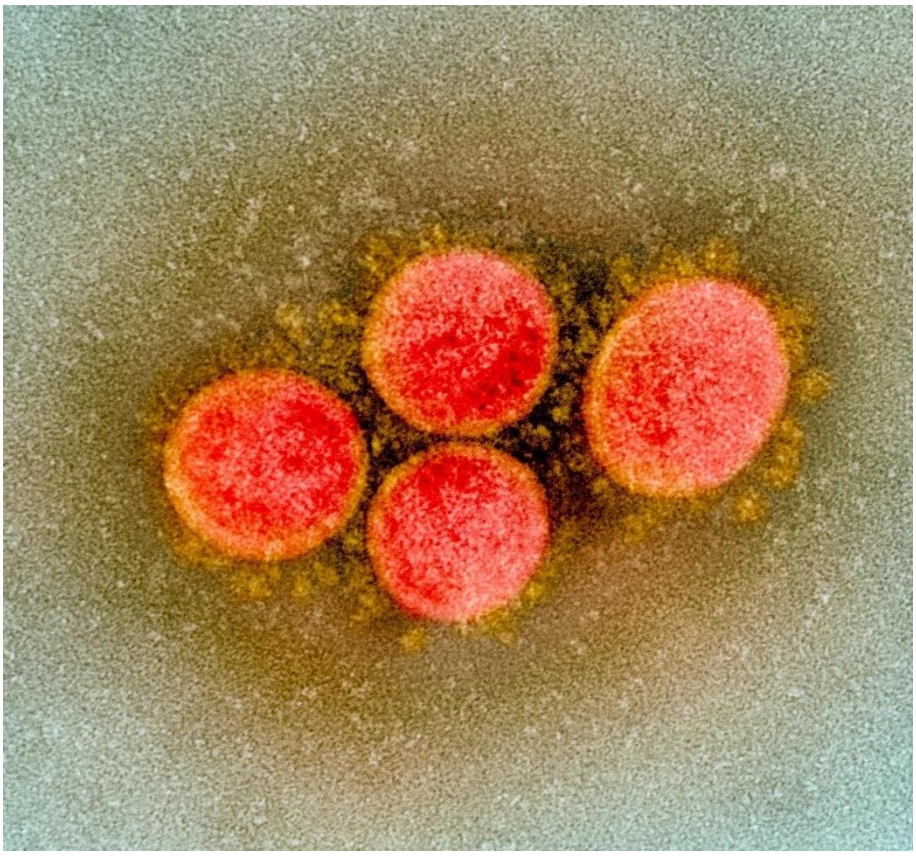
SUMÁRIO

1 REVISÃO DA LITERATURA

1.1 Coronavírus da Síndrome Respiratória Aguda Grave 2 (SARS-CoV-2)

A taxonomia do SARS-CoV-2 (Figura 1) pode ser descrita da seguinte forma (Coronaviridae Study Group Of The International Committee On Taxonomy Of Viruses, 2020): Reino *Riboviria*; Ordem *Nidovirales*; Subordem *Cornidovineae*; Família *Coronaviridae*; Subfamília *Orthocoronavirinae*; Gênero *Betacoronavirus*; Subgênero *Sarbecovirus*; Espécie *Severe Acute Respiratory Syndrome-Related Coronavirus*.

Figura 1: Imagem de microscopia eletrônica de transmissão de partículas de SARS-CoV-2 isoladas de um paciente. Imagem capturada e aprimorada em cores no *Integrated Research Facility* em Fort Detrick, Maryland, *National Institute of Allergy and Infectious Diseases* (NIAID) Fonte: https://www.niaid.nih.gov/news-events/novel-coronavirus-sarscov2-images

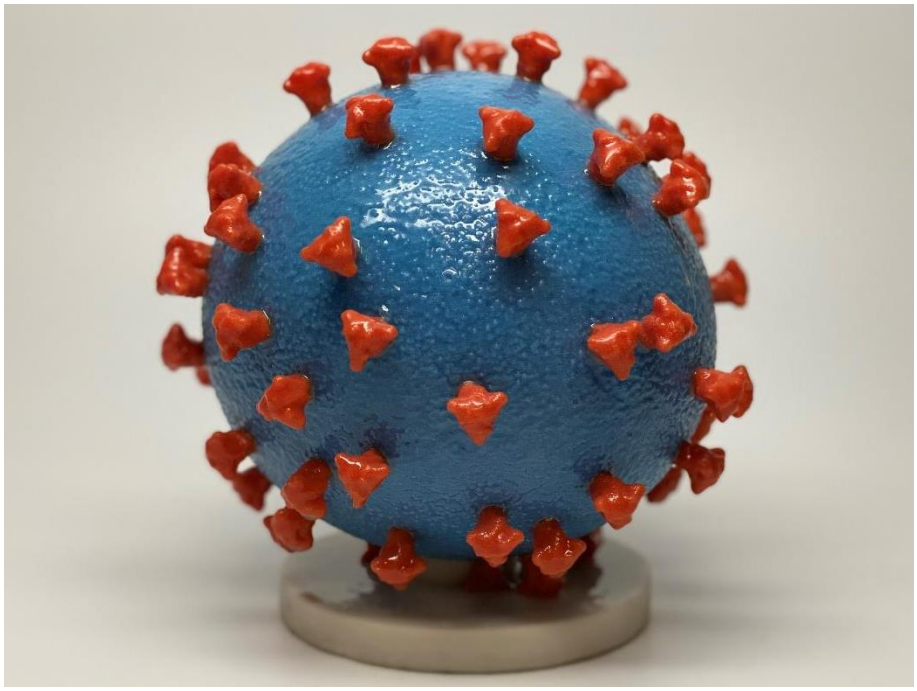


1.2 Família *Coronaviridae*

Os coronavírus (CoVs), nomeados por sua superfície em forma de coroa, são um grupo de vírus geneticamente diverso e que pode infectar várias espécies de animais, incluindo morcegos, porcos, espécies de aves, gatos, roedores e humanos (Weiss, 2020). Apenas os CoVs alfa e beta são conhecidos por infectar humanos e causam principalmente doenças respiratórias e entéricas com uma variedade de sintomas e largo espectro de severidade (Corman, 2018; Coronaviridae Study Group of the ICTV, 2020)

São vírus envelopados medindo de 80 a 160nm de diâmetro e com um genoma de RNA positivo. A associação do genoma com a proteína N forma o nucleocapsídeo em forma de hélice. O envelope compreende uma membrana bicamada lipídica e proteínas de superfície que envolvem toda a partícula viral. Dentre elas, a proteína Spike é responsável pela morfologia externa do coronavírus (Figura 1 e 2) (de Wilde, 2017; Murray, 2009), sendo o elemento que mais chama a atenção dos pesquisadores por ter relação direta com a entrada do vírus nas células animais. A síntese de proteínas ocorre em duas fases. Primeiramente, ocorre a tradução de uma poliproteína que é clivada formando uma RNA polimerase dependente de RNA. Subsequentemente, a polimerase gera uma fita de RNA sentido negativo que é usada como molde para replicar RNA mensageiro (mRNA) (Murray, 2009; de Wilde, 2017; Murgolo, 2021; V'kovski, 2021).

Figura 2: Modelo de um SARS-CoV-2. A superfície do vírus (azul) é coberta com proteínas spike (vermelho). As espículas na superfície dos coronavírus dão o nome a essa família de vírus – corona, que é latim para “coroa”, e a maioria dos coronavírus têm aparência semelhante. Fonte: NIH (https://www.niaid.nih.gov/news-events/novel-coronavirus-sarscov2-images)



O genoma de RNA dos CoVs contém pouco menos de 30 kb e codifica 14 *Open Reading Frame* (ORF), algumas das quais se sobrepõem (Snijder, 2003). Possui uma cauda poli A e sequências curtas de regiões não traduzidas (UTR) que formam estruturas regulatórias e uma única sequência líder próxima à extremidade 5’. Uma sequência reguladora da transcrição precede a maioria das ORFs. A extremidade proximal 5’ do genoma contém duas grandes ORFs: ORF1a e ORF1b que juntos resultam em uma poliproteína que compreende aproximadamente dois terços do genoma, que é clivada em proteínas não estruturais (nsp) pelas proteases nsp5 e nsp3 codificadas pelo ORF1a (Angelini, 2013; Snijder, 2003). Além das funções de protease, as nsps têm funções específicas na replicação e na modulação do ambiente da célula hospedeira, ancorando os complexos de replicação viral a domínios celulares e conduzindo a transcrição de mRNA. O terço restante do genoma contém os genes das proteínas estruturais comuns a todos os CoVs - Spike (S), Envelope (E), glicoproteína de Membrana (M) e Nucleocapsídeo (N) (Robson, 2020).

A replicação dos vírus de RNA normalmente está associada a uma alta taxa de erro que resulta em diversas populações de mutantes genômicos (Denison, 2011). Embora essa baixa fidelidade permita que os vírus de RNA se adaptem a diferentes ambientes e pressões seletivas, também está associada a uma maior chance de erros levarem à extinção de uma população. Isso sugere a necessidade de um equilíbrio entre a diversidade de mutantes e a aptidão para virulência e evolução (Denison, 2011; Gorbalenya, 2006; Smith, 2012). Esse fator desafia o desenvolvimento de antivirais contra CoVs e outros vírus de RNA, já que podem levar rapidamente ao desenvolvimento de resistência a fármacos (Perales, 2015). Outra barreira é a função de *proofreading* das exonucleases representada nos CoVs pela nsp14 (Chen, 2009). Além disso, essas proteínas também estão envolvidas em vários outros processos do ciclo de vida e patogenicidade do vírus, incluindo respostas imunes inatas e recombinação do genoma viral (Becares, 2016; Gribble, 2021) Os genomas dos vírus da ordem *Nidovirales* são os maiores e mais complexos entre os vírus de RNA conhecidos (Saberi, 2018) e a nsp14 é altamente conservada dentro da família *Coronaviridae* (Ma, 2015). A função de *proofreading* pode ter sido crucial para garantir a competência de replicação dos CoVs e na expansão e manutenção dos gêneros e das espécies hoje conhecidas (Gorbalenya, 2006).

1.3 Histórico dos Coronavírus

Uma publicação em 1931 indicou que havia uma nova doença respiratória que acometia primariamente galinhas de duas a três semanas de idade (Schalk e Hawn, 1931). Os autores se referiram à doença como "uma doença respiratória aparentemente nova". A infecção era muito contagiosa e altamente virulenta e incluía sintomas como falta de ar severa e fraqueza física. Era facilmente transmitida por contato direto entre aves ou por transferência de exsudatos brônquicos de galinhas infectadas para galinhas saudáveis - como forma de experimentação. (Fabricant, 1998) Em 1933, a doença foi reproduzida em galinhas jovens pela injeção intratraqueal, subcutânea e intraperitoneal de material tratado em filtro de Berkefeld - filtro de cerâmica largamente utilizado em laboratórios de microbiologia. Algumas empresas afirmam que esse filtro

possui a capacidade de filtrar até 100% das partículas acima de 0,9µm a até 98% das partículas acima de 0,5µm de diâmetro. Os animais desenvolveram sintomas típicos já descritos após vários períodos de incubação e recebimento do filtrado (Bushnell e Brandly, 1933; Beach, 1931). Foi sugerido que o patógeno não poderia ser bactéria ou protozoário, pois acreditava-se que estes seriam retidos no processo de filtragem. Em 1937, Beaudette e Hudson cultivaram o então nomeado vírus da bronquite infecciosa (IBV) pela primeira vez em embriões de galinha (Beaudette e Hudson, 1937). Esta cepa ficou conhecida como “cepa Beaudette” e, 50 anos mais tarde, foi o primeiro coronavírus a ter seu genoma sequenciado (Boursnell, 1987).

Os coronavírus humanos foram descobertos como um dentre muitos vírus causadores do resfriado comum. A pesquisa sobre o estudo do resfriado comum teve origem no Reino Unido com a *Common Cold Research Unit* em 1946 (*Research into the Common Cold*, 1946). O laboratório descobriu vários vírus, como vírus influenza, vírus parainfluenza e rinovírus (Andrewes, 1966). Em um experimento entre 1960 e 1961, a equipe coletou amostras de garganta de 170 meninos em idade escolar com sintomas de resfriado em um internato na Inglaterra. Entre as poucas amostras que não puderam ser cultivadas, uma amostra em especial - que foi testada com todos os métodos de cultura bacteriana e viral disponíveis e resultaram negativos - foi designada B814 (Kendall, 1962). Em 1965, puderam confirmar que o patógeno era um vírus pois: (I) passou pelo filtro Berkefeld; (II) era suscetível ao tratamento com éter (indicando a presença de envelope viral); (III) era capaz de induzir resfriado em voluntários tratados com antibióticos (indicando que não era uma bactéria); (IV) era cultivado em cultura de células epiteliais de traqueia humana. Testes sorológicos indicaram ainda que o vírus não estava relacionado a nenhum vírus conhecido na época (Tyrrell e Bynoe, 1965).

Foi somente com o desenvolvimento da microscopia eletrônica que os vírus puderam ser visualizados e estruturalmente estudados. Cientistas da Universidade de Maryland foram os primeiros a descrever a estrutura do coronavírus usando a microscopia eletrônica de transmissão. Em 1948, relataram que o IBV tinha forma esférica e alguns deles tinham projeções filamentosas (Reagan, 1948). No entanto, as imagens eram difíceis de

interpretar devido à baixa resolução e baixa ampliação da tecnologia da época (25.000x). Outro estudo que utilizou microscopia eletrônica observou IBVs como estruturas em forma de anel (Domermuth e Edwards, 1957). Imagens microscópicas com uma resolução um pouco melhor foram publicadas em 1964. Quatro cepas de IBV, incluindo a cepa Beaudette, foram comparadas com o vírus da gripe. Em todas as cepas de IBV foram observadas projeções em forma de pêra, que eram chamadas de *Spikes* (do inglês “espículas” ou “espinhos”) (Berry, 1964). Em 1965, pesquisadores também estudaram a morfologia do vírus da hepatite murina (MHV) - outro coronavírus identificado anos antes - afirmando que a superfície da partícula viral é coberta por espículas (David-Ferreira e Manaker, 1965). Almeida e Tyrrell publicaram em 1967 imagens de dois vírus respiratórios humanos - 229E e B814 - e o IBV e concluíram que são morfologicamente idênticos (Almeida e Tyrrell, 1967). No mesmo ano, outro estudo também relatou a estrutura dos vírus do resfriado comum coletado de trabalhadores do *National Intitute of Health* nos EUA durante 1965-1966. Descobriram que seis de suas amostras tinham características morfológicas comuns com B814 e compararam a estrutura de uma de suas amostras com as do 229E, IBV e influenza. A semelhança com IBV foi tão relevante que os chamaram de "vírus semelhantes ao IBV". Neste estudo, também foi relatada a presença de espículas na superfície viral, como no trabalho de June Almeida (McIntosh, 1967). Em meados de 1967, foi reconhecido que IBV, MHV, B814 e 229E eram estrutural e biologicamente semelhantes de tal modo que formam um grupo distinto (Tyrrell e Almeida, 1967; Becker, 1967). Almeida sugeriu nome "Coronavírus" devido à franja de projeções - s*pikes* ou espículas - na superfície viral que lembra uma coroa solar. Assim o vírus foi alcunhado em um breve artigo publicado na revista *Nature* em 1968 (Almeida, 1968).

Com o passar dos anos, algumas mudanças de nomenclatura e de taxonomia foram revisadas. Em 1975, “*Coronaviridae*” foi adotado oficialmente como o nome da família pelo Comitê Internacional de Taxonomia de Vírus (ICTV) (ICTV, 1975). Em 2009, a família passou por algumas reformas de nomenclatura. O IBV foi oficialmente designado como Vírus da Bronquite Infecciosa Aviária, mas renomeado para Coronavírus Aviário em ((a) ICTV, 2009); O MHV foi renomeado como Coronavírus Murino ((b) ICTV, 2009); Cepas isoladas de

humanos foram chamadas coletivamente de Coronavírus respiratório humano e foram unidas com Coronavírus humano 229E (HCoV-229E) ((c) ICTV, 2009). Devido ao crescente número e diversidade de novas espécies descobertas, o ICTV dividiu o gênero *Coronavirus* em quatro gêneros, *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus* e *Gammacoronavirus* (Carstens, 2010).

1.4 Estrutura e replicação

O SARS-CoV-2 é um vírus de RNA de fita simples com um genoma de aproximadamente 29.900 nucleotídeos que contém 11 estruturas de leitura aberta (ORFs) e codifica 27 proteínas (Figura 3). Cerca de 70% do genoma é coberto pela sobreposição dos genes ORF1a e ORF1b, que são traduzidos em poliproteínas formando 16 proteínas não estruturais (nsp1 a nsp16). O último terço do genoma codifica as proteínas estruturais (Spike [S], Envelope [E], Membrana [M] e Nucleocapsídeo [N]) e seis proteínas acessórias (ORF3a, ORF6, ORF7a, ORF7b, ORF8 e ORF10) (Helmy, 2020). A proteína N fosforilada e o RNA genômico juntos formam o nucleocapsídeo que é envolto por bicamada fosfolipídica que possui as proteínas M, S e E. M é uma proteína transmembrana responsável pelo transporte de moléculas e formação do envelope. A proteína E é necessária para a montagem e liberação do vírus, possui atividade de canal iônico e é relatada como tendo papel na patogênese do vírus (Kaur, 2021). N é a única proteína do nucleocapsídeo e é composta por dois domínios – ambos necessários para a ligação do RNA. N também se liga à proteína M e ao complexo replicase. A proteína S, um trímero, forma a estrutura “coroa solar” na superfície do vírus e a ligação ao receptor da célula hospedeira por meio das subunidades S1 e S2 (Figura 4) (Li, 2020; Greaney, 2021).

Figura 3: Organização do genoma do SARS CoV-2 e suas proteínas codificadas. Fonte: Helmy, 2020

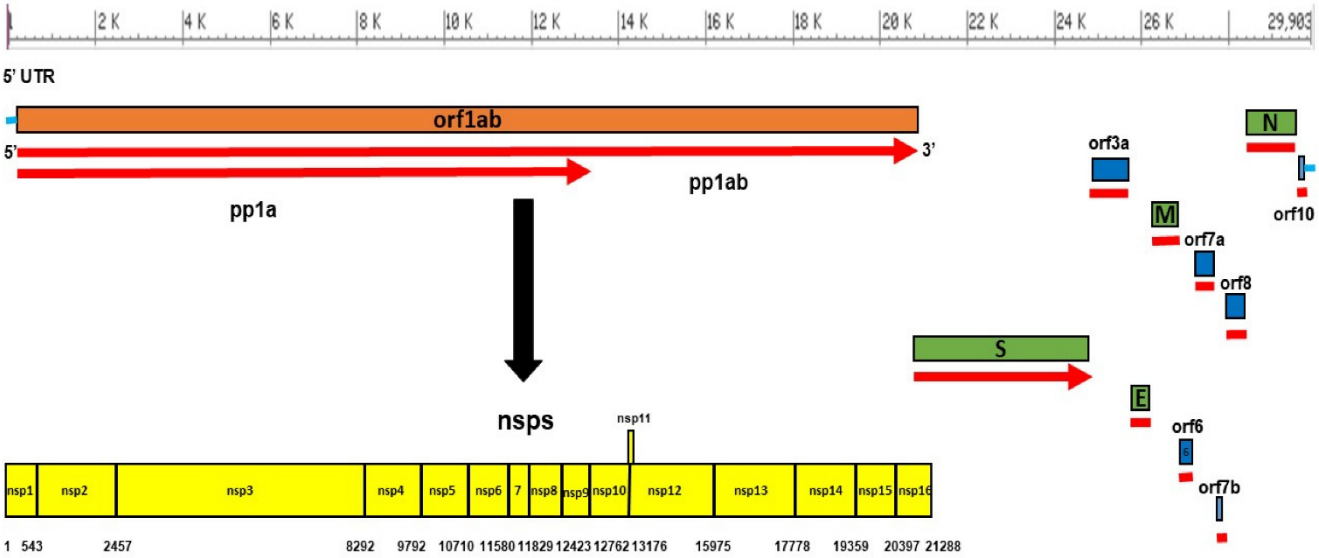
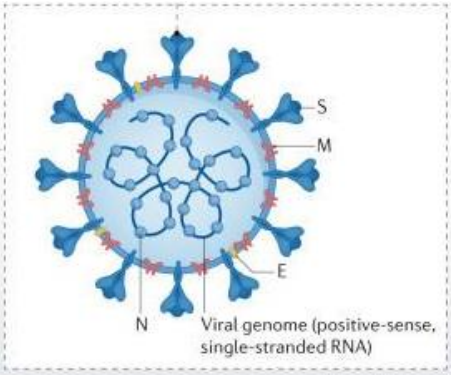
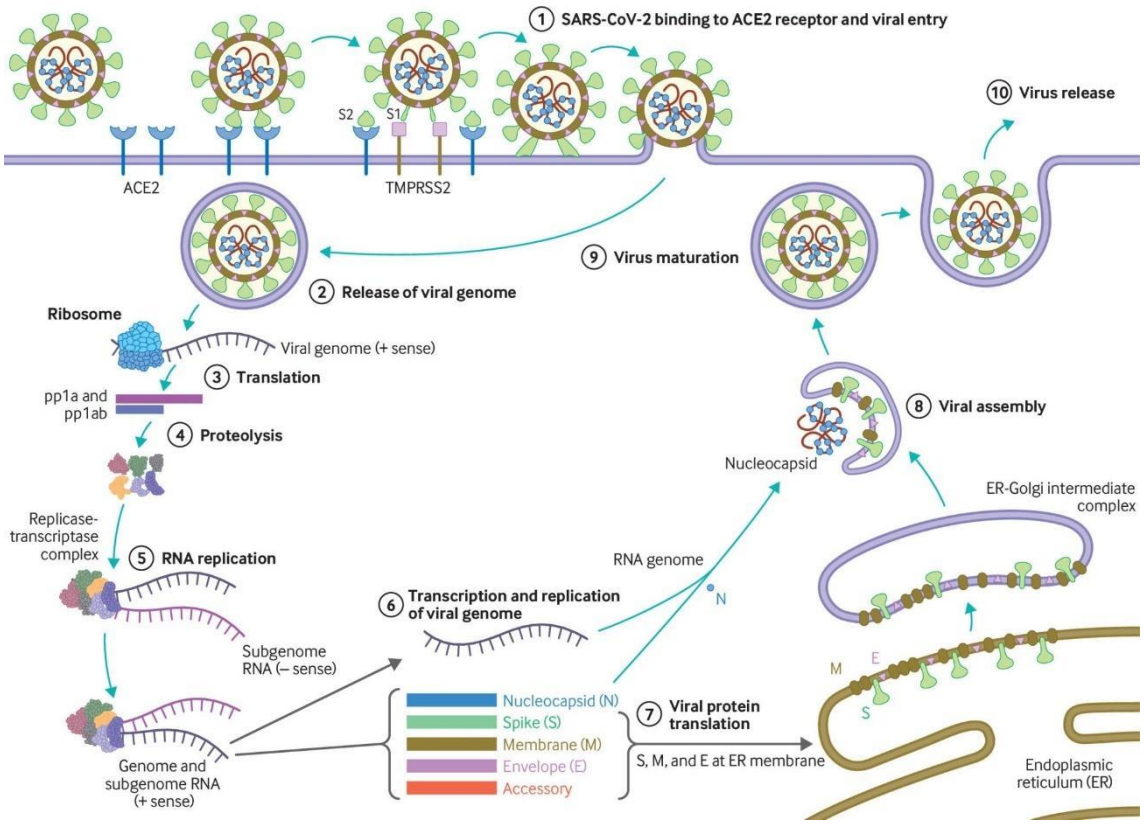


Figura 4: (A) Estrutura da partícula de SARS-CoV-2. (B) Ilustração esquemática da entrada e replicação viral uma célula animal. Fonte: Jackson, 2022; Cevik, 2020

(A)

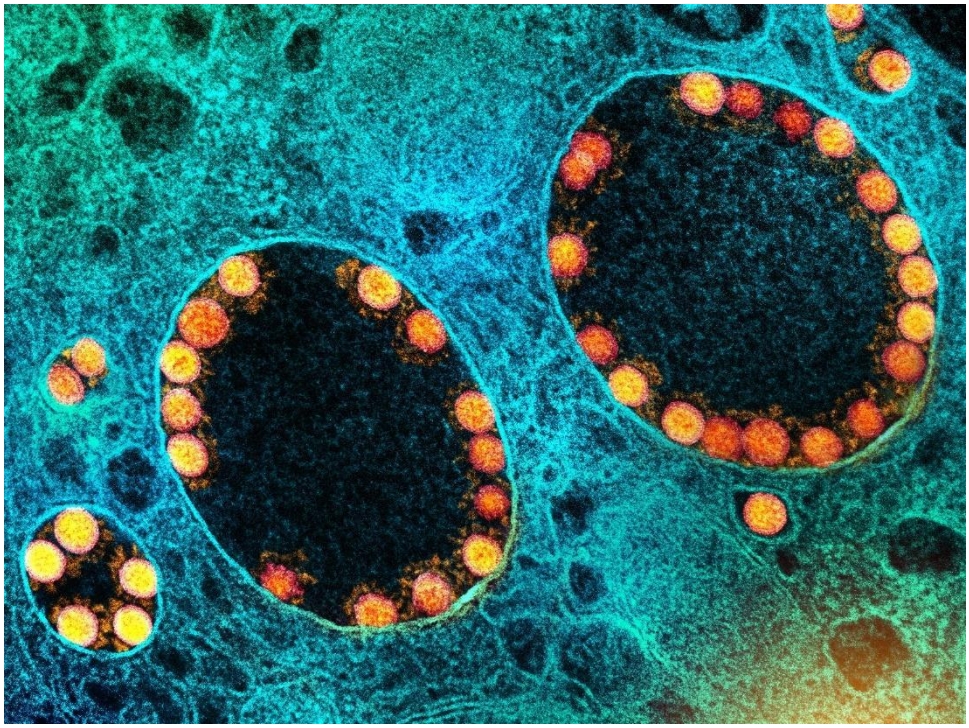




(B)

A replicação do SARS-CoV-2 é um processo complexo que envolve síntese, revisão e limitação de RNA. Começa com a entrada na célula por meio do reconhecimento pela Spike, presente na superfície do envelope viral, dos receptores da enzima conversora de angiotensina 2 humana (hACE2), conforme já observado em SARS-CoV (Luan, 2020; Wrapp, 2020). É possível que outros receptores medeiem a entrada do SARS-CoV-2 nas células hospedeiras, como linfócitos CD147 (Wang, 2020). Após a fixação, a protease transmembrana serina 2 humana (TMPRSS2) cliva e ativa a Spike (Hoffmann, 2020) permitindo que o SARS-CoV-2 entre nas células hospedeiras por endocitose ou fusão direta do envelope viral com a membrana celular (Yang, 2020; Xia, 2020). Uma vez dentro da célula, o RNA infectante atua como um mRNA, que é traduzido pelos ribossomos hospedeiros para produzir as enzimas da replicação viral (Figura 4 e 5). Estas proteínas geram novos genomas de RNA e mRNAs para a síntese dos componentes necessários para montar as novas partículas virais (Jackson, 2022; Cevik, 2020).

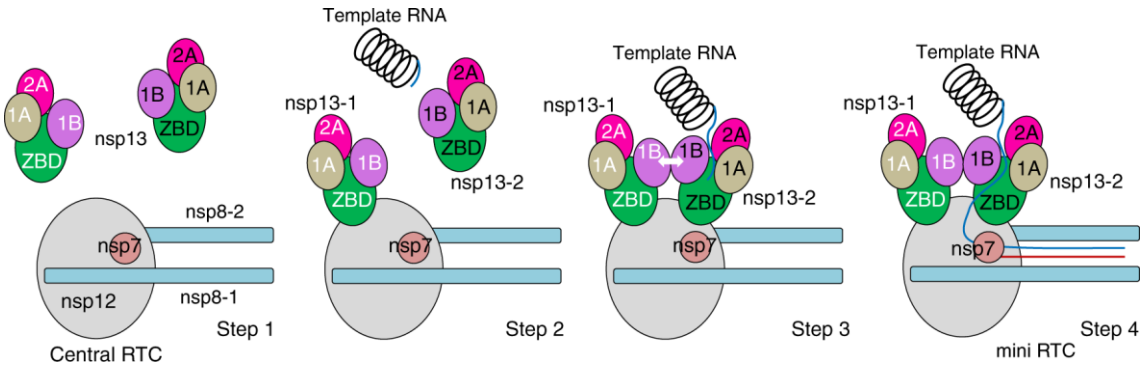
Figura 5: Imagem de microscopia eletrônica de transmissão de partículas virais de SARS-CoV-2 (em amarelo) dentro de endossomos de uma célula epitelial nasal fortemente infectada. Imagem capturada no NIAID *Integrated Research Facility* em Fort Detrick, Maryland. Fonte: https://www.niaid.nih.gov/news-events/novel-coronavirus-sarscov2-images



As nsp têm diversas utilidades, trabalhando isoladamente ou em complexos. A maior finalidade dessas proteínas é de montar o complexo de replicação e transcrição (RTC) (Ziebuhr, 2005). No RTC, a nsp12 tem a função de RNA polimerase dependente de RNA, usando a nsp7 e a nsp8 como auxiliares. A nsp13 é uma helicase dependente de ATP e catalisa o desenrolamento de oligonucleotídeos em fitas simples (Adedeji, 2012; Jia, 2019). Além disso, sugere-se que o nsp13 também desempenha um papel no capeamento do mRNA, além de vários papéis no ciclo de vida do coronavírus (Ivanov, 2004 ; Ivanov, 2004 b). Em trabalhos anteriores, foi determinado que o complexo nsp7-nsp8-nsp12-RNA é o componente central do RTC do SARS-CoV-2 (Gao, 2020; Wang, 2020). Um dos estudos propôs que durante a formação do RTC, o complexo nsp13-nsp1se liga ao complexo nsp8-nsp1 e à nsp12 para fornecer uma base para o par de nsp13-nsp2. O nsp13-2 individual tem uma atividade de helicase para desenrolar o molde de RNA com fita dupla. Após o nsp13-2 ser montado no RTC, o contato entre o par de nsp13-2 permite

a entrada de seu canal para a entrada do RNA. Subsequentemente, o molde de RNA de fita simples desenrolado passa por meio do canal e se estende até a nsp12 para a subsequente síntese de ácidos nucleicos (Yan, 2020). Na figura 6, observam-se os passos para a formação do RTC: Etapa 1, a nsp12 se liga a nsp7 e nsp8 para constituir o RTC central, aguardando o desenrolamento de um RNA modelo. A molécula nsp13 se mantém em estado inativo. Na Etapa 2, uma nsp13 (nsp13-1) entra em contato com o RTC central para formar uma plataforma e recrutar o nsp13-2. Na Etapa 3, nsp13-2 é montado no RTC e entra em contato com nsp13-1, permitindo que o RNA seja ligado e se mova em direção ao centro de atividade da nsp12. Já na Etapa 4, o molde de RNA de fita simples desenrolado passa pelo canal de ligação no nsp13-2 e se estende até o centro ativo da nsp12 para a subsequente síntese de RNA.

Figura 6: Diagrama esquemático do RTC. Fonte: Yan, 2020



A proteína Spike é responsável por mediar a interação com o receptor da hACE2 (Figura 4B) e é um alvo primário do desenvolvimento de anticorpos neutralizantes e vacinas (Baum, 2020; Gu, 2020) Esta proteína recobre a superfície do SARS-CoV-2 e é composta por duas subunidades, S1 e S2. S1 contém o domínio de ligação ao receptor (RBD) e o domínio N-terminal (NTD), responsáveis pela ligação do vírus na superfície da célula hospedeira. S2 é uma subunidade rica em alfa-hélice que contém domínios essenciais para a fusão das membranas da célula e do vírus - entrada do vírus na célula (Li, 2020; Greaney, 2021).

Como muitos outros vírus, os CoVs têm a capacidade de manipular e interromper a progressão do ciclo celular para se beneficiar do estado fisiológico das células infectadas em uma fase específica (Bagga e Bouchard 2014). O estágio pode diferir entre espécies, tipo celular e hospedeiro. Por exemplo, as células infectadas pelo IBV demonstraram entrar em parada do ciclo celular na fase S por meio da ativação da resposta ao dano do DNA celular (Xu *et al*. 2011). Isso é benéfico para a replicação do vírus, pois os fatores que normalmente são necessários para a transcrição do DNA na célula são regulados positivamente na fase S. Assim, esses elementos podem ser recrutados pelo vírus no citoplasma celular. Uma variedade de proteínas de CoVs foi associada à indução da parada do ciclo celular, mas atenção deve ser dado ao fato de que muitos estudos envolveram a expressão exacerbada de proteínas virais individuais. Portanto, devem ser interpretados com cautela, pois seus níveis de expressão, provavelmente, são diferentes dos durante a infecção (de Wilde *et al*., 2017).

1.5 A Pandemia de COVID-19

Após seu surgimento inicial em Wuhan na China no final de 2019, o SARS-CoV-2 se espalhou rapidamente pelo mundo levando à pandemia COVID-19 - oficialmente reconhecida em 11 de março de 2020 pela OMS (OMS, 2020). Até o dia nove de março, mais de 450 milhões de casos e mais de 6 milhões de mortes foram confirmados no mundo. No Brasil, o terceiro país mais afetado pela COVID-19, foram notificados 29,2 milhões de casos e 653 mil mortes (Dong, 2020). Estes grandes números podem estar relacionados, entre outros fatores, à magnitude continental do Brasil, levando a múltiplas introduções virais (Candido, 2020) e ao surgimento de novas variantes de preocupação (VOC) que apresentam maior potencial infeccioso.

A COVID-19 pode manifestar uma larga diversidade de sintomas, variando de casos assintomáticos e manifestações clínicas leves, até casos graves e críticos. De acordo com o Ministério da Saúde brasileiro, os casos podem ser classificados em:

I. Assintomático: Caracterizado por teste laboratorial positivo para COVID-19 e ausência de sintomas;

II. Leve: Caracterizado a partir da presença de sintomas não específicos, como tosse, dor de garganta ou coriza, seguido ou não de anosmia, ageusia, diarreia, dor abdominal, febre, calafrios, mialgia, fadiga e/ou cefaleia;

III. Moderado: Os sintomas mais frequentes podem incluir desde sinais leves da doença, como tosse persistente e febre persistente, até sinais de piora progressiva de outro sintoma relacionado à COVID-19 (adinamia, prostração, hiporexia, diarreia), além da presença de pneumonia sem sinais ou sintomas graves;

IV. Grave: Considera-se a Síndrome Respiratória Aguda Grave como síndrome gripal que apresente dispneia, desconforto ou angústia respiratória, pressão persistente no tórax, saturação arterial de oxigênio menor que 95% em ar ambiente ou coloração azulada de lábios ou rosto. Para crianças, os principais sintomas incluem taquipneia, hipoxemia, desconforto respiratório, alteração da consciência, desidratação, dificuldade para se alimentar, lesão miocárdica, elevação de enzimas hepáticas, disfunção da coagulação, rabdomiólise, cianose central ou saturação arterial de oxigênio <90% em repouso e ar ambiente, letargia, convulsões, dificuldade de alimentação/recusa alimentar;

V. Crítico: Os principais sintomas são sepse, síndrome da angústia respiratória aguda, insuficiência respiratória grave, disfunção de múltiplos órgãos, pneumonia grave, necessidade de suporte respiratório e necessidade de internação em unidades de terapia intensiva.

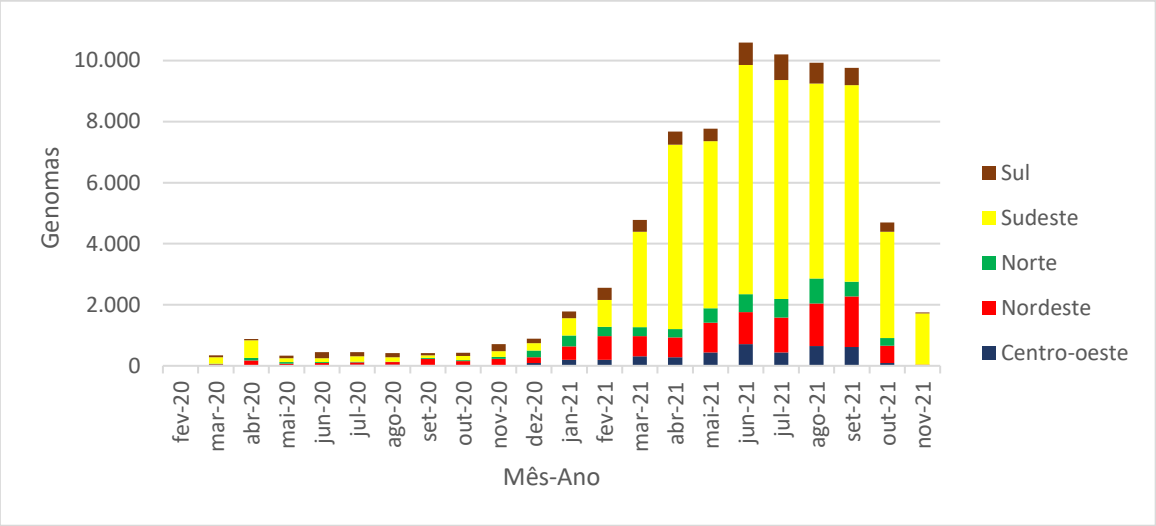
A COVID-19 pode estar também associada a sintomas e sequelas neurológicas, como delírio, encefalopatia, agitação, acidente vascular cerebral, meningoencefalite, olfato ou paladar prejudicados, ansiedade, depressão e distúrbios de sono. Em alguns casos, manifestações neurológicas foram relatadas mesmo em pacientes sem sintomas respiratórios. As manifestações clínicas da infecção pelo SARS-CoV-2 são geralmente mais leves em crianças do que em adultos. No entanto, em 26 de abril de 2020, o Sistema Nacional de Saúde Inglês lançou um alerta relatando uma nova apresentação clínica em

crianças, caracterizada como um conjunto de sintomas por inflamação exacerbada que pode levar a um quadro de falência múltipla de órgãos e choque (Ministério da Saúde).

Mudanças importantes nos epicentros do COVID-19 ocorreram durante 2020, começando com a Ásia, seguido pela Europa, América do Norte e América do Sul. Após meses de evolução relativamente lenta, novas VOCs como Alpha, Beta, Gamma, Delta e Omicron, surgiram (Mullen, 2020). Mais recentemente, a Organização Mundial da Saúde (OMS) atribuiu rótulos às VOCs com base em letras gregas. Portanto, as quatro VOCs são denominadas, respectivamente, Alfa, Beta, Gama e Delta.

Esforços internacionais de sequenciamento viral permitiram o envio de mais de 9.14 milhões de genomas para a Iniciativa Global de Compartilhamento de Todos os Dados da Influenza (GISAID) até março de 2022 (Shu, 2017) e estão disponíveis para estudos de epidemiologia genômica para acompanhar a história e a dinâmica evolutiva de SARSCoV-2. Por meio desse repositório, também é possível realizar estudos levando em consideração a geografia de um continente, país ou região. Por exemplo, algumas VOCs tiveram múltiplas inserções em países grandes como o Brasil (Candido, 2020). O Brasil sequenciou um total de 76.413 amostras até 7 de dezembro de 2021, com uma média mensal de 724 genomas por mês. Infelizmente, esse número é muito baixo se comparado ao número de casos no mesmo período (> 22.140.000), representando cerca de 0,34% (Dong, 2020). A região Sudeste é responsável pela maior parte do sequenciamento no país, seguida pelo Nordeste, Sul, Norte e Centro-Oeste (Figura 8). Até novembro de 2020, havia poucas amostras sequenciadas em comparação com a população do país. Somente a partir dos primeiros meses de 2021, a taxa de sequenciamento de SARS-CoV-2 aumentou. Isso tem a ver, provavelmente, com o rápido crescimento de casos de COVID-19 no estado do Amazonas - a segunda onda pandêmica no Brasil - e ao surgimento da linhagem P.1 e suas sublinhagens, a variante Gama (Figura 8).

Figura 7: Representação gráfica do número de genomas sequenciados no Brasil por mês e a proporção em cada região.



Apesar das regiões possuírem diferentes números de habitantes, a porcentagem de genomas sequenciados não corresponde aos tamanhos das populações de cada região brasileira. Esta disparidade é mais aparente no Sudeste, Nordeste e Sul. Eles representam, respectivamente, 42%, 27% e 14% dos brasileiros, mas contribuíram com as taxas discrepantes de sequenciamento de 66%, 14% e 8%. Além disso, o território altamente populoso do Nordeste apresenta um número grande de genomas, No entanto, representa apenas 0,018% dos casos de COVID-19 da região, a menor entre as cinco divisões brasileiras. Em contraste, a região Sudeste tem o maior percentual de sequenciamento de seus casos - 0,057%, mais de três vezes maior que o Nordeste (Tabela 1). Esta área é conhecida por ter não só alguns dos estados mais populosos, mas também por abrigar grandes institutos e unidades de pesquisa em Biologia, Ciências da Saúde e Saúde Pública, como Instituto Oswaldo Cruz, Butantan, Instituto Adolfo Lutz e grandes universidades. Estas instituições estão localizadas principalmente nos estados de São Paulo e Rio de Janeiro, são bem equipadas e recebem apoio financeiro do governo para manutenção de suas pesquisas e vigilância epidemiológica.

Tabela 1: Proporções de cada região brasileira quanto à população, sequenciamento, casos e óbitos no período estudado.

	População (2021)	Sequências	Casos cumulativos	Mortes cumulativas	Casos sequenciados
Norte	9%	7%	9%	8%	0.026%
Nordeste	27%	14%	22%	19%	0.018%
Centro-oeste	8%	5%	11%	10%	0.023%
Sudeste	42%	66%	39%	48%	0.057%
Sul	14%	8%	19%	16%	0.021%
Brasil	**100%**	**100%**	**100%**	**100%**	**0.036%**

1.6 Variantes e linhagens de SARS-CoV-2

Para estabelecer uniformidade na denominação de variantes e linhagens de SARS-CoV-2, o Centro de Controle e Prevenção de Doenças dos Estados Unidos (CDC) define:

I. Linhagem - Um grupo de vírus intimamente relacionados com um ancestral comum. O SARS-CoV-2 possui várias linhagens.
II. Variante - Um genoma viral que pode conter uma ou mais mutações, levando a diferentes variantes. Variantes do SARS-CoV-2 associadas à maior transmissibilidade, virulência ou capacidade de escapar da imunidade natural mediada por infecção ou vacinas são chamadas de Variantes de Preocupação.
III. Mutação - Uma mudança no genoma. Dependendo do organismo, as mutações acontecem com frequência, mas apenas, às vezes, podem mudar características de um organismo. No caso do SARS-CoV-2, as mutações virais que são intimamente relacionadas são agrupadas e nomeadas de acordo com as substituições mais representativas de cada grupo (Guan, 2020; Shu, 2017; Mercatelli, 2020).

Recentemente, a OMS atribuiu rótulos às variantes com base em letras gregas. Após meses de evolução gradual, surgiram novas VOCs - Alpha, Beta, Gamma, Delta e Omicron - carregando combinações singulares de substituições de aminoácidos na Spike (Mullen, 2020). Essas linhagens apareceram independentemente no Reino Unido, África do Sul, Brasil, Índia e, mais uma vez,

na África do Sul respectivamente. (Rambaut, 2020; Tegally, 2021; Faria, 2021; Naveca, 2021; Cherian, 2021; Callaway, 2021), sendo que originaram surtos secundários nos países onde emergiram, apesar de eventual alta soroprevalência (Buss, 2021).

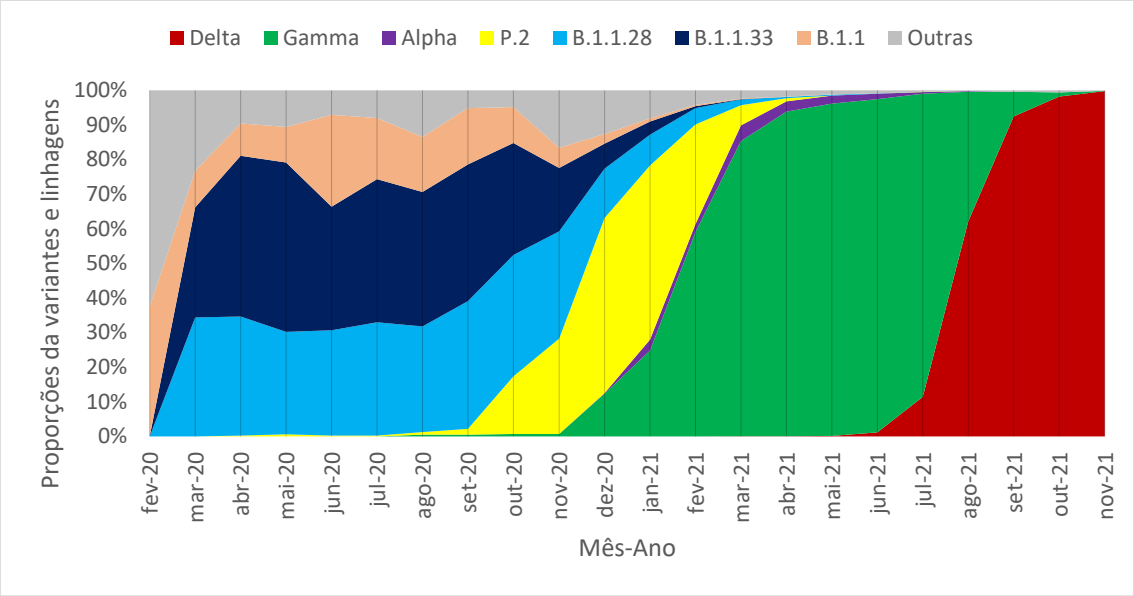
1.7 Variantes e linhagens de SARS-CoV-2 no Brasil

1.7.1 *B.1.1, B.1.1.28, B.1.1.33 e P.2*

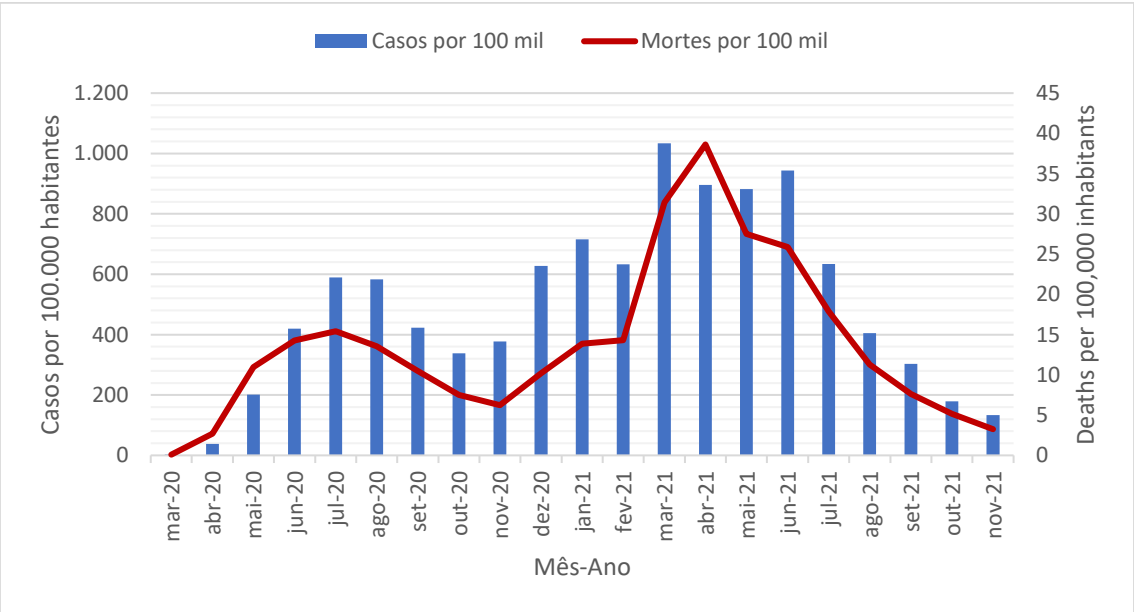
A diversidade de SARS-CoV-2 durante a primeira onda epidêmica no Brasil foi composta principalmente pelas linhagens B.1.1.33, B.1.1.28 e B.1.1. (Figura 8A) (Candido, 2020; Franceschi, 2021). Até setembro de 2020, B.1.1.33 era a mais prevalente, seguido por B.1.1.28 e B.1.1. A partir de outubro, as três começaram a diminuir suas proporções. Essas linhagens foram rapidamente substituídas por P.1 e P.2, no final de 2020 e início de 2021. Ambas são derivadas do ancestral comum B.1.1.28 e abrigam mutações importantes na Spike, como E484K e N501Y (Ferrareze, 2021; Franceschi e Caldana, 2021; Lamarca, 2021). A P.2 foi detectada pela primeira vez no Brasil em abril de 2020 e, dentre as variantes de maior proporção, chegou a mais de 57% em dezembro de 2020 e mais de 54% em janeiro de 2021. Em março de 2021, teve uma baixa significativa, pois nesse mês houve um grande aumento da proporção de P.1 que deu origem à segunda onda de COVID-19 no país. De agosto de 2021 em diante, não foi mais detectada nenhuma P.2 no Brasil. A partir de fevereiro de 2021, B.1.1 e B.1.1.33 tinham proporções abaixo de 1% e, de abril em diante, o mesmo aconteceu com a B.1.1.28 (Figura 8).

Figura 8: Frequências de linhagens e variantes do SARS-CoV-2 e relação de letalidade para o território brasileiro até 7 de dezembro de 2021. (A) Proporções de linhagens e variantes de acordo com os genomas brasileiros de SARS-CoV-2 do banco de dados GISAID. (B) Proporção de casos e óbitos por 100.000 habitantes por mês. (C) Contagem de casos e óbitos por mês.

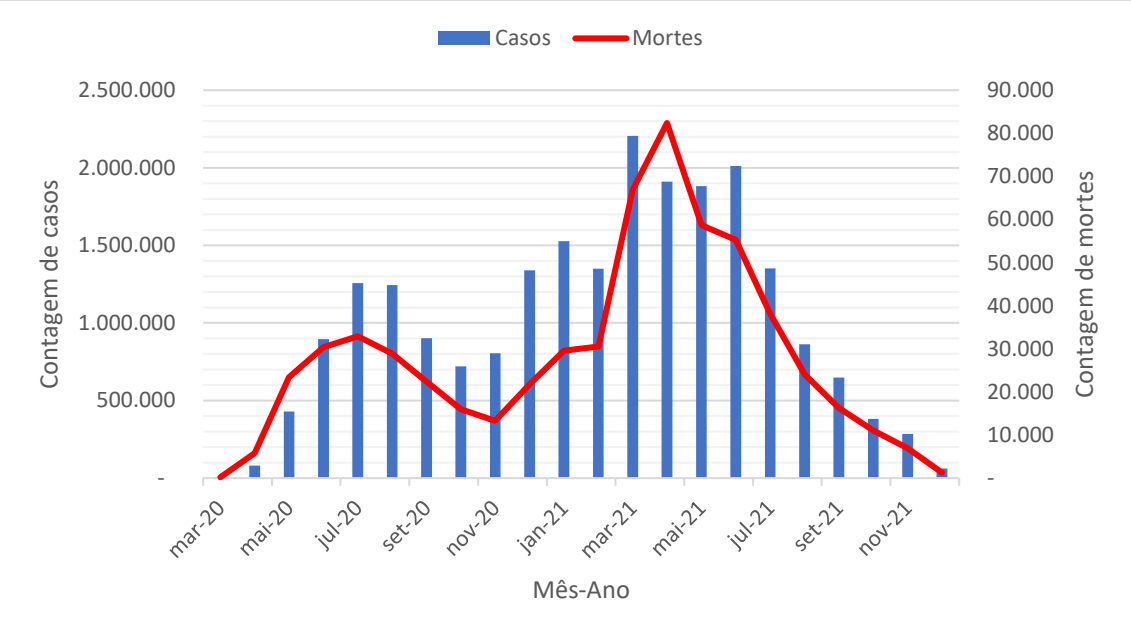
(A)



(B)



(C)



1.7.2 *Alpha*

A variante Alpha (linhagem B.1.1.7) foi apresentada em dezembro de 2020 pelas autoridades de saúde do Reino Unido (Shen, 2021). Sua maior transmissibilidade explica o aumento na incidência, hospitalizações e pressão sobre o sistema de saúde do país (Shen, 2021; Altaman, 2021; Rees-Spear, 2021). Relatórios epidemiológicos e modelagem sugerem que se espalha 56% mais rápido do que outras linhagens e leva a um aumento de 35% no risco de morte. Apesar disso, não houve relatos científicos de doenças mais graves em crianças e jovens (Altaman, 2021; Hu, 2021). Em comparação com ancestrais contendo a mutação D614G, a variante Alfa acumulou 23 mutações. Destas mutações, as que têm maior potencial para influenciar as características biológicas do vírus são H69-V70del, N501Y e P681H (Kumar, 2021; Mohammadi, 2021). Quase metade (47%) das mudanças relatadas ocorrem na proteína Spike. Essas mutações podem desempenhar um papel importante na alteração da interação com o hACE2 e no comprometimento da eficácia de anticorpos neutralizantes, anticorpos monoclonais, células T e soro convalescente (Hu, 2021; Mohammadi, 2021).

Alpha teve uma presença discreta no Brasil (Figura 8A). Foi sequenciada pela primeira vez em dezembro de 2020 e atingiu um máximo de 4,63% em março de 2020. A partir de julho de 2021, sua proporção já era menor que 1% e de setembro em diante não foi mais detectada.

1.7.3 *Beta*

A linhagem B.1.351, nomeada pela OMS como variante Beta, foi relatada pela primeira vez em dezembro de 2020 pela República da África do Sul e sua característica mais significativa é a maior taxa de transmissão (Mwenda, 2021). Possui 18 mutações em comparação à cepa de referência e 8 delas estão localizadas na proteína Spike. As substituições K417N, E484K e N501Y são as mais relevantes, pois sugere-se que a variante possa escapar da ação de anticorpos neutralizantes e comprometer a eficácia de vacinas (Wang, 2021; Zhou, 2021; Shen, 2021b; Ferrareze, 2021).

A Beta teve apenas 10 genomas sequenciados no Brasil até 7/dez/2021. Destes, nove foram detectados no estado de São Paulo, em março e abril de 2021, e um na Bahia em agosto do mesmo ano. Apesar de pouca incidência - provavelmente introduzida por viajantes - foi o suficiente para que o vírus obtivesse mais três novas mutações em relação à cepa inicial da África do Sul (Slavov, 2021). Esse fenômeno corrobora o fato de que enquanto o vírus estiver em circulação, sua constante evolução propiciará novas linhagens e variantes colocando em risco vacinas e tratamentos atuais.

1.7.4 *Gamma*

A linhagem P.1 (originalmente B.1.1.28.1) descende da B.1.1.28 e faz parte da variante Gamma, que engloba também as sublinhagens da P.1. Foi detectada pela primeira vez no Japão em janeiro de 2021 e foi isolada de quatro viajantes que chegaram a Tóquio vindos do Amazonas. (Fujino, 2021; Freitas, 2021). Gamma possui 23 mutações, sendo 12 na Spike (Mullen, 2020). As mutações mais marcantes são a tríplice K417T, E484K e N501Y (algumas destas

estão presentes também em outras variantes). Este grupo de mutações na Spike tem implicações importantes para a evasão da imunidade humoral. As primeiras sequências brasileiras de Gamma foram detectadas em agosto de 2020, tiveram aumento significativo em fevereiro de 2021 (mais de 61%) e entre abril e junho já representavam mais de 95% dos genomas sequenciados no Brasil. De agosto em diante, houve uma queda brusca da proporção da variante Gamma, pois já começava a perder espaço para a Delta.

Um número assustador de casos e mortes no Brasil pela P.1 e sublinhagens é evidente a partir de dezembro de 2020 (Figura 8). Com o surto, um grande aumento de casos, hospitalizações e óbitos resultou na segunda onda de COVID-19 vivenciada de forma semelhante à primeira onda (Ministério da Saúde). No Brasil, a segunda onda teve início em Manaus, onde a linhagem P.1 foi descrita pela primeira vez, causando muitas mortes por falta leitos de UTI e de oxigênio (Sabino, 2021). A transmissão endêmica do SARS-CoV-2 no Amazonas no ápice da segunda onda foi influenciada pela disseminação de novas sublinhagens locais P.1 que são mais transmissíveis (Naveca, 2021). As principais sublinhagens que aumentaram proporcionalmente por volta de março de 2021 contêm mutações como S:N679K (P.1.4) e S:P681H (P.1.6 e P.1.7). Foram nomeadas como “P.1 Plus” por causa de suas novas substituições adquiridas independentemente na proteína Spike, por exemplo, P.1+679K, P.1+681H, P.1+681R e P.1+NTDdels (deleções no NTD). Além disso, a primeira substituição é atualmente conhecida por estar presente em Omicron e a segunda em Omicron e Alpha. A partir de março de 2021, P.1 já começou a diminuir de proporção dando lugar às suas sublinhagens. Em agosto de 2021, a P.1 plus foi responsável por mais de 72% dos genomas sequenciados na região Norte. Um estudo indicou que a transmissão endêmica do SARS-CoV-2 no Amazonas durante a segunda onda estava ligada à evolução da Gamma por meio da aquisição de deleções de NTD e das mutações do local de clivagem da furina (P681H e N679K) (Naveca, 2021). O estudo também observou que P.1.4 e P.1.6 foram associadas a um aumento de 6 vezes na carga de RNA. Isso sugere que indivíduos infectados com P.1 Plus podem ser mais infectantes do que aqueles que abrigam apenas as mutações parentais P.1. A transmissão endêmica pode ter selecionado mais linhagens infecciosas ao invés de linhagens que são mais

aptas para evadir o sistema imunológico. No Centro-Oeste, a P.1.7 surgiu em março de 2021 e atingiu quase metade dos genomas sequenciados em agosto. Aparentemente, essa linhagem surgiu quase ao mesmo tempo que no Norte e, mais discretamente, no Nordeste. Essa semelhança coincide com a proximidade geográfica das regiões.

Com a queda da incidência da COVID-19 por volta de outubro e novembro de 2020, muitos estados e municípios relaxaram os protocolos da pandemia. Mas o surgimento do P.1 e as comemorações de fim de ano coincidiram nos meses seguintes, marcando a ascensão da segunda onda no Brasil (Figura 8). Em fevereiro de 2021, a prevalência de SARS-CoV-2 diminuiu em comparação com o mês anterior. Com exceção do Nordeste, todas as regiões tiveram queda nos casos e, logo na sequência, uma alta acentuada em março. Assim, pode-se considerar algumas circunstâncias concomitantes. Alguns meses antes, a P.1 foi introduzida primeiramente no Norte e rapidamente se espalhou por todo o país nos meses posteriores (Fujino, 2021). Ao mesmo tempo, duas outras VOCs altamente transmissíveis, Alfa e Beta, estavam surgindo em outros países (Rambaut, 2020; Tegally, 2021). Outra ocasião de destaque foi o Carnaval de 2021. Trata-se de um evento que acontece todos os anos por volta de fevereiro e março e o país recebe turistas de todo o mundo. Naquele ano, a comemoração deveria acontecer entre os dias 12 e 17 de fevereiro, mas foi oficialmente cancelada devido ao aumento dos números da Covid-19 (New York Times). Infelizmente, várias festas e carnavais clandestinos aconteceram nas semanas próximas às datas do Ano Novo e do Carnaval.

Até novembro de 2020, havia poucas amostras sequenciadas no Brasil. Os esforços no sequenciamento do SARS-CoV-2 aumentaram somente a partir de março de 2020 (Figura 7). Esse período marcou a ascensão da segunda onda, que devastou o Brasil entre março e junho de 2021, atingindo a impressionante marca de 82.392 óbitos no mês abril de 2021 (Figura 8C). Além disso, os piores números de casos e óbitos - considerando todo o período estudado e todas as regiões - pertencem à região Sul em março de 2021. Neste mês, a região atingiu um total de 539.855 casos e 15.919 óbitos, o que representa as espantosas proporções de 1.775,7 casos por 100 mil habitantes e 52,36 óbitos por 100 mil habitantes (Projeto Brasil.IO). Porto Alegre, capital do

estado do Rio Grande do Sul, foi considerada um epicentro do colapso do sistema de saúde do país durante a segunda onda. Em março de 2021, o estado apresentou uma proporção de óbitos de 64 óbitos por 100 mil habitantes, semelhante ao que ocorreu no Amazonas em janeiro e fevereiro daquele ano. Houve proporções ainda maiores de óbitos no ápice da segunda onda em comparação com a primeira onda, caracterizando a segunda onda como mais letal.

1.7.5 *Delta*

A variante Delta, detectada pela primeira vez na Índia em outubro de 2020 (OMS, "*Tracking SARS-CoV-2 variants*") e rapidamente disseminada por todo o país, compreende a linhagem B.1.617.2 e as sublinhagens AY. Abriga um total de 29 mutações, sendo oito na proteína Spike. Destas oito, as substituições L452R e P681R são as mais relevantes (Cherian, 2021). A primeira concede maior afinidade ao receptor hACE2 (Starr, 2021) e pode permitir a evasão da atividade antiviral das células-T CD8 (Koshy, 2021). A segunda substituição facilita a fusão da membrana e integração do vírus à célula hospedeira. (Haseltine, 2021; Bertram, 2013). A combinação das mutações acima mencionadas pode nomear a Delta como a variante mais transmissível dentre as existentes na época. (Shiehzadegan, 2021).

No Brasil, a variante Delta foi sequenciada pela primeira vez em março de 2021 e em agosto já representava mais da metade dos genomas sequenciados no país (62,27%). De setembro em diante, foi responsável por mais de 92% das sequencias brasileiras no período analisado (Figura 8A). Felizmente, a partir deste período o número de casos e de mortes já estava em declínio.

1.7.6 *Omicron*

Em 26 de novembro, a OMS nomeou a linhagem encontrada na África do Sul, B.1.1.529, como Omicron, juntando-se a Alpha, Beta, Gamma e Delta na lista atual de VOCs. Esta variante contém mais de 30 mutações na Spike, muitas

delas já encontradas em outros VOCs, como Delta e Alpha. Essas alterações estão possivelmente relacionadas à elevada infectividade. (Callaway, 2021; Torjesen, 2021). De acordo com o repositório do GISAID, as primeiras sequências de Omicron no país são do final de novembro de 2021. Um aumento acentuado no número de casos é perceptível já nas primeiras semanas de janeiro, mas proporcionalmente com menos mortes (Taylor, 2022). Ao mesmo tempo, a procura por doses de reforço da vacina contra COVID-19 subiu no país (Dong, 2020). A campanha de vacinação do Brasil aparentemente reduziu bastante o número de internações e óbitos no segundo semestre de 2021 e, aproximadamente, 67% dos brasileiros já estão totalmente vacinados (Taylor, 2022). Apesar disso, o número de novas infecções por SARS-CoV-2 registradas globalmente no final de dezembro de 2021 e nos primeiros dias de janeiro de 2022 aumentou 71% em relação à semana anterior. Nesse período, as infecções semanais aumentaram 65%, 78% e 100%, respectivamente, na Europa, Sudeste Ásia, nas Américas (b Taylor, 2022). Ocorre, assim, uma rápida reviravolta no declínio constante de infecções que o mundo experimentava desde outubro de 2021.

1.8 Evolução e mutações

A evolução dos CoVs ocorre em seus hospedeiros e é impulsionada pela sua diversidade genética e por pressões evolutivas. As mutações são possibilitadas pelo grande genoma viral, alta taxa de mutação (~10-4 substituições por sítio por ano) e alta frequência de recombinação (até 25% para todo o genoma *in vivo*) (Woo, 2010; Vijgen, 2005; Sánchez, 1992; Baric, 1990; Banerjee, 2020). Mutações que conferem maior *fitness* são selecionadas e, consequentemente, levam à deriva antigênica - um fenômeno evolutivo de tendência que altera constantemente a estrutura de antígenos do vírus por meio de mutações. O resultado da razão das taxas de mutações não-sinônimas/sinônimas >1 indica seleção positiva, <1 seleção negativa e =1 evolução neutra (Kryazhimskiy e Plotkin, 2008). Apesar de observar pouca diversidade viral no início da pandemia (Simmonds, 2020; Mavian, 2020), a seleção positiva com vantagens presumidas, como taxas de transmissão

aumentadas, foi documentada (Volz, 2021; Meng et al., 2021; Davies, 2021). No entanto, a caracterização funcional dessas mutações permanece pouco investigada (Singh, 2021). A deriva antigênica é mais frequentemente observada em proteínas da superfície viral que estão altamente expostas a pressões de seleção do sistema imunológico do hospedeiro (Bush, 2021). De fato, os genes da proteína Spike, particularmente as regiões de codificação S1 e RBD, têm as taxas de mutação não-sinônima mais altas (Gallagher, 2020; Tortorici, 2019), uma tendência observada na maioria dos CoVs (Singh, 2021). Estudos sugerem que sítios selecionados positivamente também foram observados no domínio N-terminal (NTD) do RBD de SARS-CoV-2 (Rambaut, 2020; Mccarthy, 2021; Weisblum, 2020). Especula-se que com a circulação ininterrupta do vírus, a vacinação e a terapia com soro convalescente, outras mutações selecionadas positivamente no NTD são prováveis de ocorrer. Outras pesquisas retrospectivas sobre a evolução de CoVs endêmicos podem ajudar a prever prováveis trajetórias evolutivas do SARS-CoV-2 (Singh, 2021). Com a rápida disseminação das diferentes variantes, é inerente o surgimento de novas mutações, que acabam selecionadas por imunidade populacional, vacinal e outros fatores ainda desconhecidos (Rambaut, 2020; European Centre for Disease Prevention and Control, 2020. Center for Disease Control and Prevention, 2021). Importante destacar a necessidade de esforços internacionais de vigilância epidemiológica e compartilhamento global de dados genômicos (Singh, 2021).

A mutação D614G - uma substituição de ácido aspártico por glicina na posição de aminoácido 614 do gene S (Zhang, 2020; Korber, 2020) - atualmente é detectada em todas as VOCs (Mullen, 2020; Korber, 2020; Lauring e Hodcroft, 2021). A D614G intensifica a infectividade do SARSCoV-2 aumentando a carga viral no tecido respiratório (Plante, 2021; Hou, 2020). Isso acontece, possivelmente, devido à maior abertura da Spike ou ao seu potencial de clivagem (Mansbach, 2021, Butowt, 2020), fazendo com que esta mutação se tornasse dominante pouco após a emergência da pandemia (Shu, 2017; Hadfield, 2018). Também existe uma correlação epidemiológica entre D614G e anosmia (Butowt, 2020).

A substituição de uma asparagina por uma tirosina na posição 501 (N501Y) apareceu pela primeira vez no Reino Unido em setembro de 2020 e é

um dos seis resíduos de contato mais importantes dentro de RBD que interage com a hACE2. Essa mutação foi associada com maior afinidade de ligação com o receptor devido à formação de uma ligação de hidrogênio extra com a hACE2 (Khan, 2021; Starr, 2020; Leung, 2020). Estudos sugerem que esta mutação pode aumentar a transmissibilidade e mortalidade do SARS-CoV-2 (Tang, 2021; Gu, 2020; Ali, 2021).

A substituição de ácido glutâmico por lisina na posição 484 do RBD (E484K) cria uma forte interação iônica com o aminoácido 75 da hACE2. Isso acontece devido à mudança eletrostática de um aminoácido carregado negativamente para um carregado positivamente (Nelson, 2021). A mutação também pode levar à evasão imune, uma vez que está localizada em uma alça flexível, outrora irrelevante para a ligação ao receptor onde o ácido glutâmico originalmente estava. Isso pode explicar a disseminação e infectividade aprimoradas do SARS-CoV-2 (Baum, 2020; Greaney, 2020), casos de reinfecção por vírus contendo a E484K (Nonaka, 2021; Resende, 2021) e a sua fixação em diferentes linhagens (Ferrareze, 2021). Além disso, a eficácia da neutralização do vírus que contém a mutação E484K é diminuída tanto no soro de pacientes vacinados como no de pacientes convalescentes (Ferrareze, 2021; Singh, 2021b; Hoffmann, 2021; Rees-Spear, 2021; Focosi e Maggi, 2021; Sabino, 2021). Estes fatos sugerem que esta substituição deve ser investigada continuamente.

A furina é uma serina protease essencial para a infecção presente em alguns coronavírus e vírus influenza (Simmons, 2011; Tse, 2014). A clivagem da furina é uma etapa inicial e crítica da entrada do SARS-CoV-2 em células suscetíveis. Mudanças no local de reconhecimento da furina podem afetar a patogenicidade do vírus (Nagy, 2021). Um trabalho demonstrou que mutações na Spike são responsáveis pela melhora do *fitness* viral da Delta. O estudo identifica a mutação P681R como fator importante para a replicação viral aprimorada da Delta em comparação com a variante Alfa - que possui a P681H (Lubinski, 2021.) A substituição P681R otimiza o processo de clivagem das subunidades S1 e S2 da Spike (Liu, 2021). Além disso, o surgimento independente da mutação em diferentes linhagens sugere evolução convergente e, consequentemente, uma possível vantagem adaptativa, uma vez que a

aquisição de um sítio de clivagem de furina foi, provavelmente, essencial para a infecção por SARS-CoV-2 em humanos (Lasek-Nesselquist, 2021; Hoffmann, 2020).

O resíduo L452, apesar de não ter contato direto com a hACE2, junto com os resíduos F490 e L492 formam uma área hidrofóbica na superfície do RBD. A substituição L452R - presente na Delta - estabiliza a interação entre a Spike e a hACE2, facilita a entrada do vírus na célula e aumenta o potencial de infectividade viral. Em contraste, o aumento na infectividade com a L452R é ligeiramente menor do que o aumento observado com a mutação N501Y (Deng, 2021). Além disso, linhagens que carregam esta mutação parecem apresentar uma resistência moderada à neutralização por anticorpos produzidos por infecção prévia ou vacinação (Deng, 2021).

1.9 Diagnóstico de COVID-19

No estudo de métodos diagnósticos é importante destacar alguns conceitos da epidemiologia analítica. A sensibilidade é a capacidade de um ensaio de identificar corretamente indivíduos doentes (positivos verdadeiros) com base na quantidade mais baixa do analito que a técnica pode detectar, A especificidade é a capacidade de um ensaio de identificar corretamente indivíduos não-doentes (negativos verdadeiros). Já a “precisão” é um termo genérico relacionado à sensibilidade e especificidade de uma técnica e sua capacidade de discriminar corretamente entre indivíduos doentes e saudáveis (Medronho, 2009).

As abordagens de detecção de SARS-CoV-2 mais utilizadas incluem testes rápidos de antígenos ou anticorpos, testes moleculares baseados em ácidos nucleicos e testes sorológicos enzimáticos imunológicos. Outras técnicas, como amplificação isotérmica de RNA mediada por grampos (LAMP), *clusters* de repetições palindrômicas curtas regularmente interespaçadas e quimioluminescência, são usadas em ambientes de pesquisa.

RT-qPCR, tomografia computadorizada (TC) e exames hematológicos e bioquímicos são testes diagnósticos básicos e os principais de COVID-19

(Roberts, 2021) atualmente disponíveis no mercado (Cui e Zhou, 2020). A TC e a RT-qPCR estão mais provavelmente limitadas a instituições maiores e mais bem equipadas, pois consomem muito tempo, são caras e exigem profissionais altamente qualificados (Kaushik, 2021). Além disso, a TC pode detectar apenas os sintomas da doença, mas não pode identificar o vírus com base nos sintomas (Li, L. 2020). Ademais, uma correlação positiva foi encontrada entre proteína C reativa, lactato desidrogenase e velocidade de sedimentação de eritrócitos com a gravidade da pneumonia vista por TC (Ai, 2020). No entanto, os testes rápidos atualmente não são confiáveis para detectar casos precoces e assintomáticos (Li, Z. 2020). Testes baseados em anticorpos, juntamente com RT-qPCR, estão sendo usados em conjunto para desenvolver técnicas de diagnóstico COVID-19 mais rápidas, precisas, específicas e altamente sensíveis. Vários ensaios sorológicos estão em desenvolvimento para o diagnóstico rápido de COVID-19, mas têm a desvantagem de menor sensibilidade em comparação com as técnicas convencionais já citadas. (Roberts, 2021).

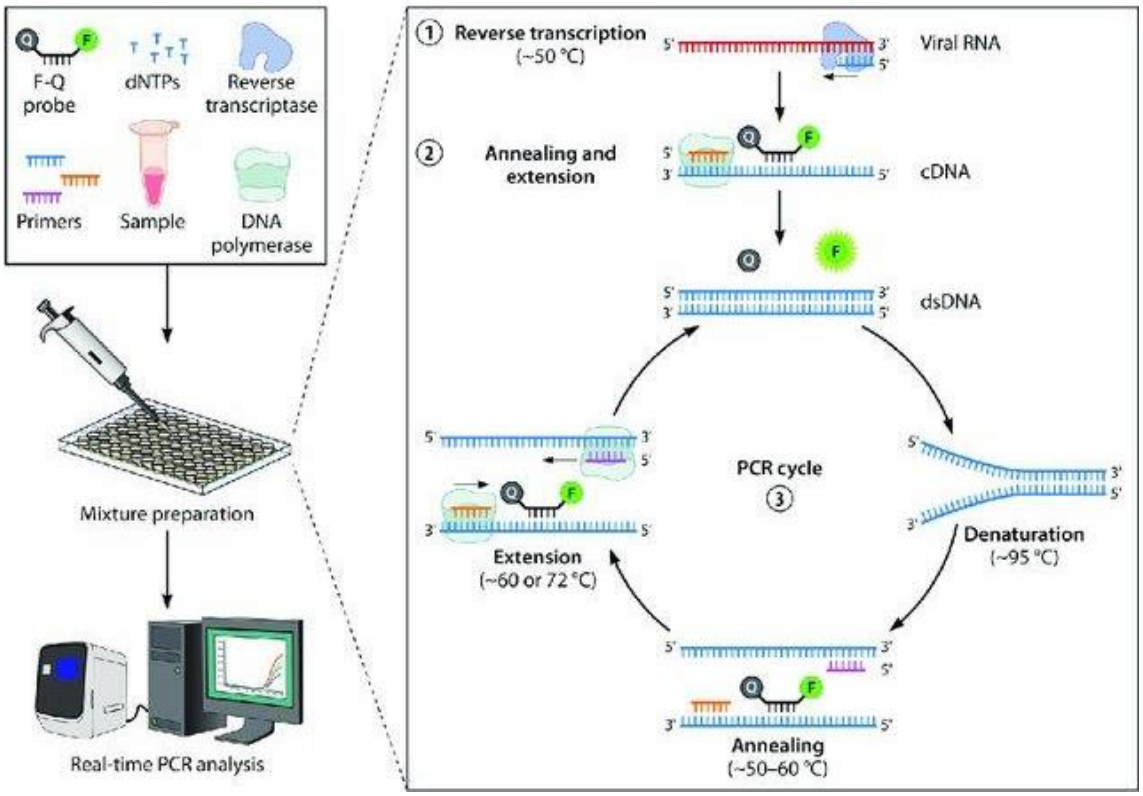
1.9.1 *RT-qPCR*

A detecção de genes alvo por meio da PCR quantitativa em tempo real tem sido a melhor escolha no diagnóstico de doenças infecciosas (Pabbaraju, 2009; Lee, 2001). Vários laboratórios vêm desenvolvendo ensaios de diagnóstico, mas o RT-qPCR continua sendo a técnica de detecção mais acurada para SARS-CoV-2, pois mede diretamente partes do genoma viral. Genes como N, E e RdRp de SARS-CoV-2 são usados para a detecção direcionada (Xu, 2020). A RT-qPCR compreende duas etapas fundamentais. A primeira delas é a transcrição reversa do RNA viral para formar DNA complementar (cDNA); a segunda consiste na amplificação de cDNA por PCR utilizando primers específicos e marcados com sondas fluorescentes (Figura 9).

Durante a transcrição reversa, a transcriptase reversa (RT) cria um cDNA a partir do molde de RNA viral, com o auxílio do primer reverso (ou oligonucleotídeos aleatórios). A atividade da RNase da RT digere o molde inicial de RNA. A atividade da DNA polimerase da RT completa a segunda fita de DNA guiada pelo primer direto e cDNA. O DNA de fita dupla recém-formado (dsDNA)

é usado como modelo a PCR. No estágio de hibridização, o primer *reverse* se liga à fita *sense* do dsDNA e o primer *forward* com uma sonda marcada com fluoróforo se ligam à fita *antisense* do DNA. Durante a etapa de extensão, a DNA polimerase estende o primer *forward* e hidrolisa a sonda resultando na liberação do fluoróforo. Em seguida, após a excitação por *laser*, a emissão de fluorescência pode ser capturada pelo termociclador em tempo real. Com cada ciclo de amplificação, o amplicon dsDNA é multiplicado por um fator de 2 vezes com um aumento proporcional no sinal de fluorescência geral. Após 20 a 40 ciclos de amplificação, o RT-PCR está completo. O ciclo no qual o sinal de fluorescência cruza o limiar para positividade é chamado de ciclo limiar (Ct), e os valores de Ct são inversamente proporcionais aos títulos virais presente nas amostras. Assim, o rendimento da reação é aumentado pela repetição dos ciclos no termociclador (Cascella, 2022; Stang, 2021; Nolan, 2006). Em 11 de janeiro de 2020, o primeiro genoma sequenciado de SARS-CoV-2 foi enviado à plataforma GISAID. A partir de então, institutos de pesquisa ao redor do mundo já começavam a elaborar primers e kits para desenvolver a técnica.

Figura 9: Ilustração esquemática do ensaio de RT-qPCR. Fonte: https://www.researchgate.net/publication/351525302_Tools_and_techniques_for_severe_acute_respiratory_syndrome_coronavirus_2_SARS-CoV-2COVID-19_detection/figures?lo=1



Embora o RT-qPCR seja considerado padrão-ouro de diagnóstico de SARS-CoV-2, é uma técnica demorada e cara. Estes são fatores limitantes que impedem que haja dados epidemiológicos confiáveis de transmissão viral em certas populações. Ademais, devido à baixa carga viral nos primeiros dias da infecção, o RT-qPCR pode não ser sensível o suficiente para detectar baixas quantidades de RNA, o que pode levar eventuais falso-negativos. Assim, testes de diferentes técnicas podem ser necessárias para a confirmação do diagnóstico (Roberts, 2021). No início da pandemia, quando as técnicas diagnósticas ainda não haviam sido otimizadas e padronizadas, uma alguns pacientes positivos para COVID-19 foram identificados como falso-negativo devido à baixa sensibilidade dos primers e das sondas utilizadas na época. (Yüce, 2020).

O RT-qPCR em amostras de saliva também tem sido estudado por ter a possibilidade de auto coleta - um método não invasivo, conveniente, indolor e fácil, sem contato direto entre profissionais de saúde e pacientes. Porém, a saliva

contém mais de 3.000 proteínas, 3.000 mRNA, ~50 microRNAs, centenas de metabólitos e mais de 700 espécies de microrganismos, como bactérias e outros vírus (Dawes, 2019), o que pode dificultar o desenvolvimento e a precisão da técnica. Em princípio, os pacientes são informados de que um resultado negativo não é garantia da ausência de infecção viral. No entanto, devido à alta especificidade da análise de RT-qPCR, a detecção de SARS-CoV-2 na saliva pode ser aceitável. Cabe adicionar que um trabalho de revisão observou uma detecção salivar substancial de SARS-CoV-2 em pacientes com *swab* nasofaríngeo negativo (11,6%), o que pode ser um indício de resultados falso-negativos em amostras padrão-ouro (Caixeta, 2021).

1.9.2 *Amplificação isotérmica (LAMP)*

A amplificação isotérmica mediada por loop (LAMP) é um método de amplificação altamente específico, sensível e rápido, levando em torno de uma a duas horas até o resultado. A técnica amplifica o RNA com por meio de quatro a seis *primers* que se ligam a regiões do genoma alvo sob uma temperatura constante de 60– 65°C (Nagamine, 2002; Sharma, 2020; Notomi, 2000). As amostras são armazenadas em meio de transporte de vírus, o RNA extraído é misturado com *primers* e o mix e aquecido até 60–65°C para a reação acontecer. Um número maior de primers pode ser usado para aumentar a sensibilidade e a especificidade e reduzir o tempo de reação (Nagamine, 2002). Os ensaios baseados em LAMP podem ser realizados em um tubo de PCR e produz resultados qualitativos. A presença de turbidez, mudança de cor ou de fluorescência são parâmetros analisados após a reação (Hardinge, 2019). Como limitação pode-se citar a potencial interação *primer-primer*, pois vários *primers* são usados em uma mesma reação (Sahoo, 2016).

1.9.3 *Imunocromatografia - testes rápidos*

Os dispositivos baseados em imunocromatografia lateral em papel (ILP) são pequenos, simples, de fácil aplicação e fornecem resultados qualitativos rapidamente (de 10 a 30 minutos) (Sher, 2017; Yetisen, 2013). Além serem

fabricados de forma econômica e em alto volume (Jiang, 2019; Deng, 2018), já foi demonstrado como ferramenta importante para o diagnóstico viral e vigilância epidemiológica (Lee, 2016; Rohman, 2012; Iqbal, 2019). A amostra (normalmente sangue) é adsorvida no papel cromatográfico e carregadas lateralmente por força capilar. O princípio de detecção é baseado na ligação colorimétrica entre anticorpos IgG e IgM ou algum antígeno de SARS-CoV-2 - usualmente Spike e/ou Nucleocapsídeo - ao seu antagonista fixado no papel cromatográfico. Assim, com base no aparecimento das linhas de teste e controle, ocorrerão resultados positivos, negativos e inválidos - no caso de a linha controle não aparecer (Carter, 2020; Sheridan, 2020). No entanto, os anticorpos podem ser detectados aproximadamente 4 a 8 dias após a infecção (Li, 2020; Luo, 2020). Além disso, reações cruzadas podem fornecer resultados falso-positivos com outros tipos de coronavírus ou vírus influenza (Santiago, 2020).

1.10 *Next Generation Sequencing* (NGS)

O sequenciamento de nova geração (do inglês: *Next Generation Sequencing*) é uma tecnologia de sequenciamento paralelo massivo de alto rendimento. A técnica é usada para determinar a ordem dos nucleotídeos de genomas (ou de regiões específicas) de DNA ou RNA. O NGS revolucionou as ciências biológicas, permitindo aos laboratórios uma ampla variedade de aplicações e estudos de sistemas biológicos. No contexto de SARS-CoV-2, tem sido fundamental na pesquisa e na descoberta das variantes virais e no desenvolvimento e padronização de ferramentas de diagnóstico molecular. Como exemplo, pode-se destacar que por meio dessa técnica o genoma do SARS-CoV-2 foi descrito de forma abrangente, possibilitando a determinação de sua classificação filogenética (Poon, 2020; Zhou, 2020; Wu, 2020; Lu, 2020). Cabe ressaltar que o NGS não está sendo empregado como método diagnóstico, mas em estudos de epidemiologia molecular e filogenética e nas ciências “ômicas” (Conesa, 2019; Chervitz, 2011). Seu uso acaba sendo restrito devido ao alto custo e à necessidade de profissionais altamente qualificados em biologia molecular e bioinformática.

As tecnologias de sequenciamento NGS se tornaram rapidamente os métodos de escolha com várias aplicações em virologia e biologia molecular - incluindo a identificação de novos vírus a partir de amostras em estudos metagenômicos (Kohl, 2015), a reconstrução de sequências completas ou quase completas de genomas e a análise da evolução viral, linhagens e variantes (Domingo, 2012; Chiu, 2019). Uma das vantagens mais relevantes das abordagens baseadas em NGS é que os genomas completos podem ser identificados mesmo em vírus desconhecidos ou pouco caracterizados. No caso do SARS-CoV-2, a segunda e a terceira geração de tecnologias NGS (Slatko, 2018) foram aplicadas com sucesso e vários protocolos específicos de preparação de bibliotecas foram desenvolvidos independentemente por diferentes fabricantes [Pillay, 2017; Paden, 2020; Campos, 2020]. O tipo de amostra (amostras clínicas, amostras ambientais, células cultivadas infectadas), carga viral, procedimentos de extração de RNA, qualidade do RNA e outras considerações devem ser conciliadas com os objetivos para guiar a escolha da estratégia de sequenciamento mais adequada. (Chiara, 2021; Sharma, 2021).

A montagem das bibliotecas, passo que precede o NGS, é feita pela colheita das amostras e a extração de material genético. Esse período pré-analítico é muito importante, pois bibliotecas de qualidade geram sequências com alta cobertura. As sequências disponíveis de SARS-CoV-2 derivam principalmente de amostras de diagnóstico clínico (após a RT-qPCR), com altas cargas virais que permitem a extração de RNA suficiente para o sequenciamento e reconstrução de genomas virais completos ou fragmentos. Alguns estudos relatam que as amostras do trato respiratório inferior podem conter uma carga viral mais alta do que as do trato respiratório superior (Li, 2020). No entanto, cabe apontar que durante o curso da infecção, a carga viral muda dinamicamente entre as diferentes alturas do sistema, bem como entre outros tecidos (Pan, 2020; Zhang, 2020; Yu, 2020; Walsh, 2020; Rokni, 2020; Wang, 2020). Um número muito limitado de sequências de SARS-CoV-2 foi obtido de urina, fezes, sangue e espécimes ambientais, como águas residuais, amostras de ar e "*swabs* ambientais" (GISAID). Nesses casos, a escolha da estratégia e tecnologia de sequenciamento é muito influenciada pela baixa carga viral e consequente

escassez e má qualidade do RNA viral (Rimoldi, 2020; Nemudryi, 2020; Lednicky, 2020).

1.10.1 *Extração de RNA*

A extração de RNA viral requer laboratórios de nível 2 de biossegurança. O RNA pode ser extraído e purificado de amostras clínicas, isolados cultivados ou amostras ambientais (Chiara, 2021). Há disponível uma grande variedade de kits para extração de RNA total ou enriquecimento de RNA viral. As metodologias padrão incluem o uso de guanidina e fenol, que, respetivamente, inibe nucleases para garantir que o RNA viral não seja degradado e desnatura e dissolve proteínas para inativar o vírus. Também é recomendada a adição de RNA transportador poli-A para aumentar o rendimento da extração e tratamento com DNAase para a preparação de bibliotecas para metatranscriptômica (Chiara, 2021). O RNA pode ser analisado qualitativamente, quantificado por espectrofotometria e armazenado a -80°C até o uso. Antes do sequenciamento, a presença e a quantidade de RNA viral podem ser avaliadas usando RT-qPCR por meio dos valores de Ct (Carter, 2020)

1.10.2 *Preparação das bibliotecas*

A preparação das bibliotecas de cDNA é um passo crucial para o sucesso do NGS. Esta etapa prepara amostras de DNA ou RNA para serem compatíveis com a reação e leitura de um sequenciador. As bibliotecas são normalmente criadas fragmentando randomicamente o DNA (em pequenas sequências chamadas de *reads*) e adicionando adaptadores em uma ou em ambas as extremidades - respectivamente, *mate-pair* que também sequencia as duas extremidades de *reads* maiores (de 2000 a 5000 nucleotídeos) e *paired-end*, que proporciona o sequenciamento nas duas extremidades de *reads* com tamanho entre 200 e 500 nucleotídeos. Esses adaptadores contêm sequências complementares que permitem que os fragmentos de DNA se liguem à *flowcell* do chip do sequenciador onde as reações acontecem. Os fragmentos podem então ser amplificados e purificados. (Meredith, 2020; Chiara, 2021; Quick, 2020)

1.10.3 *Sequenciamento*

O princípio desta metodologia, considerando algumas plataformas de sequenciamento, é similar ao método proposto por Sanger, pois há a síntese de uma fita complementar ao DNA alvo utilizando DNA polimerase e nucleotídeos terminadores marcados com diferentes fluoróforos. A fluorescência emitida após a incorporação de cada nucleotídeo é registrada como imagem e no final, por meio de uma decodificação destas imagens, tem-se a sequência estudada (Meredith, 2020).

Durante a etapa de sequenciamento, as bibliotecas são adicionadas a um chip que é colocado no sequenciador. Os *clusters* de fragmentos de DNA são amplificados resultando em milhões de cópias de DNA de fita simples. A seguir, nucleotídeos quimicamente modificados são adicionados à *flowcell* e se ligam à fita molde de DNA por meio de complementaridade. Cada nucleotídeo contém um marcador fluorescente e um terminador reversível que bloqueia a incorporação da próxima base. Posteriormente, ocorre a excitação dos fluoróforos por *laser* fazendo com que os nucleotídeos emitam fluorescência. No sequenciador, um sinal fluorescente indica qual nucleotídeo foi adicionado cada base e o terminador é clivado para que a próxima possa se ligar. Uma imagem contendo a cor fluoróforo excitado e a posição dos *clusters* é capturada na *flowcell*. Em seguida, a extremidade 3’ dos fragmentos é desbloqueada com consequente remoção dos reagentes incorporados no ciclo anterior permitindo o início de um novo ciclo (Chiara, 2021; Illumina, Understanding the NGS worflox; Turcatti et al., 2008; Fedurco et al., 2006). Este processo se repete até que todas as bases de cada fragmento sejam determinadas. Depois de ler a fita de DNA, o chip pode ser lavado para repetir o processo na fita reversa.

1.10.4 *Análise de dados*

Após o sequenciamento, o aparelho identifica os nucleotídeos e a precisão dos resultados. Para a analisar os dados, é possível importar as sequências para ferramentas de análise padrão ou configurar um pipeline próprio para o projeto. Para a formulação dos pipelines existe uma vasta gama de

pacotes, aplicativos, softwares e comandos que podem ser usados de acordo com a arquitetura de cada projeto. Essas ferramentas e pipelines fazem o controle de qualidade e alinhamento de sequências, chamada de variante para análise de mutações e visualização e interpretação de dados na forma de árvores filogenéticas, gráficos, mapas etc.

2 REFERÊNCIAS

CORONAVIRIDAE STUDY GROUP OF THE INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nature Microbiology, mar, 2020, Pg 536–544

WEISS, S.R. Forty years with coronaviruses **Journal of Experimental Medicine**, vol 217, mar, 2020

CORMAN, M.V. *et al*. Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses **Advances in Virus Research**, Vol 100, 2018, Pg 163-188

DE WILDE, A.H., SNIJDER E.J., KIKKERT M., VAN HEMERT M.J. Host Factors in Coronavirus Replication. Em: Roles of Host Gene and Non-coding RNA Expression in Virus Infection. **Current Topics in Microbiology and Immunology**, Vol 419. Springer, 2017

Organização Mundial da Saúde. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Disponível em: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 - Acessado em 17/05/2021

MURRAY P.R. *et al*. **Medical microbiology**. 6ª ed. Mosby; 2009.

MURGOLO, N. *et al*. SARS-CoV-2 tropism, entry, replication, and propagation: Considerations for drug discovery and development. **PLoS Pathogens** 17(2): e1009225, fev, 2021

V'KOVSKI, P. *et al*. Coronavirus biology and replication: implications for SARS-CoV-2. **Nature Reviews Microbiology** Vol 19, Pg 155–170, out, 2021

SNIJDER, E.J. *et al*. Unique and Conserved Features of Genome and Proteome of SARS-coronavirus, an Early Split-off From the Coronavirus Group 2 Lineage. **Journal of Molecular Biology**, Vol 331, Pg 991-1004, ago, 2003

ANGELINI, M.M. *et al*. 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. **mBio**, Vol 4, ago, 2013

ROBSON, F. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. **Molecular Cell** Vol 79, Pg 710-727, set, 2020

DENISON, M.R. *et al*. Coronaviruses. An RNA proofreading machine regulates replication fidelity and diversity, **RNA Biology**, Vol 8, Pg 270-279, mar, 2011

GORBALENYA, A. E. *et al*. Nidovirales: Evolving the largest RNA virus genome. **Virus Research**, Vol 117, Pg 17-37, Abr, 2006

SMITH, E.C. e DENISON, M.R. Implications of altered replication fidelity on the evolution and pathogenesis of coronaviruses. **Current Opinion in Virology**, Vol 2, Pg 519–524, out, 2012

PERALES C. e DOMINGO E. Antiviral Strategies Based on Lethal Mutagenesis and Error Threshold. Em: Domingo E., Schuster P. Quasispecies: From Theory to Experimental Systems. **Current Topics in Microbiology and Immunology**, vol 392. Springer, 2015

CHEN, Y. *et al*. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. **Proceedings of the National Academy of Sciences**. Vol 106, Pg. 3484–3489, mar, 2009

BECARES, M. *et al*. Mutagenesis of Coronavirus nsp14 Reveals Its Potential Role in Modulation of the Innate Immune Response. **American Society for Microbiology Journal of Virology**, Vol 90, Pg. 5399-5414, jun, 2016

GRIBBLE J. *et al*. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. **PLoS Pathogens**. Vol 17, jan, 2021

SABERI A. *et al*. A planarian nidovirus expands the limits of RNA genome size. **PLoS Pathogens**, Vol 14, nov, 2018

MA, Y. et al. Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. **Proceedings of the National Academy of Sciences**. Vol 112, Pg. 9436-9441, jul, 2015

HELMY, Y.A. FAWZY, M. ELASWAD, A. *et al.* The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. **Journal of Clinical Medicine**. Vol 9(4), 1225. 2020

M. KAUR, A. SHARMA, S. KUMAR, G. SINGH, R.P. BARNWAL. SARS-CoV-2: insights into its structural intricacies and functional aspects for drug and vaccine development. **Int. J. Biol. Macromol**. Vol 179, pg. 45-60. 2020

ZIEBUHR J. The Coronavirus Replicase. Em: Enjuanes L. (eds) Coronavirus Replication and Reverse Genetics. **Current Topics in Microbiology and Immunology**, vol 287. Springer, Berlin, Heidelberg, 2005

ADEDEJI A.O. *et al*. Mechanism of Nucleic Acid Unwinding by SARS-CoV Helicase. **PLoS ONE**, Vol 7, mai, 2012

JIA, Z. *et al*. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. **Nucleic Acids Research**. Vol 47, Pg. 6538–6550, jul, 2019

IVANOV, K.A. e ZIEBUHR, J. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. **American Society for Microbiology Journal of Virology**. Vol 78, Pg. 7833-7838, jul, 2004

(b) IVANOV, K.A. *et al*. Multiple Enzymatic Activities Associated with Severe Acute Respiratory Syndrome Coronavirus Helicase. **American Society for Microbiology Journal of Virology**. Vol 78, Pg. 5619-5632, jun, 2004

GAO Y. *et al*. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. **Science**. Vol 368, Pg. 779-782, mai, 2020

WANG, Q. *et al*. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. **Cell**. Vol 182, Pg. 417-428, jul, 2020

YAN, L. *et al*. Architecture of a SARS-CoV-2 mini replication and transcription complex. **Nature Communications**. Vol 11, nov, 2020

DONG E, DU H, GARDNER L. An interactive web-based dashboard to track COVID-19 in real time. **Lancet Infectious Diseases**. Vol 20(5), Pg 533-534, mai 2020

CANDIDO, D.S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. **Science**. Vol 369, Pg. 1255-1260, set, 2020

CDC. Variant classifications. Disponível em: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html - Acessado em 28/12/2021

MINISTÉRIO DA SAÚDE DO BRASIL. Sintomas da COVID-19. Disponível em: https://www.gov.br/saude/pt-br/coronavirus/sintomas - Acessado dia 02/09/2021

SHU, Y. e MCCAULEY, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. **Eurosurveillance**. Vol 22, mar, 2017

ORGANIZAÇÃO MUNDIAL DA SAÚDE. World Health Organization COVID-19 International Dashboard, 2020. Disponível em: https://covid19.who.int/ - Acessado em 12/12/2021

SCHALK, A.F. e HAWN, M.C. An Apparently New Respiratory Disease of Baby Chicks. **Journal of the American Veterinary Medical Association**. Vol 78, Pg. 413-422, 1931

FABRICANT, J. The Early History of Infectious Bronchitis. **Avian Diseases**, Vol 42, American Association of Avian Pathologists, Allen Press, Pg 648–50, 1998

BUSHNELL, L.D. e BRANDLY, C.A. Laryngotracheitis in Chicks. **Poultry Science**. Vol 12, Pg 55-60, 1933

BEACH J.R. A filtrable virus, the cause of infectious laryngotracheitis of chickens. **Journal of Experimental Medicine**. Vol 54, Pg 809-16, Nov, 1931

BEAUDETTE F.R E HUDSON C.B. Cultivation of the virus of infectious bronchitis. **Journal of American Veterinary Medicine Association**. Vol 90, Pg 51–8, 1937

BOURSNELL, M.E.G. *et al*. Completion of the Sequence of the Genome of the Coronavirus Avian Infectious Bronchitis Virus. **Journal Of General Virology**. Vol 68, jan, 1987

Research into the Common Cold. **Nature**. Vol 157, Pg 726–727, 1946

ANDREWES, C. Twenty years' work on the common cold. **Proceedings of the Royal Society of Medicine**. Vol 59, Pg 635–7, Jul, 1966

KENDALL, E.J., BYNOE, M.L., TYRRELL, D.A. Virus isolations from common colds occurring in a residential school. **British Medical Journal**. Vol 2, Pg, 82–6, Jul, 1962

TYRRELL, D.A. e BYNOE, M.L. Cultivation of a Novel Type of Common-cold Virus in Organ Cultures. **British Medical Journal**. Vol 1, Pg 1467–70, Jun, 1965

REAGAN R.L. *et al*. Electron micrograph of the virus of infectious bronchitis of chickens. **The Cornell Veterinarian**. Vol 38, abr 1948

DOMERMUTH, C.H. e EDWARDS, O.F. An Electron Microscope Study of Chorioallantoic Membrane Infected with the Virus of Avian Infectious Bronchitis. **The Journal of Infectious Diseases**. Vol 100, Pg 74–81, Jan 1957

BERRY, D.M. *et al*. The structure of infectious bronchitis virus. **Virology**. Vol 23, Pg 403-407, Jul 1964

DAVID-FERREIRA, J.F. E MANAKER, R.A. An electron microscope study of the development of a mouse hepatitis virus in tissue culture cells. **Journal of Cell Biology**. Vol 24, Pg 57–78, Jan 1965

ALMEIDA, J.D. E TYRRELL, D.A.J. The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. **Journal of General Virology**. Vol 1, Pg 175-8, Abr 1967

MCINTOSH, K. *et al.* Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. **Proceedings of the National Academy of Sciences**. Vol 57, Pg 933-940, Abr 1967

TYRRELL, D.A.J. E ALMEIDA, J.D. Direct electron-microscopy of organ cultures for the detection and characterization of viruses. **Archiv für die gesamte Virusforschung**. Vol 22, Pg 417–425, Set 1967

BECKER W.B. *et al*. Morphogenesis of avian infectious bronchitis virus and a related human virus (strain 229E). **Journal of Virology**. Vol 1, Pg 1019-27, Out 1967

ALMEIDA, J.D. *et al*. Virology: Coronaviruses. **Nature**. Vol 220, Pg 650, Nov 1968

INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES. 1975. Disponível em: https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=201901880 em 10/09/2021)

(a) INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES. 2009. Disponível em: (https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=20140897 Acessado em 10/09/2021)

(b) INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES. 2009. Disponível em: (https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=201901852 Acessado em 10/09/2021)

(c) INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES. 2009. Disponível em: (https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=201901846 Acessado em 10/09/2021)

CARSTENS, E.B. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2009). **Archives of Virology**. Vol 155, Pg 133–146, Jan 2010

WOO, P.C.Y. et al. Coronavirus genomics and bioinformatics analysis. **Viruses**. Vol 2, Pg 1804–20, Ago 2010

VIJGEN, L. *et al*. Genetic variability of human respiratory coronavirus OC43. **Journal of Virology**. Vol 79, Pg 3223–5, 2005

SÁNCHEZ, C.M. *et al*. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. **Virology**. Vol 190, Pg 92–105, Set 1992

BARIC R.S. *et al*. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. **Virology**. Vol 177, Pg 646–56, Ago 1990

BANERJEE A. *et al*. Predicting the recombination potential of severe acute respiratory syndrome coronavirus 2 and Middle East respiratory syndrome coronavirus. **Journal of General Virology**. Vol 101, Pg 1251–60, 2020

KRYAZHIMSKIY, S. PLOTKIN, J.B. The Population Genetics of dN/dS. **PLOS Genetics**. Vol 4(12), 2008

SIMMONDS, P. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. **mSphere**. Vol 24, Jun 2020

MAVIAN, C. MARINI, S. MANES, C. CAPUA, I. PROSPERI, M. SALEMI, M. Regaining perspective on SARS-CoV-2 molecular tracing and its implications **medRxiv**. Mar 2020.

VOLZ, E. MISHRA, S. CHAND, M. BARRETT, J.C. JOHNSON, R. GEIDELBERG, L. HINSLEY, W.R. LAYDON, D.J. DABRERA, G. O'TOOLE, A. AMATO, R. RAGONNET-CRONIN, M. HARRISON, I. JACKSON, B. ARIANI, C.V. BOYD, O. LOMAN, N.J. MCCRONE, J.T. GONÇALVES, S. JORGENSEN, D. MYERS, R. HILL, V. JACKSON, D.K. GAYTHORPE, K. GROVES, N. SILLITOE, J. KWIATKOWSKI, D.P. THE COVID-19 GENOMICS UK (COG-UK) CONSORTIUM, FLAXMAN, S. RATMANN, O. BHATT, S. HOPKINS, S. GANDY, A. RAMBAUT, A. FERGUSON, N.M. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. **medRxiv**. Jan 2021

MENG, B. KEMP, S.A. PAPA, G. DATIR, R. FERREIRA, I.A.T.M. MARELLI, S. HARVEY, W.T. LYTRAS, S. MOHAMED, A. GALLO, G. THAKUR, N. COLLIER, D.A. MLCOCHOVA, P. THE COVID-19 GENOMICS UK (COG-UK) CONSORTIUM. DUNCAN, L.M. CARABELLI, A.M. KENYON, J.C. LEVER, A.M. DE MARCO, A. SALIBA, C. CULAP, K. CAMERONI, E. MATHESON, N.J.

PICCOLI, L. CORTI, D. JAMES, L.C. ROBERTSON, D.L. BAILEY, D. GUPTA, R.K. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. **Cell Reports**. Vol 35. Jun 2021

DAVIES, N.G. ABBOTT, S. BARNARD, R.C. JARVIS, C.I. KUCHARSKI, A.J. MUNDAY, J.D. PEARSON, C.A.B. RUSSELL, T.W. TULLY, D.C. WASHBURNE, A.D. WENSELEERS, T. GIMMA, A. WAITES, W. WONG, K.L.M. VAN ZANDVOORT, K. SILVERMAN, J.D. CMMID COVID-19 WORKING GROUP. COVID-19 GENOMICS UK (COG-UK) CONSORTIUM. DIAZ-ORDAZ, K. KEOGH, R. EGGO, R.M. FUNK, S. JIT, M. ATKINS, K.E. EDMUNDS, W.J. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. **Science**. Vol 9, Abr 2021

SINGH, J. PANDIT, P. MCARTHUR, A.G. Banerjee, A. Mossman, K. Evolutionary trajectory of SARS-CoV-2 and emerging variants. **Virology Journal**. Vol 18, Ago 2021

BUSH, R.M. Predicting adaptive evolution. **Nature Reviews Genetics**. Vol 2, Pg. 387-92. Mai 2021

GALLAHER, W.R. A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-COV-2. **Archives of Virology**. Vol 165, Pg. 2341-2348, Jul 2020

TORTORICI, M.A. VEESLER, D. Structural insights into coronavirus entry. **Advances in Virus Research**. Vol 105, Pg. 93-116. Ago 2019

MCCARTHY, K.R. RENNICK, L.J. NAMBULLI, S. ROBINSON-MCCARTHY, L.R. BAIN, W.G. HAIDAR, G. DUPREX, W.P. Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. **Science**. Vol 371, pg. 1139-1142. Mar 2021

WEISBLUM, Y. SCHMIDT, F. ZHANG, F. DASILVA, J. POSTON, D. LORENZI, J.C. MUECKSCH, F. RUTKOWSKA, M. HOFFMANN, H.H. MICHAILIDIS, E. GAEBLER, C. AGUDELO, M. CHO, A. WANG, Z. GAZUMYAN, A. CIPOLLA, M. LUCHSINGER, L. HILLYER, C.D. CASKEY, M. ROBBIANI, D.F. RICE, C.M. NUSSENZWEIG, M.C. HATZIIOANNOU, T. BIENIASZ, P.D. Escape from

neutralizing antibodies by SARS-CoV-2 spike protein variants. **Elife**. Vol 28. Out 2020

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL. Threat Assessment Brief: Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom. Disponível em: https://www.ecdc.europa.eu/en/publications-data/threat-assessment-brief-rapid-increase-sars-cov-2-variant-united-kingdom - Acessado em 06/04/2022

CENTERS FOR DISEASE CONTROL AND PREVENTION. Science Brief: Emerging SARS-CoV-2 Variants. Disponível em: https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html?CDC_AA_refVal=https%3A%2F%2F

MERCATELLI, D. GIORGI, F.M. Geographic and genomic distribution of SARS-CoV-2 mutations. **Frontiers in Microbiology**. Vol 11. 2020

GUAN, Q. SADYKOV, M. MFARREJ, S. HALA, S. NAEEM, R. NUGMANOVA, R. AL-OMARI, A. SALIH, S. MUTAIR, A.A. CARR, M.J. HALL, W.W. AROLD, S.T. PAIN, A. A genetic barcode of SARS-CoV-2 for monitoring global distribution of different clades during the COVID-19 pandemic. **International Journal of Infectious Diseases**. Vol 100 Pg 216-223. Nov 2020

ANDERSEN, K.G. RAMBAUT, A. LIPKIN, W.I. HOLMES, E.C. GARRY, R.F. The proximal origin of SARS-CoV-2. **Nature Medicine**. Vol 26. Pg 450–452. Mar 2020

TANG, X. WU, C. LI, X. SONG, Y. YAO, X. WU, X. DUAN, Y. ZHANG, H. WANG, Y. QIAN, Z. CUI, J. LU, J. On the origin and continuing evolution of SARS-CoV-2. **National Science Review**. Vol 7. Pg 1012–1023. Jun 2020

CERAOLO, C. GIORGI, F.M. Genomic variance of the 2019-nCoV coronavirus. **Journal of Medical Virology**. Vol 92. Pg. 522-528. Mai 2020

BAUM, A. FULTON, B. O. WLOGA, E. COPIN, R. PASCAL, K. E. RUSSO, V. GIORDANO, S. LANZA, K. NEGRON, N. NI, M. WEI, Y. ATWAL, G. MURPHY, S. A. STAHL, J. N. YANCOPOULOS, G. D. KYRATSOUS, C. A. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. **Science**. Vol 369. Pg. 1014–1018. 2020

GU, H. CHEN, Q. YANG, G. HE, L. FAN, H. DENG, Y. *et al*. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. **Science**. Vol 369. Pg 1603-1607. 2020

LI, Q. WU, J. NIE, J. ZHANG, L. HAO, H. LIU, S. ZHAO, C. ZHANG, Q. LIU, H. NIE, L. QIN, H. WANG, M. LU, Q. LI, X. SUN, Q. LIU, J. ZHANG, L. LI, X. HUANG, W. WANG, Y. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. **Cell**. Vol 182. Pg. 1284-1294. Set 2020

GREANEY, A.J. STARR, T.N. GILCHUK, P. ZOST, S.J. *et al*. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. **Cell Host and Microbe**. Vol 29. Pg 44–57. Jan 2021

JACKSON CB, FARZAN M, CHEN B, CHOE H. Mechanisms of SARS-CoV-2 entry into cells. **Nature Reviews. Molecular Cell Biology**. Vol23(1) Pg 3-20. Jan 2022

CEVIK M, KUPPALLI K, KINDRACHUK J, PEIRIS M. Virology, transmission, and pathogenesis of SARS-CoV-2 **British Medical Journal**. Vol 371, 2020

LUAN, J.; LU, Y.; JIN, X.; ZHANG, L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. **Biochem. Biophys. Res. Commun**. Vol 526(1) Pg 165-169. mar 2020.

WRAPP, D.; WANG, N.; CORBETT, K.S.; GOLDSMITH, J.A.; HSIEH, C.L.; ABIONA, O.; GRAHAM, B.S.; MCLELLAN, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. **Science.** Vol 367, Pg 1260–1263. 2020

WANG K, CHEN W, ZHANG Z, DENG Y, LIAN JQ, DU P, WEI D, ZHANG Y, SUN XX, GONG L, YANG X, HE L, ZHANG L, YANG Z, GENG JJ, CHEN R, ZHANG H, WANG B, ZHU YM, NAN G, JIANG JL, LI L, WU J, LIN P, HUANG W, XIE L, ZHENG ZH, ZHANG K, MIAO JL, CUI HY, HUANG M, ZHANG J, FU L, YANG XM, ZHAO Z, SUN S, GU H, WANG Z, WANG CF, LU Y, LIU YY, WANG QY, BIAN H, ZHU P, CHEN ZN. CD147-spike protein is a novel route for SARS-CoV-2 infection to host cells. **Signal Signal Transduction and Targeted Therapy**. Vol 5(1) Pg 283 Dez 2020

HOFFMANN, M.; KLEINE-WEBER, H.; SCHROEDER, S.; KRÜGER, N.; HERRLER, T.; ERICHSEN, S.; SCHIERGENS, T.S.; HERRLER, G.; WU, N.H.; NITSCHE, A. *et al*. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. **Cell**. Vol 181(2) Pg 271-280. mar 2020.

YANG, N.; SHEN, H.M. Targeting the Endocytic Pathway and Autophagy Process as a Novel Therapeutic Strategy in COVID-19. **Int. J. Biol. Sci**. Vol 16. Pg 1724–1731. 2020

XIA, S.; ZHU, Y.; LIU, M.; LAN, Q.; XU, W.; WU, Y.; YING, T.; LIU, S.; SHI, Z.; JIANG, S. *et al*. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. **Cell. Mol. Immunol**. Vol 17(7). Pg 765-767. Jul 2020.

BAGGA, S. BOUCHARD, M.J. Cell cycle regulation during viral infection. **Methods in Molecular Biology**. Vol 1170. Pg 165-227. 2014

XU, L.H. HUANG, M. FANG, S.G. LIU, D.X. Coronavirus infection induces DNA replication stress partly through interaction of its nonstructural protein 13 with the p125 subunit of DNA polymerase δ. **Journal of Biological Chemistry**. Vol 286(45) Pg 39546-59. nov 2011

ZHANG, L. JACKSON, C.B. MOU, H. OJHA, A. PENG, H. QUINLAN, B.D. RANGARAJAN, E.S. PAN, A. VANDERHEIDEN, A. SUTHAR, M.S. LI, W. IZARD, T. RADER, C. FARZAN, M. CHOE, H. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. **Nature Communications**. Vol 26. Nov 2020

KORBER, B. FISCHER, W.M. GNANAKARAN, S. YOON, H. THEILER, J. ABFALTERER, W. HENGARTNER, N. GIORGI, E.E. BHATTACHARYA, T. FOLEY, B. HASTIE, K.M. PARKER, M.D. PARTRIDGE, D.G. EVANS, C.M. FREEMAN, T.M. DE SILVA, T.I. SHEFFIELD COVID-19 GENOMICS GROUP. MCDANAL, C. PEREZ, L.G. TANG, H. MOON-WALKER, A. WHELAN, S.P. LABRANCHE, C.C. SAPHIRE, E.O. MONTEFIORI, D.C. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. **Cell**. Vol 182. Pg 812-827. Ago 2020

LAURING, A.S. HODCROFT, E.B. Genetic Variants of SARS-CoV-2-What Do They Mean? **Journal of the American Medical Association**. Vol 9. Pg 529-531. Fev 2021

PLANTE, J.A. LIU, Y. LIU, J. XIA, H. JOHNSON, B.A. LOKUGAMAGE, K.G. ZHANG, X. MURUATO, A.E. ZOU, J. FONTES-GARFIAS, C.R. MIRCHANDANI, D. SCHARTON, D. BILELLO, J.P. KU, Z. AN, Z. KALVERAM, B. FREIBERG, A.N. MENACHERY, V.D. XIE, X. PLANTE, K.S. WEAVER, S.C. SHI, P.Y. Spike mutation D614G alters SARS-CoV-2 fitness. **Nature**. Vol 592 Pg 116-121. Abr 2021

HOU, Y.J. CHIBA, S. HALFMANN, P. EHRE, C. KURODA, M. DINNON, K.H. 3RD. LEIST, S.R. SCHÄFER, A. NAKAJIMA, N. TAKAHASHI, K. LEE, R.E. MASCENIK, T.M. GRAHAM, R. EDWARDS, C.E. TSE, L.V. OKUDA, K. MARKMANN, A.J. BARTELT, L. DE SILVA, A. MARGOLIS, D.M. BOUCHER, R.C. RANDELL, S.H. SUZUKI, T. GRALINSKI, L.E. KAWAOKA, Y. BARIC, R.S. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. **Science**. Vol 370. Pg 1464-1468. Dez 2020

MANSBACH, R.A. CHAKRABORTY, S. NGUYEN, K. MONTEFIORI, D.C. KORBER, B. GNANAKARAN, S. The SARS-CoV-2 Spike variant D614G favors an open conformational state. **Science Advances**. Vol 7. Abr 2021

HADFIELD J, MEGILL C, BELL SM, HUDDLESTON J, POTTER B, CALLENDER C, SAGULENKO P, BEDFORD T, NEHER RA. Nextstrain: real-time tracking of pathogen evolution. **Bioinformatics**. Vol 34(23). Pg 4121-4123. Dez 2018

BUTOWT, R. BILINSKA, K. VON BARTHELD, C.S. Chemosensory Dysfunction in COVID-19: Integration of Genetic and Epidemiological Data Points to D614G Spike Protein Variant as a Contributing Factor. **American Chemical Society Chemical Neuroscience**. Vol 11. Pg 3180-3184. Out 2020

RAMBAUT, A. *et al*. Preliminary genomic characterization of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Disponível em https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 - Acessado em 10/11/2021

TEGALLY, H. WILKINSON, E. GIOVANETTI, M. IRANZADEH, A. WILLIAMSON, C. DE OLIVEIRA, T. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. **MedRxiv**. Dez 2020

NEW YORK TIMES. At Rio's Carnival, Vaccine Jabs Are In, Glitter and Samba Out. Disponível em: https://www.nytimes.com/2021/02/16/world/americas/rio-carnival-coronavirus.html - Acessado em 1º de fevereiro de 2022

PROJETO BRASIL.IO ESPECIAL COVID-19. Disponível em: https://brasil.io/covid19/ - Acessado em 15/12/2021

TU, H. AVENARIUS, M.R. KUBATKO, L. HUNT, M. PAN, X. RU, P. GAREE, J. THOMAS, K. MOHLER, P. PANCHOLI, P. JONES, D. Distinct Patterns of Emergence of SARS-CoV-2 Spike Variants including N501Y in Clinical Samples in Columbus Ohio. **BioRxiv**. Jan 2021

WAN, Y. SHANG, J. GRAHAM, R. BARIC, R.S. LI, F. Receptor Recognition by the Novel Coronavirus from Wuhan: An Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. **Journal of Virology**. Vol 94 (7). Mar 2020

SANTOS, J.C. PASSOS, G.A. The high infectivity of SARS-CoV-2 B.1.1.7 is associated with increased interaction force between Spike-ACE2 caused by the viral N501Y mutation. **BioRxiv**. Jan 2021

GRABOWSKI, F. PREIBISCH, G. GIZIŃSKI, S. KOCHAŃCZYK, M. LIPNIACKI, T. SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations. **Viruses**. Vol 13. Mar 2021

CHAW, S.M. TAI, J.H. CHEN, S.L. HSIEH, C.H. CHANG, S.Y. YEH, S.H. YANG, W.S. CHEN, P.J. WANG, H.Y. The origin and underlying driving forces of the SARS-CoV-2 outbreak. **Journal of Biomedical Science**. Vol 27. Jun 2020

KEMP, S.A. COLLIER, D.A. DATIR, R.P. et al. SARS-CoV-2 evolution during treatment of chronic infection. **Nature**. Vol 592. Pg 277–282. Fev 2021

MCCALLUM, M. MARCO, A.D. LEMPP, F. TORTORICI, M.A. PINTO, D. WALLS, A.C. *et al*. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. **BioRxiv**. 2021.

BASCOS, N.A.D. MIRANO-BASCOS, D. SALOMA, C.P. Structural analysis of spike protein mutations in an emergent SARS-CoV-2 variant from the Philippines. **BioRxiv**. 2021.

CANDIDO, D.; CLARO, I.M.; DE JESUS, J.G.; SOUZA,W.M.; MOREIRA, F.R.R.; DELLICOUR, S.; MELLAN, T.A.; DU PLESSIS, L.; PEREIRA, R.H.M.; SALES, F.C.S.; *et al*. Evolution and Epidemic Spread of SARS-CoV-2 in Brazil. **Science**. Vol 369, Pg 1255–1260. 2020

FRANCESCHI VB, FERRAREZE PAG, ZIMERMAN RA, CYBIS GB, THOMPSON CE. Mutation hotspots and spatiotemporal distribution of SARS-CoV-2 lineages in Brazil, February 2020-2021. **Virus Research**. Vol 304. Out, 2021

LAMARCA, A.P.; DE ALMEIDA, L.G.P.; FRANCISCO, R. DA S.; LIMA, L.F.A.; SCORTECCI, K.C.; PEREZ, V.P.; BRUSTOLINI, O.J.; SOUSA, E.S.S.; SECCO, D.A.; SANTOS, A.M.G.; *et al*. Genomic Surveillance of SARS-CoV-2 Tracks Early Interstate Transmission of P.1 Lineage and Diversification within P.2 Clade in Brazil. **medRxiv**. 2021

FERRAREZE PAG, FRANCESCHI VB, MAYER AM, CALDANA GD, ZIMERMAN RA, THOMPSON CE. E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. **Infection, Genetics and Evolution**. Vol 93. Set, 2021

SHEN X, TANG H, MCDANAL C, WAGH K, FISCHER W, THEILER J, *et a*l. SARS-CoV-2 variant B. 1.1. 7 is susceptible to neutralizing antibodies elicited by ancestral Spike vaccines. **Cell Host Microbe**. Vol 29(4) Pg 529-539. 2021

(b) SHEN X, TANG H, PAJON R *et al*. Neutralization of SARS-CoV-2 variants B.1.429 and B.1.351 **New England Journal of Medicine.** Vol 384(24). Pg. 2352-2354. 2021

ALTMANN DM, BOYTON RJ, BEALE R. Immunity to SARS-CoV-2 variants of concern. **Science**. Vol 371(6534). Pg 1103. 2021

REES-SPEAR C, MUIR L, GRIFFITH SA, HEANEY J, ALDON Y, SNITSELAAR JL, THOMAS P, GRAHAM C, SEOW J, LEE N, ROSA A, ROUSTAN C, HOULIHAN CF, SANDERS RW, GUPTA RK, CHEREPANOV P, STAUSS HJ, NASTOULI E; SAFER Investigators, DOORES KJ, VAN GILS MJ, MCCOY LE. The effect of spike mutations on SARS-CoV-2 neutralization. **Cell Reports**. Vol 34(12). mar 2021

J HU, P PENG, K WANG, L FANG, F-Y LUO, A-S JIN, *et al*. Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. **Cell Molecular Immunology**. Vol 18(4). Pg 1061-1063. 2021

MOHAMMADI M, SHAYESTEHPOUR M, MIRZAEI H. The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines**. Brazilian Journal of Infectious Diseases**. Vol 25(4). Ago 2021

CHERIAN S, POTDAR V, JADHAV S, YADAV P, GUPTA N, DAS M, RAKSHIT P, SINGH S, ABRAHAM P, PANDA S, TEAM N. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. **Microorganisms**. Vol 9(7). Jul 2021

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Tracking SARS-CoV-2 variants. Disponível em: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ - Acessado dia 4 de Abril de 2022.

KUMAR, V. SINGH, J. HASNAIN, S. E. SUNDAR, D. Possible Link between Higher Transmissibility of Alpha, Kappa and Delta Variants of SARS-CoV-2 and Increased Structural Stability of Its Spike Protein and hACE2 Affinity. **International journal of molecular sciences**. Vol 22(17), 9131.

MWENDA M, SAASA N, SINYANGE N, BUSBY G, CHIPIMO PJ, HENDRY J, KAPONA O, YINGST S, HINES JZ, MINCHELLA P, SIMULUNDU E, CHANGULA K, NALUBAMBA KS, SAWA H, KAJIHARA M, YAMAGISHI J,

KAPIN'A M, KAPATA N, FWOLOSHI S, ZULU P, MULENGA LB, AGOLORY S, MUKONKA V, BRIDGES DJ. Detection of B.1.351 SARS-CoV-2 Variant Strain - Zambia, December 2020. **Morbidity and Mortality Weekly Report**. Vol 70(8). Pg 280-282. fev 2021

WANG P, NAIR MS, LIU L, *et al*. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7 **Nature**, Vol 593(7857). Pg 130-135. 2021

SLAVOV SN, PATANÉ JS LEISTER, BEZERRA RD, SANTOS B *et al*. Genomic monitoring unveil the early detection of the SARS-CoV-2 B.1.351 lineage (20H/501Y.V2) in Brazil **Journal of Medical Virology**. Vol 93(12). Pg 6782-6787. Dez 2021

ZHOU D, DEJNIRATTISAI W, SUPASA P *et al*. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. **Cell**. Vol 184(9). Pg 2348-2361. 2021

(b) SINGH J, SAMAL J, KUMAR V, SHARMA J, AGRAWAL U, EHTESHAM NZ, SUNDAR D, RAHMAN SA, HIRA S, HASNAIN SE. Structure-Function Analyses of New SARS-CoV-2 Variants B.1.1.7, B.1.351 and B.1.1.28.1: Clinical, Diagnostic, Therapeutic and Public Health Implications. **Viruses**. Vol 9;13(3):439. Mar 2021

HOFFMANN M, ARORA P, GROSS R *et al*. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies **Cell**. Vol 184(9). Pg 2384-2393. 2021

FOCOSI D, MAGGI F. Neutralising antibody escape of SARS-CoV-2 spike protein: Risk assessment for antibody-based Covid-19 therapeutics and vaccines. **Reviews in Medical Virology**. Vol 31(6):e2231. Mar 2021.

FUJINO T, NOMOTO H, KUTSUNA S, UJIIE M, SUZUKI T, SATO R, FUJIMOTO T, KURODA M, WAKITA T, OHMAGARI N. Novel SARS-CoV-2 Variant in Travelers from Brazil to Japan. **Emerging Infectious Diseases**. Vol 27(4). Pg 1243–5. Abr 2021

FREITAS ARR, BECKEDORFF OA, CAVALCANTI LPG, SIQUEIRA AM, CASTRO DB, COSTA CFD, LEMOS DRQ, BARROS ENC. The emergence of novel SARS-CoV-2 variant P.1 in Amazonas (Brazil) was temporally associated

with a change in the age and sex profile of COVID-19 mortality: A population based ecological study. **Lancet Regional Health Americas**. Vol 1:100021. Set 2021

KHAN, A.; ZIA, T.; SULEMAN, M.; KHAN, T.; ALI, S.S.; ABBASI, A.A.; MOHAMMAD, A.; WEI, D.-Q. Higher Infectivity of the SARS-CoV-2 New Variants Is Associated with K417N/T, E484K, and N501Y Mutants: An Insight from Structural Data. **Journal of Cellular Physiology**. Vol 236(10). Pg 7045-7057 Out 2021

STARR, T.N.; GREANEY, A.J.; HILTON, S.K.; ELLIS, D.; CRAWFORD, K.H.D.; DINGENS, A.S.; NAVARRO, M.J.; BOWEN, J.E.; TORTORICI, M.A.; WALLS, A.C. *et al*. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. **Cell**. Vol 182. Pg 1295-1310.e20. 2020

LEUNG, K.; SHUM, M.H.; LEUNG, G.M.; LAM, T.T.; WU, J.T. Early Transmissibility Assessment of the N501Y Mutant Strains of SARS-CoV-2 in the United Kingdom, October to November 2020. **Eurosurveillance** Vol 26(1):2002106. Jan 2021

TANG, J.W.; TAMBYAH, P.A.; HUI, D.S. Emergence of a New SARS-CoV-2 Variant in the UK. **Journal of Infection**. Vol 82(4). e27–e28. 2021

ALI, F.; KASRY, A.; AMIN, M. The New SARS-CoV-2 Strain Shows a Stronger Binding Affinity to ACE2 Due to N501Y Mutant. **Medical Drug Discovery**. Vol 10, 100086. 2021

NELSON, G.; BUZKO, O.; SPILMAN, P.; NIAZI, K.; RABIZADEH, S.; SOON-SHIONG, P. Molecular Dynamic Simulation Reveals E484K Mutation Enhances Spike RBD-ACE2 Affinity and the Combination of E484K, K417N and N501Y Mutations (501Y.V2 Variant) Induces Conformational Change Greater than N501Y Mutant Alone, Potentially Resulting in an Escape Mutant. **BioRxiv.** 2021

MULLEN, J.L.; TSUENG, G.; LATIF, A.A.; ALKUZWENY, M.; CANO, M.; HAAG, E.; ZHOU, J.; ZELLER, M.; MATTESON, N.; ANDERSEN, K.G.; *et al*. 2020. Outbreak.Info. Disponível em: https://outbreak.info - Acessado em 9/7/2021).

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; *et al*. Genomic Evidence of SARS-CoV-2 Reinfection Involving E484K Spike Mutation, Brazil. **Emerging Infectious Diseases**. Vol 10(21). e00410-21. Mai 2021

RESENDE, P.C.; BEZERRA, J.F.; VASCONCELOS, R.H.T.; ARANTES, I.; APPOLINARIO, L.; MENDONÇA, A.C.; PAIXAO, A.C.; DUARTE, A.C.; SILVA, T.; ROCHA, A.S.; *et al*. Severe Acute Respiratory Syndrome Coronavirus 2 P.2 Lineage Associated with Reinfection Case, Brazil, June–October 2020. **Emerging Infectious Diseases**. Vol 27(7). Pg 1789-1794. Jul 2021

LUBINSKI B, FERNANDES MHV, FRAZIER L, TANG T, DANIEL S, DIEL DG, JAIMES JA, WHITTAKER GR. Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. **iScience**. Vol 25(1):103589. Jan 2022.

SIMMONS G, BERTRAM S, GLOWACKA I *et al.* Different host cell proteases activate the SARS-coronavirus spike-protein for cell–cell and virus-cell fusion. **Virology**. Vol 413(2). Pg 265–274. Mai 2011

TSE LV, HAMILTON AM, FRILING T, WHITTAKER GR. A novel activation mechanism of avian influenza virus H9N2 by furin. **Journal of Virology**. Vol 88(3). Pg. 1673–1683. Fev 2014

DENG, X.; GARCIA-KNIGHT, M.A.; KHALID, M.M.; SERVELLITA, V.; WANG, C.; MORRIS, M.K.; SOTOMAYOR-GONZÁLEZ, A.; GLASNER, D.R.; REYES, K.R.; GLIWA, A.S.; *et al*. Transmission, Infectivity, and Neutralization of a Spike L452R SARS-CoV-2 Variant. **Cell**. Vol 184(13). Pg 3426–3437.e8.

MEDRONHO R.A. et al. Epidemiologia. 2ª. ed. Rio de Janeiro: Atheneu, 2009

ROBERTS A, CHOUHAN RS, SHAHDEO D, SHRIKRISHNA NS, KESARWANI V, HORVAT M, GANDHI S. A Recent Update on Advanced Molecular Diagnostic Techniques for COVID-19 Pandemic: An Overview. **Front Immunol**. Vol 14;12:732756. Dez 2021

CUI F, ZHOU HS. Diagnostic methods and potential portable biosensors for coronavirus disease 2019. **Biosensors and Bioelectronics**. Vol 1;165:112349. Out 2020

KAUSHIK A, KHAN R, SOLANKI P, GANDHI S, GOHEL H, MISHRA YK. From Nanosystems to a Biosensing Prototype for an Efficient Diagnostic: A Special Issue in Honor of Professor Bansi D. Malhotra. **Biosensors**. Vol 11(10):359. 2020

LI L, QIN L, XU Z, YIN Y, WANG X, KONG B, BAI J, LU Y, FANG Z, SONG Q, CAO K, LIU D, WANG G, XU Q, FANG X, ZHANG S, XIA J, XIA J. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. **Radiology**. Vol 296(2):E65-E71. Ago 2020

LI Z, YI Y, LUO X, XIONG N, LIU Y, LI S, *et al*. Development and Clinical Application of a Rapid IgM-IgG Combined Antibody Test for SARS-CoV-2 Infection Diagnosis. **J Med Virol** Vol 92. Pg 1518–24. 2020

AHIRWAR R, GANDHI S, KOMAL K, TRIPATHI PP, VYAS MS, KUMAR K, *et al.* Biochemical Composition, Transmission and Diagnosis of SARS-CoV-2. **Bioscience Reports** 41(8):BSR20211238. Ago 2021

ROBERTS A, CHAUHAN N, ISLAM S, MAHARI S, GHAWRI B, GANDHAM RK, *et al*. Graphene Functionalized Field-Effect Transistors for Ultrasensitive Detection of Japanese Encephalitis and Avian Influenza Virus. **Science Reports** Vol 10:14546. 2020

CHUNG M, BERNHEIM A, MEI X, ZHANG N, HUANG M, ZENG X, *et al*. CT Imaging Features of 2019 Novel Coronavirus (2019-NCoV). **Radiology**. Vol 295. Pg 202–7. 2020

AI T, YANG Z, HOU H, ZHAN C, CHEN C, LV W, *et al*. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. **Radiology**. Vol 296. Pg 32–40. 2020

LIU Y, LIU G, ZHANG Q. Deep Learning and Medical Diagnosis. **Lancet**. Vol 394. Pg 1709–10. 2020

NAGY A, BASIOUNI S, PARVIN R, HAFEZ HM, SHEHATA AA. Evolutionary insights into the furin cleavage sites of SARS-CoV-2 variants from humans and animals. **Archives of Virology**. Vol 166(9). Pg 2541-2549. Set 2021

LIU Y, LIU J, JOHNSON BA, XIA H, KU Z, SCHINDEWOLF C, WIDEN SG, AN Z, WEAVER SC, MENACHERY VD, XIE X, SHI PY. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. **BioRxiv**. 2021

LASEK-NESSELQUIST, E.; PATA, J.; SCHNEIDER, E.; GEORGE, K.S. A Tale of Three SARS-CoV-2 Variants with Independently Acquired P681H Mutations in New York State. **MedRxiv.** 2021

HOFFMANN, M.; KLEINE-WEBER, H.; PÖHLMANN, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. **Molecular Cell**. Vol 78(4). Pg 779-784.e5. mai 2020

NAVECA FG, NASCIMENTO V, DE SOUZA VC, CORADO AL, NASCIMENTO F, SILVA G, COSTA Á, DUARTE D, PESSOA K, MEJÍA M, BRANDÃO MJ, JESUS M, GONÇALVES L, DA COSTA CF, SAMPAIO V, BARROS D, SILVA M, MATTOS T, PONTES G, ABDALLA L, SANTOS JH, ARANTES I, DEZORDI FZ, SIQUEIRA MM, WALLAU GL, RESENDE PC, DELATORRE E, GRÄF T, BELLO G. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. **Nature Medicine.** Vol 27(7). Pg 1230-1238. Jul 2021

TAYLOR, L. Covid-19: Omicron drives weekly record high in global infections. **The British Medical Journal**. 376 :o66. Jan 2022

b TAYLOR, L. Covid-19: Brazil sees omicron cases soar but data blackout obscures true impact. **The British Medical Journal**. 376 :o133. Jan 2022

REUTERS. Brazil reels as COVID-19 cases soar; hospitals, economy under pressure. Disponível em: https://www.reuters.com/world/americas/brazil-reels-omicron-spreads-weighing- hospitals-economy-2022-01-14/ - Acessado em 7/2/ 2021.

Pabbaraju K, Wong S, Wong AA, Appleyard GD, Chui L, Pang XL, Yanow SK, Fonseca K, Lee BE, Fox JD, Preiksaitis JK. Design and validation of real-time

reverse transcription-PCR assays for detection of pandemic (H1N1) 2009 virus. **J Clin Microbiol**. 2009 Nov;47(11):3454-60.

LEE MS, CHANG PC, SHIEN JH, CHENG MC, SHIEH HK. Identification and subtyping of avian influenza viruses by reverse transcription-PCR. **Journal of Virology Methods**. Vol 97(1-2). Pg 13-22. set 2001

XU L, LI D, RAMADAN S, LI Y, KLEIN N. Facile biosensors for rapid detection of COVID-19**. Biosens Bioelectron**. Vol 15. 170:112673. Dez 2020

CASCELLA M, RAJNIK M, ALEEM A, *et al*. Features, Evaluation, and Treatment of Coronavirus (COVID-19). Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK554776/ - Acessado em 08/março/2022

STANG A, ROBERS J, SCHONERT B, JÖCKEL KH, SPELSBERG A, KEIL U, CULLEN P. The performance of the SARS-CoV-2 RT-PCR test as a tool for detecting SARS-CoV-2 infection in the population. **The Journal of Infection**. Vol 83(2). Pg 237-279. Ago 2021

NOLAN T, HANDS RE, BUSTIN SA. Quantification of mRNA using real-time RT-PCR. **Nature Protocols**. Vol 1(3). Pg 1559-82. 2006

YÜCE M, FILIZTEKIN E, ÖZKAYA KG. COVID-19 diagnosis -A review of current methods. **Biosensors and Bioelectronics**. Vol 15(172). Pg 112752. Jan 2021

DAWES C, WONG DTW. Role of saliva and salivary diagnostics in the advancement of oral health. **Journal of Dental Research**. Vol 98. Pg 133–41. 2019

CAIXETA DC, OLIVEIRA SW, CARDOSO-SOUSA L, CUNHA TM, GOULART LR, MARTINS MM, MARIN LM, JARDIM ACG, SIQUEIRA WL, SABINO-SILVA R. One-Year Update on Salivary Diagnostic of COVID-19. **Front Public Health**. Vol 21(9):589564. Mai 2021

NAGAMINE K, HASE T, NOTOMI T. Accelerated reaction by loop-mediated isothermal amplification using loop primers. **Molecular Cell Probes**. Vol 16(3). Pg 223-9. Jun 2022

SHARMA S, KABIR MA, ASGHAR W. Lab-on-a-Chip Zika Detection With Reverse Transcription Loop-Mediated Isothermal Amplification-Based Assay for Point-of-Care Settings. **Archives of Pathology and Laboratory Medicine**. Vol 1;144(11). Pg 1335-1343. Nov 2020

NOTOMI T, OKAYAMA H, MASUBUCHI H, YONEKAWA T, WATANABE K, AMINO N, HASE T. Loop-mediated isothermal amplification of DNA. **Nucleic Acids Research**. Vol 15;28(12):E63. Jun 2000

HARDINGE P, MURRAY JAH. Reduced False Positives and Improved Reporting of Loop-Mediated Isothermal Amplification using Quenched Fluorescent Primers. **Sci Rep**. Vol 14;9(1):7400. Mai 2019

SAHOO PR, SETHY K, MOHAPATRA S, PANDA D. Loop mediated isothermal amplification: An innovative gene amplification technique for animal diseases. **Vet World**. Vol 9(5). Pg 465-9. Mai 2016

SHER M, ZHUANG R, DEMIRCI U, ASGHAR W. Paper-based analytical devices for clinical diagnosis: recent advances in the fabrication techniques and sensing mechanisms**. Expert Reviews in Molecular Diagnostics**. Vol 17(4). Pg 351-366. Abr 2017

YETISEN AK, AKRAM MS, LOWE CR. Paper-based microfluidic point-of-care diagnostic devices. **Lab Chip**. Vol 13(12). Pg 2210-51. Jun 2013

JIANG N, AHMED R, DAMAYANTHARAN M, ÜNAL B, BUTT H, YETISEN AK. Lateral and Vertical Flow Assays for Point-of-Care Diagnostics**. Advanced Healthcare Materials**. Vol 8(14):e1900244. Jul 2019

DENG X, WANG C, GAO Y, LI J, WEN W, ZHANG X, WANG S. Applying strand displacement amplification to quantum dots-based fluorescent lateral flow assay strips for HIV-DNA detection. **Biosensors and Bioelectronics**. Vol 15;105. Pg 211-217. Mai 2018

LEE D, SHIN Y, CHUNG S, HWANG KS, YOON DS, LEE JH. Simple and Highly Sensitive Molecular Diagnosis of Zika Virus by Lateral Flow Assays. **Analytical Chemistry**. Vol 20;88(24). Pg 12272-12278. Dec 2016

ROHRMAN BA, LEAUTAUD V, MOLYNEUX E, RICHARDS-KORTUM RR. A lateral flow assay for quantitative detection of amplified HIV-1 RNA. **PLoS One**. Vol 7(9):e45611. 2012

IQBAL, S.M.A., BUTT, N.Z. Design and analysis of microfluidic cell counter using spice simulation. **SN Applied Science**. 1, 1290. Set 2019

CARTER LJ, GARNER LV, SMOOT JW, LI Y, ZHOU Q, SAVESON CJ, SASSO JM, GREGG AC, SOARES DJ, BESKID TR, JERVEY SR, LIU C. Assay Techniques and Test Development for COVID-19 Diagnosis. **ACS Central Science**. Vol 27;6(5). Pg 591-605. Mai 2020

SHERIDAN C. Fast, portable tests come online to curb coronavirus pandemic. **Nature Biotechnology**. Vol 38(5). Pg 515-518. Mai 2020

LI Z, YI Y, LUO X, XIONG N, LIU Y, LI S, SUN R, WANG Y, HU B, CHEN W, ZHANG Y, WANG J, HUANG B, LIN Y, YANG J, CAI W, WANG X, CHENG J, CHEN Z, SUN K, PAN W, ZHAN Z, CHEN L, YE F. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. **Journal of Medical Virology**. Vol 92(9). Pg 1518-1524. Set 2020

LUO Z, ANG MJY, CHAN SY, YI Z, GOH YY, YAN S, TAO J, LIU K, LI X, ZHANG H, HUANG W, LIU X. Combating the Coronavirus Pandemic: Early Detection, Medical Treatment, and a Concerted Effort by the Global Community. **Research (Washington D. C.)**. Vol 16;2020:6925296. Jun 2020

SANTIAGO I. Trends and Innovations in Biosensors for COVID-19 Mass Testing. **Chembiochemistry**. Vol 15;21(20). Pg 2880-2889. Out 2020

POON LLM, PEIRIS M. Emergence of a novel human coronavirus threatening human health. **Nature Medicine**. Vol 26(3):317-319. Mar 2020

ZHOU P, YANG XL, WANG XG, HU B, ZHANG L, ZHANG W, SI HR, ZHU Y, LI B, HUANG CL, CHEN HD, CHEN J, LUO Y, GUO H, JIANG RD, LIU MQ, CHEN Y, SHEN XR, WANG X, ZHENG XS, ZHAO K, CHEN QJ, DENG F, LIU LL, YAN B, ZHAN FX, WANG YY, XIAO GF, SHI ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**. Vol 579(7798). Pg 270-273. Mar 2020

WU F, ZHAO S, YU B, CHEN YM, WANG W, SONG ZG, HU Y, TAO ZW, TIAN JH, PEI YY, YUAN ML, ZHANG YL, DAI FH, LIU Y, WANG QM, ZHENG JJ, XU L, HOLMES EC, ZHANG YZ. A new coronavirus associated with human respiratory disease in China. **Nature**. Vol 579(7798). Pg 265-269. Mar 2020

LU R, ZHAO X, LI J, NIU P, YANG B, WU H, WANG W, SONG H, HUANG B, ZHU N, BI Y, MA X, ZHAN F, WANG L, HU T, ZHOU H, HU Z, ZHOU W, ZHAO L, CHEN J, MENG Y, WANG J, LIN Y, YUAN J, XIE Z, MA J, LIU WJ, WANG D, XU W, HOLMES EC, GAO GF, WU G, CHEN W, SHI W, TAN W. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. **The Lancet**. Vol 395(10224) Pg 565-574. Fev 2020

CONESA A, BECK S. Making multi-omics data accessible to researchers. **Scientific Data.** Vol 6(1):251. Out 2019

CHERVITZ SA, DEUTSCH EW, FIELD D, PARKINSON H, QUACKENBUSH J, ROCCA-SERRA P, SANSONE SA, STOECKERT CJ JR, TAYLOR CF, TAYLOR R, BALL CA. Data standards for Omics data: the basis of data sharing and reuse. **Methods in Molecular Biology**. Vol 719. Pg 31-69. 2011

KOHL C, BRINKMANN A, DABROWSKI PW, RADONIĆ A, NITSCHE A, KURTH A. Protocol for metagenomic virus detection in clinical specimens. **Emerging Infectious Diseases**. Vol 21(1). Pg 48-57. Jan 2015

DOMINGO E, SHELDON J, PERALES C. Viral quasispecies evolution. **Microbiology and Molecular Biology Reviews**. Vol 76(2). Pg 159-216. Jun 2012

CHIU CY, MILLER SA. Clinical metagenomics. **Nature Reviews, Genetics**. Vol 20(6). Pg 341-355. Jun 2019

SLATKO BE, GARDNER AF, AUSUBEL FM. Overview of Next-Generation Sequencing Technologies. **Current Protocols in Molecular Biology**. Vol 122(1):e59. Abr 2018

PILLAY S, GIANDHARI J, TEGALLY H, WILKINSON E, CHIMUKANGARA B, LESSELLS R, MOOSA Y, MATTISON S, GAZY I, FISH M, SINGH L, KHANYILE

KS, SAN JE, FONSECA V, GIOVANETTI M, ALCANTARA LC JR, DE OLIVEIRA T. Whole Genome Sequencing of SARS-CoV-2: Adapting Illumina Protocols for Quick and Accurate Outbreak Investigation during a Pandemic. **Genes (Basel)**. Vol 11(8):949. Ago 2020

PADEN CR, TAO Y, QUEEN K, ZHANG J, LI Y, UEHARA A, TONG S. Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2. **Emerging Infectious Diseases**. Vol 26(10). Pg 2401-2405. Out 2020

CAMPOS GS, SARDI SI, FALCAO MB, BELITARDO EMMA, ROCHA DJPG, ROLO CA, MENEZES AD, PINHEIRO CS, CARVALHO RH, ALMEIDA JPP, AGUIAR ERGR, PACHECO LGC. Ion torrent-based nasopharyngeal swab metatranscriptomics in COVID-19. **Journal of Virology Methods**. Vol 282:113888. Mai 2020

CHIARA M, D'ERCHIA AM, GISSI C, MANZARI C, PARISI A, RESTA N, ZAMBELLI F, PICARDI E, PAVESI G, HORNER DS, PESOLE G. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. **Briefings in Bioinformatics**. Vol 22(2). Pg 616-630. Mar 2021

SHARMA A, BALDA S, APREJA M, KATARIA K, CAPALASH N, SHARMA P. COVID-19 Diagnosis: Current and Future Techniques. **International Journal of Biological Macromolecules**. Vol 193(Pt B). Pg 1835-1844. Dez 2021

LI C, ZHAO C, BAO J, TANG B, WANG Y, GU B. Laboratory diagnosis of coronavirus disease-2019 (COVID-19). **Clinical Chimica Acta, International Journal of Clinical Chemistry**. Vol 510. Pg 35-46. Nov 2020

PAN Y, ZHANG D, YANG P, POON LLM, WANG Q. Viral load of SARS-CoV-2 in clinical samples. **Lancet Infectious Diseases**. Vol 20(4). Pg 411-412. Abr 2020

ZHANG W, DU RH, LI B, ZHENG XS, YANG XL, HU B, WANG YY, XIAO GF, YAN B, SHI ZL, ZHOU P. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. **Emerging Microbes Infection**. Vol 9(1). Pg 386-389. Fev 2020

YU F, YAN L, WANG N, YANG S, WANG L, TANG Y, GAO G, WANG S, MA C, XIE R, WANG F, TAN C, ZHU L, GUO Y, ZHANG F. Quantitative Detection and Viral Load Analysis of SARS-CoV-2 in Infected Patients. **Clinical Infectious Diseases**. Vol 71(15). Pg 793-798. Jul 2020

WALSH KA, JORDAN K, CLYNE B, ROHDE D, DRUMMOND L, BYRNE P, AHERN S, CARTY PG, O'BRIEN KK, O'MURCHU E, O'NEILL M, SMITH SM, RYAN M, HARRINGTON P. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. **Journal of Infection**. Vol 81(3). Pg 357-371. Set 2020

ROKNI M, GHASEMI V, TAVAKOLi Z. Immune responses and pathogenesis of SARS-CoV-2 during an outbreak in Iran: Comparison with SARS and MERS. **Reviews in Medical Virology**. Vol 30(3):e2107. Mai 2020

WANG Y, ZHANG L, SANG L, YE F, RUAN S, ZHONG B, SONG T, ALSHUKAIRI AN, CHEN R, ZHANG Z, GAN M, ZHU A, HUANG Y, LUO L, MOK CKP, AL GETHAMY MM, TAN H, LI Z, HUANG X, LI F, SUN J, ZHANG Y, WEN L, LI Y, CHEN Z, ZHUANG Z, ZHUO J, CHEN C, KUANG L, WANG J, LV H, JIANG Y, LI M, LIN Y, DENG Y, TANG L, LIANG J, HUANG J, PERLMAN S, ZHONG N, ZHAO J, MALIK PEIRIS JS, LI Y, ZHAO J. Kinetics of viral load and antibody response in relation to COVID-19 severity. **Journal of Clinical Investigation**. Vol ;130(10). Pg 5235-5244. Out 2020

RIMOLDI SG, STEFANI F, GIGANTIELLO A, POLESELLO S, COMANDATORE F, MILETO D, MARESCA M, LONGOBARDI C, MANCON A, ROMERI F, PAGANI C, CAPPELLI F, ROSCIOLI C, MOJA L, GISMONDO MR, SALERNO F. Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. **Science of The Total Environment**. Vol 744:140911. Nov 2020

NEMUDRYI A, NEMUDRAIA A, WIEGAND T, SURYA K, BUYUKYORUK M, CICHA C, VANDERWOOD KK, WILKINSON R, WIEDENHEFT B. Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. **Cell Reports Medicine**. Vol 1(6):100098. Set 2020

LEDNICKY, J.A., SHANKAR, S.N., ELBADRY, M.A., GIBSON, J.C., ALAM, M.M., STEPHENSON, C.J., EIGUREN-FERNANDEZ, A., MORRIS, J.G., MAVIAN, C.N., SALEMI, M., CLUGSTON, J.R. AND WU, C.Y. Collection of

SARS-CoV-2 Virus from the Air of a Clinic within a University Student Health Care Center and Analyses of the Viral Genomic Sequence. **Aerosol Air Quality Research**. Vol 20(6). Pg 1167-1171. Jun 2020

MEREDITH LW, HAMILTON WL, WARNE B, *et al*. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of healthcare associated COVID-19: a prospective genomic surveillance study. **Lancet Infectious Diseases** Vol 20(11). Pg 1263–1272. Jul 2020

QUICK J. nCoV-2019 sequencing protocol v3 (LoCost). Disponível em: https://protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye - Acessado em 2/3/2022

TURCATTI, G.; ROMIEU, A.; FEDURCO, M. *et al*. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. **Nucleic Acids Research**. Vol 36(4):e25. Mar 2008.

FEDURCO, M.; ROMIEU, A.; WILLIAMS, S. *et al*. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. **Nucleic Acids Research**. Vol 34(3):e22. Fev 2006.

ILLUMINA. Understanding the NGS workflow. Disponível em: https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html - Acessado em 2/3/2022.

3 OBJETIVOS

3.1 Objetivo geral

O objetivo principal deste trabalho é, além de contextualizar o estado da arte de SARS-CoV-2 em tópicos pertinentes ao escopo do trabalho, estudar a genômica de amostras do RS e apresentar o histórico epidemiológico viral no Brasil.

3.2 Objetivos específicos

- Identificar a prevalência de linhagens de SARS-CoV-2 no RS no período estudado;
- Caracterizar as mutações encontradas;
- Construir e analisar a árvore filogenética das sequências obtidas;
- Entender possíveis rotas de dispersão viral usando abordagens filogenéticas;
- Estabelecer o histórico completo de linhagens e variantes no Brasil;
- Correlacionar o surgimento de linhagens, ondas de COVID-19 e estatísticas populacionais do país;
- Apresentar histórico epidemiológico pormenorizado da pandemia de COVID-19 no período estudado.

4 CONCLUSÃO

O primeiro estudo realizado corrobora a substituição virtual total de linhagens anteriores por P.1 no sul do Brasil nos casos de COVID-19 sequenciados em março de 2020. Além disso, vários casos foram relacionados à nova sublinhagem P.1.2 e sua provável origem foi estabelecida no estado do Rio Grande do Sul. Na época do estudo, a evolução contínua da P.1 era algo muito preocupante, considerando seu impacto clínico e epidemiológico, justificando uma vigilância genômica aprimorada. Algumas limitações devem ser consideradas. Em primeiro lugar, o tamanho da amostra é baixo e não necessariamente representativo do estado do RS. Considerando o número de sequências de cada região intermediária do RS disponíveis no GISAID, é muito provável que a distribuição observada seja consequência de amostragem em diferentes momentos nessas localidades ou simples aleatoriedade. Assim, não deve ser assumida como uma representação verdadeira da diversidade espacial do estado. Como os genomas publicamente disponíveis são resultado de esforços de sequenciamento episódico, especialmente no Brasil, e não contínuo, inferências mais precisas sobre introduções e processos de difusão em contextos regionais e mundiais são limitadas devido à falta de distribuição geográfica e temporal adequada das amostras. Portanto, pesquisas e vigilância constantes são essenciais para desvendar uma caracterização genômica mais precisa do SARS-CoV-2 no Brasil, identificando novas variantes para melhor resposta e controle da sua disseminação.

A revisão dos dados de COVID-19 no Brasil revelou o curso da pandemia no país no período estudado. É um estudo retrospectivo da evolução epidemiológica do SARS-CoV-2 no Brasil envolvendo todos os territórios. Esse tipo de exame minucioso é pertinente para entender o comportamento do vírus. O trabalho possibilitou a identificação das ondas e proporções de casos e mortes em cada região brasileira. Ademais, foi possível observar a ascensão e declínio das principais linhagens e variantes analisadas e, ao mesmo tempo, sua incidência sobre a população. Entende-se que a variante Gamma foi a mais letal na história da pandemia no país, até o momento. Em todas as regiões, foi a

aparente causa da segunda onda e de números alarmantes. Também é evidente a discrepância da contagem e das proporções das variantes entre as diferentes regiões brasileiras. O Brasil possui magnitude continental e isso possibilita a introdução de linhagens em diferentes territórios e em diferentes momentos. Além disso, o país detém larga diversidade genética, cultural e geográfica, colaborando com a heterogeneidade dos resultados apresentados. Um fato que chama a atenção - e pode ser entendido como uma limitação da análise - é o número particularmente baixo de genomas sequenciados em relação ao número de casos confirmados no país. Logo, os resultados apresentados podem não refletir a epidemiologia exata da pandemia no país. No entanto, este estudo ainda é uma rica fonte de informações e poderá ajudar a guiar o desenvolvimento de intervenções não medicamentosas e possíveis soluções para prevenção da doença e proteção da população.

APÊNDICE A Artigo 1- 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

Revista

Pathogens

JCR 3,406 - DOI: https://doi.org/10.3390/pathogens10080988

Título

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

Autores

Vinícius Bonetti Franceschi[1,#], Gabriel Dickin Caldana[2,#], Christiano Perin[3], Alexandre Horn[3], Camila Peter[4], Gabriela Bettella Cybis[5], Patrícia Aline Gröhs Ferrareze[2], Liane Nanci Rotta[2], Flávio Adsuara Cadegiani[6], Ricardo Ariel Zimerman[3,*], Claudia Elizabeth Thompson[1,2,7,*]

[1] Center of Biotechnology, Graduate Program in Cell and Molecular Biology (PPGBCM), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

[2] Graduate Program in Health Sciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre, RS, Brazil

[3] Department of Infection Control and Prevention, Hospital da Brigada Militar, Porto Alegre, RS, Brazil

[4] Laboratório Exame, Novo Hamburgo, RS, Brazil

[5] Department of Statistics, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

[6] Corpometria Institute, Brasília, DF, Brazil

[7] Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brazil

[#] V.B. Franceschi and G.D. Caldana equally contributed to this work, the order of authorship being arbitrary.

* R.A. Zimerman and C.E. Thompson are both responsible for the work coordination, the order of senior authorship being arbitrary.

Abstract

Almost a year after the COVID-19 pandemic had begun, new lineages (B.1.1.7, B.1.351, P.1, and B.1.617.2) associated with enhanced transmissibility, immunity evasion, and mortality were identified in the United Kingdom, South Africa, and Brazil. The previous most prevalent lineages in the state of Rio Grande do Sul (RS, Southern Brazil), B.1.1.28 and B.1.1.33, were rapidly replaced by P.1 and P.2, two B.1.1.28-derived lineages harboring the E484K mutation. To perform a genomic characterization from the metropolitan region of Porto Alegre, we sequenced viral samples to: (i) identify the prevalence of SARS-CoV-2 lineages in the region, the state, and bordering countries/regions; (ii) characterize the mutation spectra; (iii) hypothesize viral dispersal routes by using phylogenetic and phylogeographic approaches. We found that 96.4% of the samples belonged to the P.1 lineage and approximately 20% of them were assigned as the novel P.1.2, a P.1-derived sublineage harboring signature substitutions recently described in other Brazilian states and foreign countries. Moreover, sequences from this study were allocated in distinct branches of the P.1 phylogeny, suggesting multiple introductions in RS and placing this state as a potential diffusion core of P.1-derived clades and the emergence of P.1.2. It is uncertain whether the emergence of P.1.2 and other P.1 clades is related to clinical or epidemiological consequences. However, the clear signs of molecular diversity from the recently introduced P.1 warrant further genomic surveillance.

Keywords: COVID-19; high-throughput nucleotide sequencing; infectious diseases; molecular epidemiology; molecular evolution; phylogeny; severe acute respiratory syndrome coronavirus 2.

pathogens

MDPI

Article

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

Vinícius Bonetti Franceschi [1,†], Gabriel Dickin Caldana [2,†], Christiano Perin [3], Alexandre Horn [3], Camila Peter [4], Gabriela Bettella Cybis [5], Patrícia Aline Gröhs Ferrareze [2], Liane Nanci Rotta [2], Flávio Adsuara Cadegiani [6], Ricardo Ariel Zimerman [3,*,‡] and Claudia Elizabeth Thompson [1,2,7,*,‡]

1 Graduate Program in Cell and Molecular Biology (PPGBCM), Center of Biotechnology, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre 91501-970, RS, Brazil; vinicius.franceschi@ufrgs.br
2 Graduate Program in Health Sciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), Porto Alegre 90050-170, RS, Brazil; gabrield@ufcspa.edu.br (G.D.C.); p.ferrareze@gmail.com (P.A.G.F.); lnrotta@gmail.com (L.N.R.)
3 Department of Infection Control and Prevention, Hospital da Brigada Militar, Porto Alegre 91900-590, RS, Brazil; drchristianoperin@gmail.com (C.P.); amariantehorn@gmail.com (A.H.)
4 Laboratório Exame, Novo Hamburgo 93510-250, RS, Brazil; camilapptr@gmail.com
5 Department of Statistics, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, RS, Brazil; gcybis@gmail.com
6 Corpometria Institute, Brasília 70390-150, DF, Brazil; flavio.cadegiani@unifesp.br
7 Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre 90050-170, RS, Brazil
* Correspondence: ricardoarielzimerman@gmail.com (R.A.Z.); cthompson@ufcspa.edu.br or thompson.ufcspa@gmail.com (C.E.T.); Tel.: +55-(51)-3288-3500 (R.A.Z.); +55-(51)-3303-8889 (C.E.T.)
† These authors contributed equally.
‡ These authors share the senior authorship.

Abstract: Almost a year after the COVID-19 pandemic had begun, new lineages (B.1.1.7, B.1.351, P.1, and B.1.617.2) associated with enhanced transmissibility, immunity evasion, and mortality were identified in the United Kingdom, South Africa, and Brazil. The previous most prevalent lineages in the state of Rio Grande do Sul (RS, Southern Brazil), B.1.1.28 and B.1.1.33, were rapidly replaced by P.1 and P.2, two B.1.1.28-derived lineages harboring the E484K mutation. To perform a genomic characterization from the metropolitan region of Porto Alegre, we sequenced viral samples to: (i) identify the prevalence of SARS-CoV-2 lineages in the region, the state, and bordering countries/regions; (ii) characterize the mutation spectra; (iii) hypothesize viral dispersal routes by using phylogenetic and phylogeographic approaches. We found that 96.4% of the samples belonged to the P.1 lineage and approximately 20% of them were assigned as the novel P.1.2, a P.1-derived sublineage harboring signature substitutions recently described in other Brazilian states and foreign countries. Moreover, sequences from this study were allocated in distinct branches of the P.1 phylogeny, suggesting multiple introductions in RS and placing this state as a potential diffusion core of P.1-derived clades and the emergence of P.1.2. It is uncertain whether the emergence of P.1.2 and other P.1 clades is related to clinical or epidemiological consequences. However, the clear signs of molecular diversity from the recently introduced P.1 warrant further genomic surveillance.

Keywords: COVID-19; severe acute respiratory syndrome coronavirus 2; infectious diseases; high-throughput nucleotide sequencing; molecular evolution; molecular epidemiology; phylogeny

check for updates

Citation: 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.; et al. 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* **2021**, *10*, 988. https://doi.org/10.3390/pathogens10080988

Academic Editor: Tomomi Takano

Received: 29 June 2021
Accepted: 29 July 2021
Published: 5 August 2021

1. Introduction

After its initial emergence in Wuhan (China) in late 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread rapidly around the world leading to the

COVID-19 pandemic officially recognized in March 2020 [1]. As of May 17, 2021, >163 million cases and >3.3 million deaths have been confirmed. In Brazil, the third most affected country by COVID-19, >15.6 million cases and >435,000 deaths have been reported. From a virological standpoint, this could be related to the continental magnitude of Brazil, leading to multiple viral introductions [2] and the recent emergence of a novel variant of concern (VOC) presenting enhanced infectiousness.

Rio Grande do Sul (RS) is the southernmost state in Brazil. It is bordered southerly by Uruguay, westerly by Argentina, and northerly by the state of Santa Catarina, Brazil. With an estimated population of 11.5 million inhabitants and 39.79 inhabitants per square kilometer, RS is the 6th most populous and the 13th most densely populated state in the country [3]. Since 2017, the Brazilian Institute of Geography and Statistics (IBGE) has divided RS into eight intermediate geographic regions: Porto Alegre, Pelotas, Uruguaiana, Santa Maria, Santa Cruz do Sul/Lajeado, Ijuí, Passo Fundo, and Caxias do Sul [4]. The municipality of Porto Alegre is the state capital, and its metropolitan region comprises 34 municipalities aggregating >4 million inhabitants (~2% of the country's population) and is characterized by intense transit of people. COVID-19 was firstly confirmed in RS on 10 March 2020, in a returning traveler from Italy [5]. The state implemented in May 2020 the "controlled distancing system", which divided the state into 21 regions and 26 areas (R01–R26) and consisted of a flag system establishing restrictions and flexibilizations of non-essential activities based on the weekly occupation of intensive care unit beds and expected deaths. However, due to economic losses, the more amenable "shared management system" allowed mayors to appeal in court and adopt less restrictive flag protocols [6].

Important shifts of COVID-19 epicenters have occurred during 2020, starting with Asia and followed by Europe, North America, and South America. After months of relatively slow evolution, novel VOCs (e.g., B.1.1.7, B.1.351, P.1, and B.1.617.2) harboring a constellation of signature mutations in the spike protein have emerged [7]. More recently, the World Health Organization (WHO) assigned labels for the VOCs based on Greek letters. Therefore, the four VOCs are named, respectively, Alpha, Beta, Gamma, and Delta. These lineages independently arose in the United Kingdom [8], South Africa [9], Brazil [10,11], and India [12,13] and have fueled secondary outbreaks in places where they have emerged, despite previous rates of seroprevalence of up to 75% [14]. The city of Manaus (Amazonas, Brazil), the probable place of origin of the P.1 lineage, faced a major second wave of COVID-19. An explosive resurgence of cases and deaths became evident in mid-December 2020. Since the P.1 variant carries multiple mutations of potential biological significance (especially E484K, K417T, and N501Y in the receptor-binding domain (RBD) of the spike protein): (i) some key substitutions may lead to the immunity evasion; (ii) higher transmissibility when compared with pre-existing lineages has been characterized; (iii) this VOC has been the focus of increased surveillance and deserves being studied in greater detail [15]. After this outbreak, almost all Brazilian states experienced increases in the number of cases, hospitalizations, intensive care unit (ICU) admissions, and deaths, resulting in a reemergence of the public health crisis previously experienced in the first wave of COVID-19 [16].

The diversity of SARS-CoV-2 during the first epidemic wave in Brazil was mainly composed of B.1.1.28 and B.1.1.33 lineages [2,17], although the very low sequencing rate across the country has limited these estimates [17]. However, these previous lineages were rapidly replaced by P.1 and P.2 in late 2020 and early 2021, which are both derived from the common ancestor B.1.1.28 and harbor concerning mutations in the spike protein (e.g., E484K and N501Y) [17,18]. In RS state, the most common lineages identified up to May 2021 are still B.1.1.33 (n = 290) and B.1.1.28 (n = 238) [19,20]. Nevertheless, P.1 has emerged as the most prevalent lineage sequenced in more recent samples [19]. Recently, newer mutations were detected in addition to the original set presented in P.1, giving rise to the sublineage P.1.2 [21]. P.2 probably emerged in Rio de Janeiro state (Southeast) [22], but it was also found in several municipalities of RS state as of October 2020 [23,24]. The first P.1 infection in the state was once thought to be in a patient from Gramado city in

February 2021 [25]. However, in a more recent study, the actual first P.1 was detected on 30 November 2020. This happened in a patient with comorbidities from Campo Bom city, who was reinfected by the P.2 lineage on 11 March 2021 [26].

Even though RS was one of the least affected Brazilian states in the first epidemic wave, it suffered a pronounced increase in cases in late 2020 [16]. In February 2021, the progressive increases in cases and hospitalizations (3.8-fold) led to the collapse of the local state healthcare system. Since recent findings of the widespread dissemination of the SARS-CoV-2 lineage P.1 in Brazil have been confirmed, we sequenced samples from patients from the metropolitan region of Porto Alegre to: (i) identify the prevalence of SARS-CoV-2 lineages in the region, the state, and bordering countries/regions; (ii) characterize the mutation spectra; (iii) hypothesize possible viral dispersal routes by using phylogenetic and phylogeographic approaches.

2. Results

2.1. Epidemiological Information

Of the 56 samples of hospitalized patients between March 9 and 17 2021, 75.0% (n = 42) of them were male, and the mean age was 37.2 years (interquartile range (IQR): 13.5 years). The mean cycle threshold (Ct) value for the first RT-qPCR conducted at Laboratório Exame was 19.12 cycles (median: 18.00; IQR: 6.00 cycles). Forty-seven (83.9%) had contact with a confirmed or suspected case. The majority of them were from the RS state capital, Porto Alegre (n = 32; 57.1%). In total, 51 (91.1%) were from the intermediate geographic region of Porto Alegre and 5 (8.9%) from the intermediate region of Santa Maria (Table 1 and Figure S2C).

2.2. SARS-CoV-2 Mutations and Lineages

Consensus SARS-CoV-2 genomes were obtained with an average coverage depth of 813.2× (median: 820.6×; IQR: 184.7×) (Figure S6). We detected 175 different mutations comprising all samples (Figure 1A). The ORF1ab carried 102 (58.3%) replacements followed by spike (n = 24; 13.7%), nucleocapsid (n = 18; 10.3%), ORF3a (n = 14, 8.0%), ORF7a (n = 6; 3.43%), ORF8 (n = 5; 2.86%), and membrane (n = 3; 1.7%) genes. Remarkably, 50% of the spike substitutions occurred in only one genome, and, of these, nine (75.0%) were missense (Table S1). Fifty-nine (33.7%) mutations were identified in two or more sequences. From these, 36 (61.0%) are non-synonymous (missense), 21 (35.6%) are synonymous, 1 (1.7%) is intergenic at 5′ Untranslated Region (UTR), and 1 (1.7%) is an inframe deletion. Highly frequent (≥10 genomes) mutations were found in 34 genomic positions, 24 (70.5%) being missense and 9 (26.5%) synonymous. Fifteen substitutions (10 in the spike protein: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, and T1027I) are P.1 lineage-defining mutations (Figure 1B and Table 2). The only P.1 defining replacement not found at high frequency in our study was the deletion in ORF1ab (del 11288:11296), called in only four genomes. Deletions overlapping annealing sites of amplicon primers are associated with a strong decrease in the PCR efficiency prior to sequencing, leading to low genomic coverage [27]. Then, after applying a stringent coverage depth filter (DP > 50) for calling the genomic positions in the consensus sequences, this deleted region was flagged as low coverage.

Most importantly, other positions presenting single nucleotide polymorphisms (SNPs) reached the appropriate threshold, since a point mutation is generally unable to cause dropouts.

After comparing the frequency of mutations from the recently sequenced samples and the Brazilian P.1 genomes, we observed a combination of mutations that stood out in a significant proportion (n = 11; 19.6%) compared with previous P.1 sequences from Brazil. This combination was previously described [21] and gave rise to the P.1.2 lineage, which harbors three ORF1ab replacements (synC1912T, D762G, and T1820I), one in ORF3a (D155Y), and one in N protein (synC28789T) (Table 2). Additionally, two of these genomes (18.2%) carry T11296G (ORF1ab nsp6: F3677L) and eight (72.7%) harbor G25641T (ORF3a: L83F) substitutions. Another cluster, made of four local genomes and subsequently named

Clade 2, was also detected. This clade possesses three defining mutations (ORF1ab nsp4: V2862L, synC10507T, and ORF3a: M260K), but it does not fall into a lineage designation at this moment but deserves further monitoring (Figure S1).

Table 1. Epidemiological characteristics of the 56 sequenced samples from Rio Grande do Sul, Southern Brazil.

Study ID (HBM-RS)	GISAID ID (EPI_ISL_)	Cycle Threshold	Municipality of Residence	Gender	Age Group	Lineage	Contact with Confirmed Case
39468	2139494	16	São Leopoldo	Male	30–39	P.1	Yes
39469	2139495	19	Porto Alegre	Female	20–29	P.1.2	Yes
39470	2139496	19	Porto Alegre	Male	60–69	P.1	No
39471	2139497	18	Porto Alegre	Male	20–29	P.1	Yes
39472	2139498	17	Gravataí	Male	30–39	P.1	No
39473	2139499	26	Cachoeira do Sul	Female	20–29	P.1	Yes
39474	2139500	18	Gravataí	Male	30–39	P.1	Yes
39475	2139501	18	Porto Alegre	Female	20–29	P.1	No
39476	2139502	15	Porto Alegre	Male	20–29	P.1	Yes
39477	2139503	21	Porto Alegre	Male	30–39	P.1	Yes
39478	2139504	15	Cachoeira do Sul	Male	30–39	P.1	Yes
39479	2139505	22	Porto Alegre	Male	50–59	P.1	Yes
39480	2139506	17	Novo Hamburgo	Male	40–49	P.1	Yes
39481	2139507	14	Porto Alegre	Female	70–79	P.1	Yes
39482	2139508	14	Porto Alegre	Female	80–89	P.1.2	No
39483	2139509	13	Gravataí	Male	30–39	P.1	Yes
39484	2139510	20	Porto Alegre	Male	20–29	P.1	Yes
39485	2139511	16	Porto Alegre	Male	50–59	P.1	Yes
39486	2139512	27	Porto Alegre	Male	30–39	P.2	No
39487	2139513	14	São Sebastião do Caí	Male	40–49	P.1.2	Yes
39488	2139514	28	Santo Antônio da Patrulha	Male	70–79	P.1	Yes
39489	2139515	27	Porto Alegre	Female	20–29	P.1.2	Yes
39490	2139516	18	Porto Alegre	Male	20–29	P.1	Yes
39491	2139517	15	Alvorada	Female	20–29	B.1.1.28	Yes
39492	2139518	17	Gravataí	Female	30–39	P.1	Yes
39493	2139519	22	Canoas	Male	30–39	P.1.2	Yes
39494	2139520	17	Porto Alegre	Female	30–39	P.1	No
39495	2139521	17	Porto Alegre	Male	30–39	P.1	Yes
39496	2139522	17	Canoas	Female	30–39	P.1	Yes
39497	2139523	21	Porto Alegre	Male	40–49	P.1.2	Yes
39498	2139524	20	Porto Alegre	Female	40–49	P.1	Yes
39499	2139525	22	Portão	Male	30–39	P.1	Yes
39500	2139526	11	Porto Alegre	Male	20–29	P.1.2	Yes
39501	2139527	14	Santa Maria	Male	20–29	P.1.2	Yes
39502	2139528	21	Porto Alegre	Male	30–39	P.1	Yes
39503	2139529	16	Porto Alegre	Male	30–39	P.1	No
39504	2139530	21	Gravataí	Male	40–49	P.1	Yes
39505	2139531	13	Porto Alegre	Male	30–39	P.1.2	Yes
39506	2139532	23	Porto Alegre	Female	40–49	P.1	Yes
39507	2139533	28	Canoas	Female	30–39	P.1	Yes
39508	2139534	22	Porto Alegre	Male	20–29	P.1	Yes
39509	2139535	23	Alvorada	Male	20–29	P.1	Yes
39510	2139536	19	Canoas	Male	50–59	P.1.2	Yes
39511	2139537	22	Porto Alegre	Male	30–39	P.1	No
39512	2139538	25	Cachoeira do Sul	Female	40–49	P.1	Yes
39513	2139539	23	Santa Maria	Male	40–49	P.1	Yes
39514	2139540	15	Porto Alegre	Male	30–39	P.1	No
39515	2139541	21	Porto Alegre	Male	20–29	P.1	Yes
39516	2139542	28	Porto Alegre	Male	50–59	P.1	Yes
39517	2139543	17	Sapiranga	Male	30–39	P.1	Yes
39518	2139544	17	Porto Alegre	Male	30–39	P.1.2	Yes
39519	2139545	23	Porto Alegre	Male	20–29	P.1	Yes
39520	2139546	15	Campo Bom	Male	20–29	P.1	Yes
39521	2139547	15	Porto Alegre	Male	20–29	P.1	Yes
39522	2139548	21	Porto Alegre	Male	50–59	P.1	Yes
39523	2139549	18	Porto Alegre	Male	20–29	P.1	Yes

All samples were nasopharyngeal swabs collected during 9–17 March 2021, from residents of RS state. Study ID: Study identifier only known by study investigators. The Ct values are related to the first RT-qPCR conducted at Laboratório Exame.

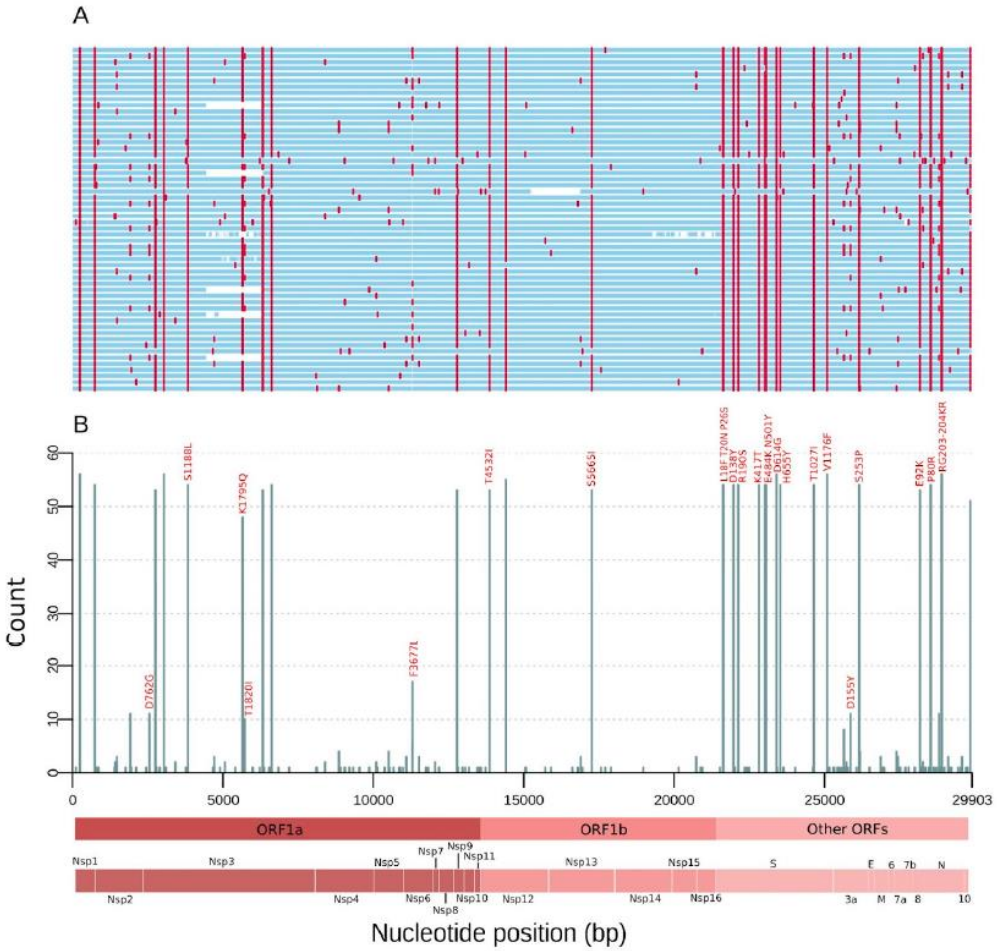


Figure 1. Mutations of the SARS-CoV-2 genomes from RS state, Southern Brazil sampled in March 2021. (**A**) Genome map for the 56 genomes sequenced. Nucleotide substitutions are colored in red and blank regions represent low sequencing coverage. (**B**) Count of single nucleotide polymorphisms (SNPs) per SARS-CoV-2 genome position along the 56 genomes. These mutations are corresponding to the red lines in (**A**), and only missense substitutions represented by >10 sequences have their respective amino acid changes indicated above the bars. Main open reading frames (ORFs) and SARS-CoV-2 proteins are indicated at the bottom to allow a rapid visualization of the viral proteins affected.

Considering PANGO lineages, 54 genomes (96.4%) were designated as P.1, one (1.8%) as P.2, and one (1.8%) as B.1.1.28. Even without being classified according to the Pango-designation's most updated version, the P.1.2 lineage was present in 11/54 (20.4%) of the P.1 sequences (https://github.com/cov-lineages/pango-designation/issues/56; accessed on 6 May 2021) (Figure S1).

2.3. Lineage Distribution in Neighboring Countries and Brazilian Regions

RS state shares borders with Argentina to its west (Figure S2), leading to the transit of people at the frontiers. From March to April 2020, B.1 was the most prevalent lineage in this bordering country. B.1.499 and N.3 were abundant from May to July when the N.5 started to rise and surpassed B.1.499 in November 2020 (Figure 2A). Importantly, N.3 and N.5 are derived from the B.1.1.33 lineage widespread in Brazil (https://cov-lineages.org/lineages.html; accessed on 4 May 2021). The P.2 lineage, which initially emerged in

Brazil [22] and derived from another Brazilian disseminated lineage (B.1.1.28), was firstly found in November 2020, and, by January and February 2021, it had already outnumbered the other lineages in Argentina (Figure 2A).

Table 2. Detailed description and frequency of mutations found in our 56 sequences compared with all Brazilian P.1 sequences until 26 April 2021.

Genomic Position	Effect	Amino Acid Change	Gene/Region	Product	Frequency Our Study (%)	Frequency in Brazilian's P.1 (%)
C241T	Intergenic	NA	5′ UTR	NA	100.0	97.2
T733C	Synonymous	D156D	ORF1ab	Leader Protein	96.4	99.9
C1912T	Synonymous	S549S		nsp2	19.6	1.4
A2550G	Missense	D762G			19.6	1.5
C2749T	Synonymous	D828D			94.6	99.7
C3037T	Synonymous	F924F			100.0	99.9
C3828T	**Missense**	**S1188L**			96.4	95.3
A5648C	**Missense**	**K1795Q**		nsp3	85.7	100.0
C5724T	Missense	T1820I			17.9	2.3
A6319G	Synonymous	P2018P			87.5	99.5
A6613G	Synonymous	V2116V			96.4	99.8
T11296G	Missense	F3677L		nsp6	30.4	8.2
C12778T	Synonymous	Y4171Y		nsp9	94.6	98.9
C13860T	Missense	T4532I		RdRp	94.6	99.8
C14408T	Synonymous	L4715L			98.2	96.8
G17259T	Missense	S5665I		Helicase	94.6	99.7
C21614T	**Missense**	**L18F**	S	Surface Glycoprotein	96.4	99.9
C21621A	**Missense**	**T20N**			94.6	99.8
C21638T	**Missense**	**P26S**			96.4	99.1
G21974T	**Missense**	**D138Y**			96.4	100.0
G22132T	**Missense**	**R190S**			96.4	98.4
A22812C	**Missense**	**K417T**			96.4	83.4
G23012A	**Missense**	**E484K**			96.4	99.9
A23063T	**Missense**	**N501Y**			96.4	99.8
A23403G	Missense	D614G			100.0	97.7
C23525T	**Missense**	**H655Y**			96.4	100.0
C24642T	**Missense**	**T1027I**			96.4	99.9
G25088T	Missense	V1176F			100.0	99.9
G25855T	Missense	D155Y	ORF3a	ORF3a Protein	19.6	1.6
T26149C	**Missense**	**S253P**			94.6	98.7
G28167A	**Missense**	**E92K**	ORF8	ORF8 Protein	94.6	99.8
C28512G	**Missense**	**P80R**	N	Nucleocapsid Phosphoprotein	96.4	98.3
C28789T	Synonymous	Y172Y			19.6	1.3
AGTAGGG 28877–28883 TCTAAAC	Missense	RG203-204KR			96.4	99.8

Original bases or amino acids are represented before the genomic coordinate, while the mutated ones are presented after. Only mutations observed in more than 10 genomes from this study are shown. P.1 lineage-defining mutations are highlighted in bold. P.1.2 (new lineage) defining replacements are underlined and marked with gray background color. UTR, untranslated region; ORF, open reading frame; S, spike; N, nucleocapsid; nsp, nonstructural protein; RdRp, RNA-dependent RNA polymerase.

In the entire Brazil, despite early introductions of B.1 and B.1.1, lineages B.1.1.28 and B.1.1.33 were most abundant from March to October 2020. In October, P.2 already represented an important portion of the sequences, and, by November, it had already surpassed B.1.1.33. In December 2020 and January 2021, with the emergence of P.1, this lineage and P.2 already became the most prevalent, while, between February and April, P.1 replaced all other lineages (Figure 2A,B). Some fluctuations evidently occurred in different Brazilian regions, such as a prevalence of more local lineages (e.g., B.1.195 and B.1.1.378 in the Northern region, B.1.1 and N.9 in the Northeast, and B.1.1.7 in the Southeast and

Centre-West regions) (Figure S3). In RS, a similar landscape was observed compared to the Brazilian scenario. B.1.1.28 and B.1.1.33 were most prevalent until October 2020, when P.2 emerged and remained until January 2021 along with B.1.1.28 as the most prevalent. After the introduction of P.1 (January 2021), this lineage practically supplanted the others in February and March 2021 (Figure 2A,B).

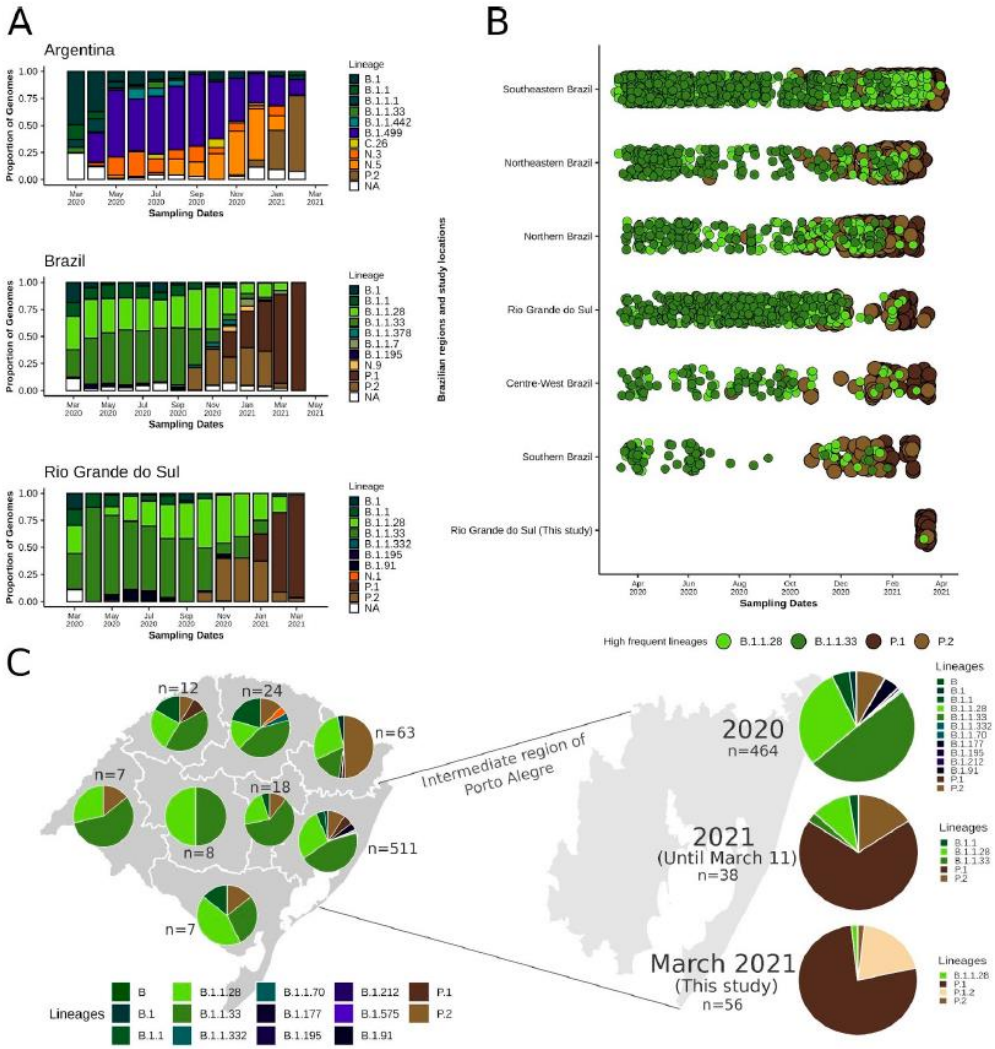


Figure 2. Distribution of SARS-CoV-2 lineages across time in Argentina, other Brazilian regions, and RS state. The other bordering country (Uruguay) was not included due to the limited number of samples available (n = 135). (**A**) Per month lineage distribution in Argentina, all of Brazil, and RS state. Only the 10 most prevalent lineages were considered. (**B**) Timeline showing the distribution of the most prevalent lineages until the end of 2020 (B.1.1.28 and B.1.1.33) and from the end of 2020 onward (P.1 and P.2) for the five Brazilian regions, RS state, and considering only this study (March 2021). (**C**) Map of RS state divided into eight intermediate regions, as defined by IBGE, displaying the proportion of the lineages in each area. The region of Porto Alegre was amplified and the lineage frequencies from 2020, 2021 and the present study are presented on the right. NA, other lineages.

After dividing RS into the intermediate regions proposed by IBGE (Figure S2), it was possible to gain insights into the dynamics of the lineages in the state, despite the low sample size of some regions (Figure 2C). In most regions, the lineages B.1.1.28 and B.1.1.33 were more prevalent, but P.2 was also detected. In fact, in the Caxias do Sul region, more P.2 (n = 31; 49.2%) were sequenced in relation to other lineages. Since the Porto Alegre region has a larger sample size, we divided its results by year to check the most recent (2021) evolutionary abundance. In 2020, B.1.1.33 (n = 229; 49.3%) and B.1.1.28 (n = 137, 29.5%) were the most abundant, followed by P.2 (n = 37; 8.0%). In 2021, P.1 (n = 26, 68.4%) and P.2 (n = 6, 15.8%) have already outperformed the other lineages (Figure 2C). In our study from March 2021, 96.4% of the samples were classified as P.1. We were able to identify a new P.1 sublineage (P.1.2) in 11 (20.4%) genomes from four different municipalities (Porto Alegre, Canoas, São Sebastião do Caí, and Santa Maria), demonstrating the possible diversification of P.1 and its spread within RS (Figure 2C and Figure S1).

2.4. Maximum Likelihood Phylogenomic Analysis

After running the Nextstrain workflow using quality control and subsampling approaches, we obtained a dataset of 8635 time- and geographical-representative genomes. From these, 861 were from Africa, 1370 from Asia, 2219 from Europe, 481 from North America, 218 from Oceania, and 3486 from South America. Brazil was represented by 2608 sequences and RS state by 730 sequences (56 from this study and 674 available in GISAID) (Table S2).

The time-resolved ML phylogenetic tree confirmed the PANGO lineages assigned, since 54 genomes (96.4%) grouped with P.1 representatives, 1 (1.8%) with B.1.1.28, and 1 (1.8%) with P.2 sequences. We also observed a strong correlation between genetic distances and sampling dates ($R^2 = 0.71$). The P.1 sequences were grouped above the regression line, showing higher evolutionary rates than the other lineages in the SARS-CoV-2 phylogeny, as observed in other studies [11,14]. We highlighted the most abundant global lineages present in RS state that passed the quality control criteria (B.1.1 (n = 32), P.1 (n = 83), P.2 (n = 83), B.1.1.28 (n = 203), and B.1.1.33 (n = 286)). We also noticed the high abundance of B.1.1.28 and B.1.1.33 lineages until October and November 2020, followed by the rise and establishment of P.2 and P.1, respectively (Figure 3A).

The only B.1.1.28 sequence identified in this study (RS-HBM-39491) branched in a clade represented by 30 sequences, mostly represented by Southeastern Brazilian (n = 17; 56.6%) genomes. This clade is supported by the ORF1ab:synC15810T mutation and includes a subclade characterized by the ORF1ab:L4182F mutation, where the local sequence is placed together with four samples from São Paulo (SP), one from Portugal, and one from Chile (Figure S4). Most importantly, this local genome harbors seven other mutations: ORF1ab:T2087I (nsp3), D3022N (nsp4), N3970S (nsp8), V4436A (RNA-dependent RNA polymerase), synC13724T, synG18973A, and intergenic:G29736T. Additionally, the only P.2 sequence from this study (RS-HBM-39486) formed a separate clade composed of 20 sequences from several Brazilian states (10 from RS, 1 from Paraná, 1 from SP, and 1 from Maranhão), 5 from Argentina, 1 from USA, and 1 from Norway (Figure S5). This clade is characterized by the ORF1ab:synT6218C (nsp3) mutation. Moreover, this local sequence accrued seven specific mutations: ORF1ab:synA7201G, S2926F (nsp4), V6871A (2′-O-ribose methyltransferase), S:G1251V, ORF7a:G38V, N:synC28333T, and intergenic:G29688T.

To get a more detailed understanding of the P.1 diffusion throughout Rio Grande do Sul, other Brazilian regions, and worldwide, we built an ML tree of 4499 genomes belonging to this lineage (Table S3). P.1 sequences from this study were allocated into several distinct branches, suggesting multiple introductions and the formation of different P.1-derived clades and clusters.

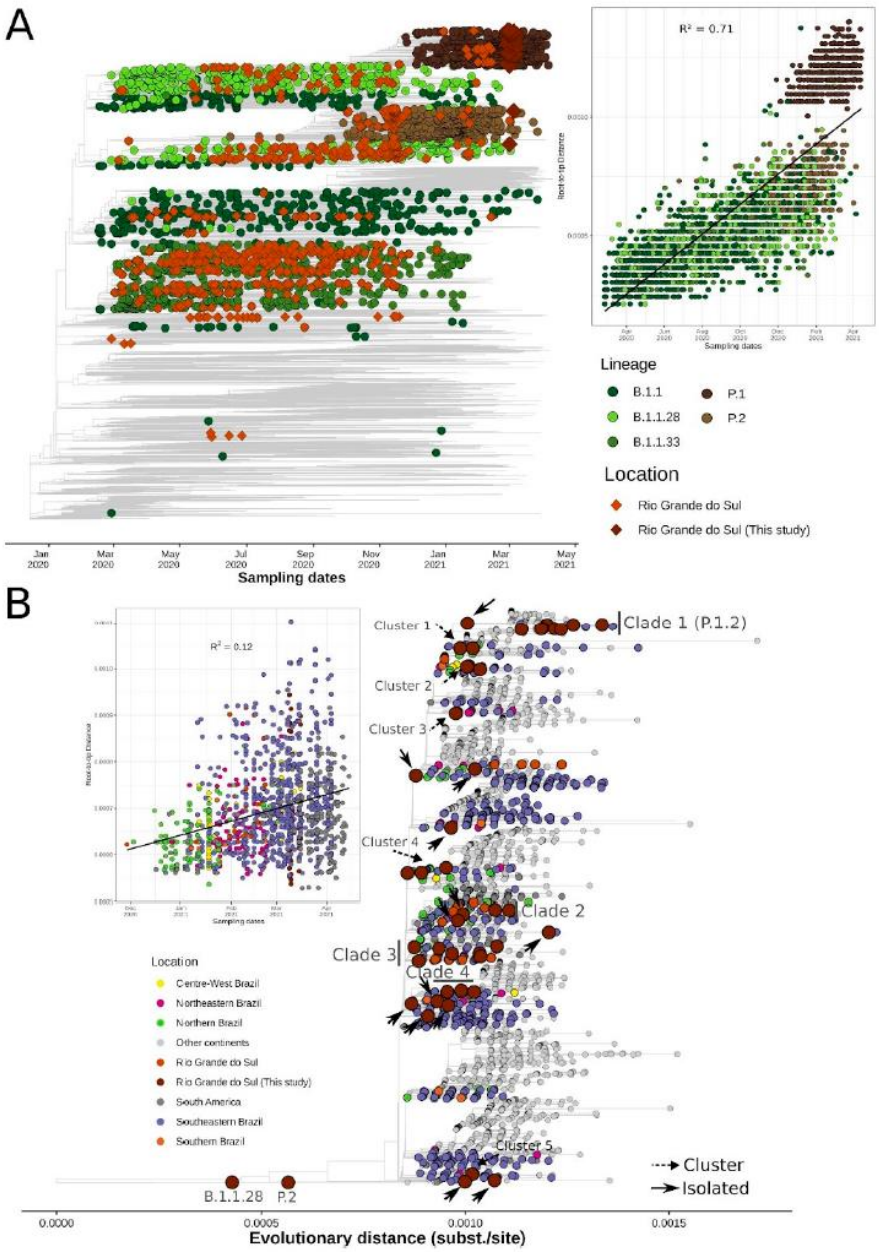


Figure 3. Phylogenetic analysis of genomes sequences in RS state in a global context. (**A**) Time-resolved ML tree of 8635 global representative SARS-CoV-2 genomes. Circles represent global sequences belonging to the five most abundant lineages in RS state that passed quality control criteria: B.1.1 (n = 32), P.1 (n = 83), P.2 (n = 83), B.1.1.28 (n = 203), and B.1.1.33 (n = 286). Diamonds represent RS genomes (available in GISAID and sequenced in this study). Root-to-tip regression is represented on the right of the tree. (**B**) ML tree of 4499 SARS-CoV-2 genomes belonging to the P.1 lineage. Tips are colored by Brazilian regions, South America, or other continents. Introductions, clusters, and clades are annotated in the tree (see Methods). Root-to-tip regression is depicted on the left of the tree and sequences from "other continents" were dropped to improve visualization.

We identified 4 clades, 5 clusters, and 13 isolated sequences (Figure 3B and Figure S7). Most importantly, Clade 1 had high branch-support (SH-aLRT = 76.3) and was composed of 11 sequences originated in this study that shared five lineage-defining mutations as previously described (Table 2) and were recently attributed to the P.1.2 sublineage (https://github.com/cov-lineages/pango-designation/issues/56; accessed on 6 May 2021). As of April 26, 2021, this sublineage is already distributed worldwide in 93 sequences (the Netherlands, Spain, England, and USA) and in other Brazilian states (Rio de Janeiro (RJ) and SP) [21]. Clade 2 sequences harbored two mutations in ORF1ab:V2862L (nsp4) and synC10507T and one in ORF3a:M260K, and it comprised 81 genomes. Four samples are from this study. The majority are from Amazonas (n = 15), São Paulo (n = 11), RS (n = 8), and Bahia (BA) (n = 4), and worldwide sequences are mainly represented by French Guiana, USA, Spain, Japan, and Jordan. Clade 3 is represented by three ORF1ab mutations (synC1420T, D1600N [nsp3], and synT8392A) in three of the seven local genomes. It is composed of sequences from RS (n = 25), SP (n = 15), Maranhão (n = 10), and RJ (n = 8), as well as other countries (mainly Spain, French Guiana, and USA). Clade 4 is characterized by two ORF1ab substitutions (G400S (nsp2) and S6822I (2′-O-ribose methyltransferase)), one N:synT26861C in three genomes, and carries other additional mutations (ORF1ab: synG10096A, G3676S (nsp6), F3677L (nsp6)) and M:synT26861C. This clade is mainly found in SP (n = 11), RS (n = 7), Santa Catarina (n = 5), BA (n = 4), and Goiás (n = 4), as well as other countries (mainly USA, Chile, and England).

Clusters 1 and 3 have, respectively, one (ORF1ab: G3676S (nsp6)) and two (ORF1ab: synC1471T and A1049V (nsp3)) shared mutations. Among all identified clusters, the most diverse was Cluster 5, which contains three samples from this study and has five defining mutations: four in ORF1ab (synT4705C, synC11095T, syn11518, and T5541I (helicase)) and one in ORF7a: E16D. Moreover, two sequences share one distinct mutation (ORF1ab: F3677L (nsp6)).

2.5. Bayesian Molecular Clock and Phylogeographic Analysis

To date the time of the most recent common ancestor (TMRCA) and the diffusion of the four P.1 clades identified in our ML analysis, we used coalescent and phylogeographic methods. For Clade 1, which is correspondent to the recently labeled P.1.2 lineage, sequences showed a moderate correlation of genetic distances and sampling dates (correlation coefficient: 0.59, $R^2 = 0.34$) (Figure 4A). We estimated a median evolutionary rate of 7.68×10^{-4} (95% highest posterior density interval [HPD]: 4.18×10^{-4} to 1.14×10^{-3} subst/site/year) and the TMRCA on 18 December 2020 (95% HPD: 29 October 2020 to 31 January 2021). Interestingly, the tree's root was placed in RS, between a sequence from RS (the oldest sequence from this clade: EPI_ISL_983865) and a subclade from USA. The divergence of these subclades was dated on 15 January 2021 (95% HPD: 15 January to 26 March 2021). The subclade composed of the RS sequences formed two separate clusters, one with three sequences from this study and one Australian genome and another composed of sequences from RS, SP, UK, Portugal, USA, and transmission clusters from RJ and Netherlands (Figure 4B). The emergence of an important cluster in RJ carrying additional mutations [21] was dated here on 11 March 2021 (95% HPD: 11 March to 6 April 2021). As American sequences formed a separate subclade, local transmission is probably occurring in the country. The divergence of the American subclade was dated 7 February 2021 (95% HPD: 1 February to 11 May 2021). In accordance with the root being placed in RS, the BSSVS procedure identified well-supported rates of diffusion from RS to other Brazilian states such as São Paulo (Bayes Factor (BF): 6.82; posterior probability (PP): 0.52), Rio de Janeiro (BF: 39.18; PP: 0.86), and other countries such as USA (BF: 31.94; PP: 0.84) and Netherlands (BF: 80.16; PP: 0.93).

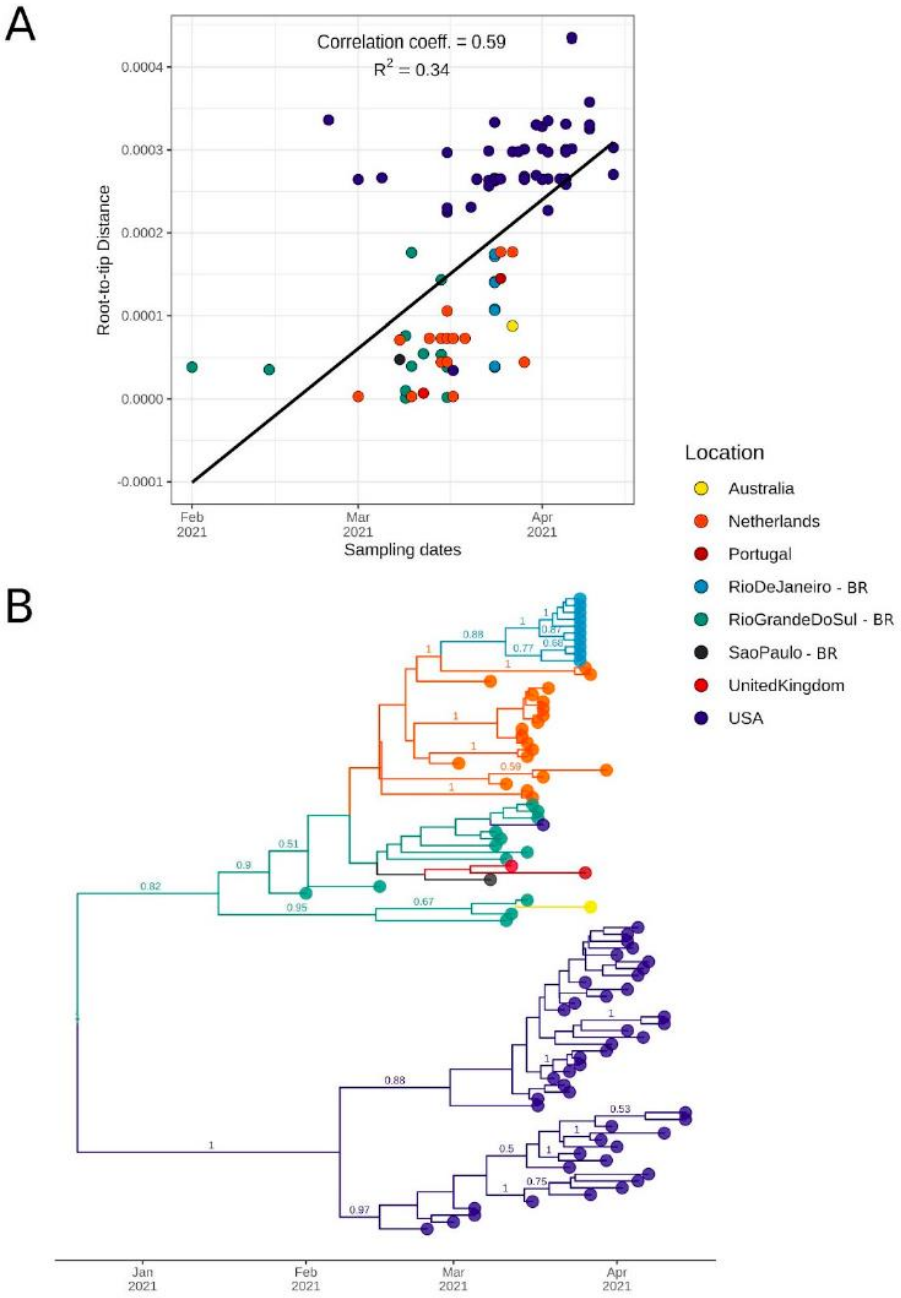


Figure 4. Bayesian discrete asymmetric phylogeographic analysis of the identified Clade 1 (lineage P.1.2). (**A**) Root-to-tip regression of genetic distances and sampling dates for Clade 1. Correlation coefficient and R squared are depicted above the graph. (**B**) MCC tree of the 93 sequences included in this analysis up to 26 April 2021 (82 from GISAID and 11 from this study). Numbers above branches represent the posterior probability of each branch. Only posteriors > 0.5 are shown. Circles indicate countries outside Brazil and Brazilian states (BR suffix).

However, it is possible that this lineage emerged in another Brazilian state, but its earlier representatives were not sampled. This is a strong hypothesis since this sequence is associated with community transmission after contact with tourists in a city of RS (Gramado) that receives numerous visitors annually [25].

For Clade 2, we estimated a median evolutionary rate of 5.85×10^{-4} (95% HPD: 4.18×10^{-4} to 7.71×10^{-4} subst/site/year), and the TMRCA was dated 30 November 2020 (95% HPD: 2 November to 21 December 2020). This clade includes sequences from 11 Brazilian states from all 5 regions and 9 other countries. We were able to detect at least five introductions from Amazonas, where this clade probably emerged. These introductions ranged from December 28, 2020 (95% HPD: 28 December 2020 to 5 January 2021) to 28 January 2021 (95% HPD: 28 January to 7 March 2021). Importantly, we identified a well-supported subclade (PP = 1) of four genomes from this study (Figure 5A).

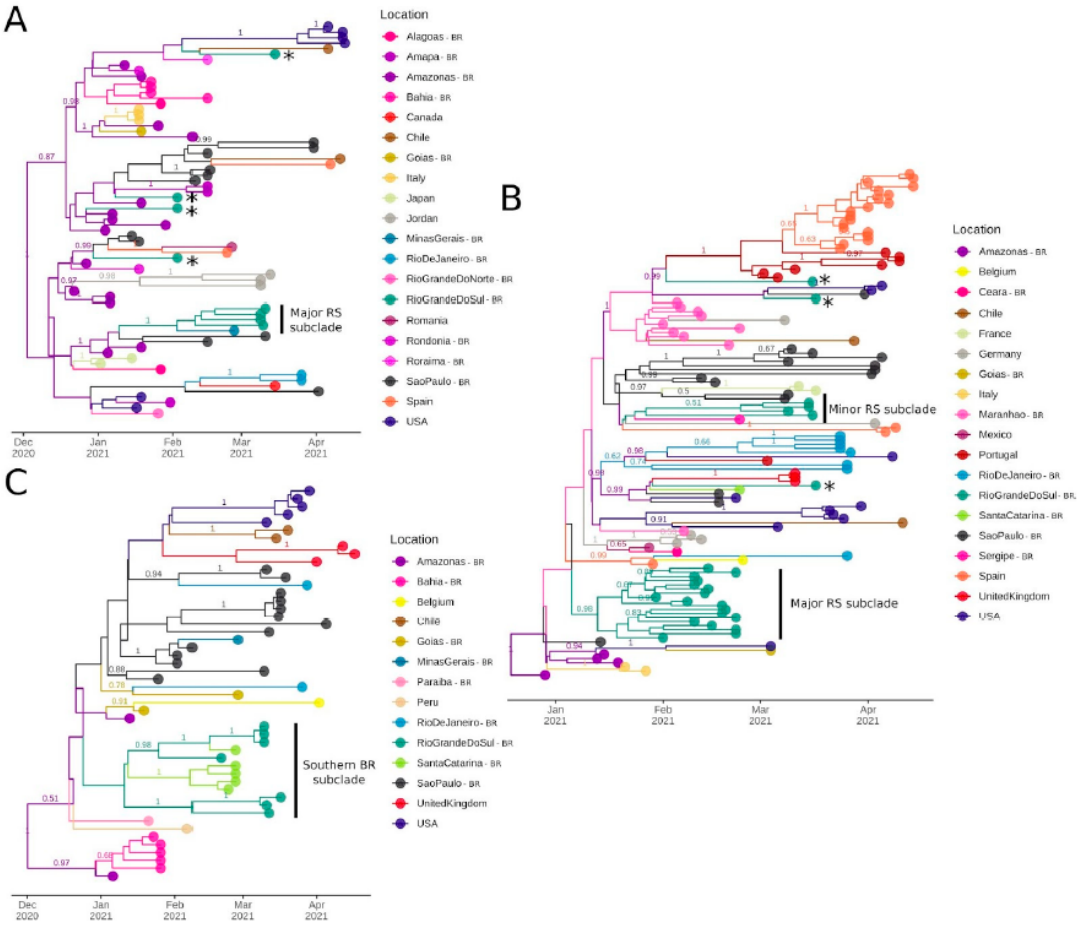


Figure 5. Bayesian discrete asymmetric phylogeographic analysis of the identified Clades 2–4 up to 26 April, 2021. (**A**) Clade 2: MCC tree of the 71 sequences included in this analysis. (**B**) Clade 3: MCC tree of the 122 genomes included in this analysis. (**C**) Clade 4: MCC tree of the 50 sequences included in this analysis. For all MCC trees, numbers above branches represent the posterior probability of each branch. Only posteriors > 0.5 are shown. Asterisks represent potential introductions in RS state and subclades cited in the text are indicated. Circles indicate countries outside Brazil and Brazilian states (BR suffix).

For Clade 3, the TMRCA was estimated on 20 December 2020 (95% HPD: November 25 to 29 December 2020) and the median evolutionary rate was 7.85×10^{-4} (95% HPD: 6.06×10^{-4} to 1.02×10^{-3} subst/site/year). This clade harbors sequences from 9 Brazilian states and 10 other countries. Amazonas is the most probable source of its emergence. From then onwards, multiple transmission clusters were established in foreign countries (e.g., Spain, Portugal, and USA) and Brazilian states (especially Maranhão, SP, and RS). This clade was introduced at least 5 times in RS, leading to 2 major subclades represented by 18 and 4 sequences, respectively. The major subclade (n = 18, PP = 0.98) was dated 11 January 2021 (95% HPD: 11 January to 1 February 2021) (Figure 5B).

For Clade 4, the TMRCA was dated on 2 December 2020 (95% HPD: 7 October 2020 to 3 January 2021), and the median evolutionary rate was 6.26×10^{-4} (95% HPD: 3.51×10^{-4} to 1.01×10^{-3}). This clade comprises nine Brazilian states and five foreign countries. After its initial emergence and spread in Amazonas, it had already formed transmission clusters in SP, BA, United Kingdom, and USA. Most importantly, a subclade containing sequences from two neighboring states from Southern Brazil (seven from RS and five from Santa Catarina (SC)) indicates its diffusion from RS to SC, probably leading to two separate introductions. The divergence of this subclade was estimated on 16 December 2020 (95% HPD: 16 December 2020 to 19 January 2021) (Figure 5C).

Phylogenetic and molecular clock approaches suggest the wide circulation of the VOC P.1 both nationally and internationally between late 2020 and early 2021. This lineage has already diversified into some clades that bear characteristic mutations, although they exhibit similar evolutionary rates. We inferred that P.1 (and its derived clades) was introduced multiple times in the southernmost Brazilian state (RS) between mid-December 2020 and January 2021. Remarkably, this date is close to the first P.1 detection in Manaus, which is located ~4000 km away. These early introductions led to the formation of local subclades that could be identified even using a reduced set of sequenced samples.

3. Discussion

In this study, the analysis of 56 samples from the state of Rio Grande do Sul (RS), Southern Brazil, confirmed that the P.1 lineage was already highly prevalent. Interestingly, we demonstrated that P.1 is already showing signs of diversification and has originated a new sublineage (P.1.2). Herein, we indicate the likely origin and the first clusters of this novel lineage. This sublineage was detected in three Brazilian states, and other countries, and its most recent common ancestor was dated on mid-December, 2020 (95% HPD: 29 October 2020 to 31 January 2021). In accordance with the majority of the states from Brazil, this state experienced significant increases in hospitalizations in early 2021. This scenario was related to the emergence and rapid spread of the P.1 variant across the country.

After almost one year of relatively slow SARS-CoV-2 evolution, the emergence of multiple and convergent lineages harboring a constellation of mutations in the spike protein raised concern in the scientific community. This protein is responsible for mediating interaction with the human Angiotensin-Converting Enzyme 2 receptor (hACE2) and is a primary target of neutralizing antibodies and vaccines [28]. The variants harboring different mutational signatures, including spike protein substitutions, were classified as VOCs and "variants of interest" (VOIs), depending on their growing relevance in the current pandemic. The first three VOCs emerged in England (B.1.1.7) [8], South Africa (B.1.351) [9], and Brazil (P.1) [10]. More recently, B.1.617.2 (India) [12,13] also was characterized as a VOC. By May 2020, B.1.427/429 [29], B.1.526 (New York, USA), B.1.617 (India), and P.2 (Brazil) [22] were categorized as VOIs. B.1.1.7, B.1.351, and P.1, the most studied VOCs, have the D614G and N501Y mutations in common. B.1.351 and P.1 share a mutation in the K417 site (K417N and K417T, respectively) and the E484K replacement, which is also observed in the P.2 lineage. Additionally, B.1.1.7 carries the P681H substitution in the furin-cleavage site and multiple VOIs bear the L452R substitution [7]. The presence of common substitutions in different SARS-CoV-2 lineages suggests co-evolutionary and convergent mutational processes [8–10,30].

In the present study, we noticed that B.1.1.33 and B.1.1.28 lineages, detected at the beginning of the pandemic in Brazil [2], had been similarly prevalent in different regions until September 2020, before the appearance of P.2 (in October) and P.1 (in December 2020). The B.1.1.33 lineage shows variable abundance in different Brazilian states (ranging from 2% in Pernambuco to 80% in Rio de Janeiro), with moderate prevalence in South American countries (5–18%). Surprisingly, this lineage was firstly detected in early March 2020 in other American countries (e.g., Argentina and USA). Apparently, an intermediate strain probably emerged in Europe and subsequently spread to Brazil, where its spread gave rise to B.1.1.33 [31] and possibly triggered secondary outbreaks in Argentina and Uruguay [31,32]. We found that N.3 and N.5, both derived from B.1.1.33, represented an important proportion of the sequences from Argentina from May to December 2020, when it was replaced by the P.2 lineage, which probably emerged in Rio de Janeiro (Southeastern Brazil). The B.1.1.28 lineage, despite apparently being less abundant than B.1.1.33 in several Brazilian regions, quickly diversified into two variants: VOC P.1 and VOI P.2 [33]. Since the end of 2020, these two lineages have lead the diversity of SARS-CoV-2 in Brazil [17] and have caused concern in other countries after several introductions. Regarding the distribution of sequenced samples across RS state, the cumulative frequency of B.1.1.28 and B.1.1.33 was higher until mid-April 2021 [16]. However, since the end of 2020 and beginning of 2021, a rise in P.1 and P.2 sequences was observed. Our study supported that P.1 outperformed other lineages in RS as of March 2021, although the collection of samples in hospitalized patients and low geographic representativeness does not allow the extrapolation of these findings.

The emergence of a B.1.1.28 derived lineage carrying the S:E484K mutation (P.2) was dated, in a retrospective study, late February 2020 in the Southeast (São Paulo and Rio de Janeiro), followed by transmission to the South (especially RS). Since then, multiple dispersion routes were observed between Brazilian states, especially in late 2020 and early 2021 [18]. However, this lineage went unreported until October 2020, when it was first detected in the state of Rio de Janeiro [22] and in the small municipality of Esteio in RS [23]. The increased frequency of B.1.1.28 and derived lineages was corroborated by another study that included samples from several municipalities of RS in November 2020. This study found that 86% of the genomes could be classified as B.1.1.28 and ~50% of these, in fact, belong to the new lineage P.2 [24]. Nonetheless, our current study suggests that P.2 has already been nearly entirely replaced by the P.1 lineage or is not particularly well represented among the analyzed patients seeking emergency consultation or requiring hospitalization.

Between June and October 2020, an extremely high seroprevalence (44–76%) was observed in Manaus (Amazonas, Brazil) in a study from blood donors [14]. However, despite these numbers, Manaus faced a resurgence of cases and a six-fold increase in hospitalizations between December 2020 and January 2021. The most plausible hypotheses that would justify this condition are: (i) the previous overestimation of seroprevalence in Manaus; (ii) the immune evasion property of some SARS-CoV-2 mutations found in VOCs; (iii) higher transmissibility and pathogenicity of SARS-CoV-2 lineages circulating in the second wave compared with pre-existing lineages [15].

A genomic epidemiology study that used 250 SARS-CoV-2 genomes from 25 different municipalities from Amazonas sampled between March 2020 and January 2021 shows that the first exponential phase in the state was driven mainly by the spread of lineage B.1.195, which was gradually replaced by B.1.1.28. The second wave coincided with the emergence of P.1 in November, which rapidly replaced the parental lineage (<2 months) [11] and whose emergence was preceded by a period of rapid molecular evolution [10]. Importantly, rapid accumulation of mutations over short timeframes have been reported in chronically infected or immunocompromised hosts [34,35]. However, preliminary findings pointed to the existence of P.1 intermediate lineages, suggesting that the constellation of mutations defining P.1 were acquired at sequential steps during multiple rounds of infections instead of within a single long-term infected individual [36]. The VOC P.1 carries three deletions,

four synonymous substitutions, a four base-pair nucleotide insertion, and at least 17 other lineage-defining replacements, including 10 missense mutations in the spike protein (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, and T1027I), 8 of which are subjected to positive selection [10].

Regarding infectiousness, transmissibility, and case fatality, the viral load was ~10-fold higher in P.1 infections than in non-P.1 infections [11]. Although another study points to uncertainties regarding viral load and duration of infection after accounting for confounding effects [10]. Moreover, it was estimated to be 1.7–2.4-fold more transmissible, raising the probability that reinfections would be caused more frequently in hosts infected by P.1 rather than by older lineages. Remarkably, infections were 1.2–1.9 times more likely to result in death in the period following the emergence of P.1 compared to previous time frames [10]. These findings support that successive lineage replacements in Amazonas were driven by a complex combination of factors, including the emergence of the more transmissible VOC P.1 virus [11].

A study conducted in RS described a P.1 lineage infection on 30 November 2020 followed by a P.2 lineage reinfection on 11 March 2021 in a patient with comorbidities. This report was the first detected P.1 in the state [26]. Other analyses suggest that the P.1 lineage presumably emerged in Manaus, Brazil, in mid-November 2020 [10,11]. Therefore, the P.1 lineage was present in Southern Brazil about a few days after its emergence in Manaus, Northern Brazil. Our molecular clock analysis supported this scenario. Another study, once thought to be the first P.1 report in RS, documented local transmission of P.1 from a person who had close contact with tourists and was positive for COVID-19 in early February 2021 [25]. This happened in the city of Gramado, a town in the mountains that receives around 6.5 million tourists every year and belongs to the Caxias do Sul intermediate region. Interestingly, this sample from Gramado was the earliest representative of a new P.1-derived lineage (P.1.2), described in 11 patients from our study and found in transmission clusters from the RJ state in Southeastern Brazil, USA, and the Netherlands. Remarkably, our local sequences are more similar to genomes from other countries compared to the RJ cluster, which acquired at least four additional mutations (including S:A262S) [21].

Whether P.1.2 has worse clinical outcomes than its prior variant (P.1) is unknown. However, as described above, the missense mutations characteristic of the new sublineage are located at nsp2 and nsp3 (ORF1ab), ORF3a, and nucleocapsid. These sites are known for their interaction with human proteome, potentially influencing the immunological and inflammatory response against SARS-CoV-2 infection [37]. The ORF3a:D155Y substitution is located near SARS-CoV caveolin-binding Domain IV. The binding interaction of viral ORF3a protein to host caveolin-1 is essential for entry and endomembrane trafficking of SARS-CoV-2. Since this mutation breaks the salt bridge formation between Asp155 and Arg134, it can interfere with the binding affinity of ORF3a to host caveolin-1 and change virulence properties. Most importantly, this disrupted interaction may be associated with improved viral fitness, since it can avoid the induction of host cell apoptosis or extend the asymptomatic phase of infection [38]. We hypothesize that these new substitutions could, therefore, influence epidemiological and clinical outcomes favoring P.1.2 evolution. This is elusive at best at this time, however, and further sublineage characterization is needed to further explore its real relevance.

Some limitations should be considered. Firstly, the sample size is low and not necessarily representative of RS state. Considering the number of sequences from each intermediate region in RS available in GISAID, it is very likely that the distribution seen on the map (Figure 2C) is a consequence of sampling at different times in these localities or simple randomness. Thus, it should not be assumed as a true representation of the spatial diversity in the state. Since publicly available genomes are a result of episodic sequencing efforts, especially in Brazil, more precise inferences about introductions and diffusion processes in regional and worldwide contexts are restricted due to the lack of proper geographical and temporal distribution of the samples. Therefore, more research and surveillance are

essential to unravel a more precise genomic characterization of SARS-CoV-2 in Brazil, promptly identifying novel variants to better respond and control its spread.

In summary, our study corroborates the total virtual substitution of previous lineages by P.1 in Southern Brazil in COVID-19 cases sequenced in March 2020. Moreover, we confirmed various cases caused by the novel P.1.2 sublineage and placed its origin in the state of Rio Grande do Sul. The continuous evolution of the VOC P.1 is concerning, considering its clinical and epidemiological impact, and warrants enhanced genomic surveillance.

4. Materials and Methods

4.1. Sample Collection and Clinical Testing

Samples were obtained from Hospital da Brigada Militar patients, both admitted or visiting the emergency ward, from Porto Alegre, RS, Brazil. Nasopharyngeal swabs were collected and placed in saline solution. Samples were transported to the clinical laboratory (Laboratório Exame) and tested on the same day for SARS-CoV-2 using Real-Time Reverse-transcriptase Polymerase Chain Reaction (Charité RT-qPCR assays). The RTq-PCR assay used primers and probes recommended by the World Health Organization targeting the nucleocapsid (N1 and N2) genes [39]. Remnant samples were stored at −20 °C.

Between 9 March and 17 March 2021, all routinely tested samples of the clinical laboratory provenient of the Hospital da Brigada Militar patients and yielded positive RT-qPCR were selected. Subsequently, those positive clinical samples were submitted to a second RT-qPCR performed by BiomeHub (Florianópolis, Santa Catarina, Brazil), using the same protocol (charite-berlin). Only samples with quantification cycle (Cq) below 30 for at least one primer were submitted to the SARS-CoV-2 genome sequencing. In total, 56 patients who presented symptoms such as fever, cough, sore throat, dyspnea, anosmia, fatigue, diarrhea, and vomiting (moderate and severe clinical status) [40] were included in the study.

4.2. RNA Extraction, Library Preparation, and Sequencing

Total RNAs were prepared as in the reference protocol [41] using SuperScript IV (Invitrogen, Carlsbad, CA, USA) for cDNA synthesis and Platinum Taq High Fidelity (Invitrogen, Carlsbad, CA, USA) for specific viral amplicons. Subsequently, cDNA was used for the library preparation with Nextera Flex (Illumina, San Diego, CA, USA) and quantified with Picogreen and Collibri Library Quantification Kit (Invitrogen, Carlsbad, CA, USA). The sequencing was performed on the Illumina MiSeq (Illumina, San Diego, CA, USA) 150 × 150 runs with 500xSARS-CoV-2 coverage (50–100 thousand reads/sample).

4.3. Quality Control and Consensus Calling

Quality control, reference mapping, and consensus calling were performed using an in-house pipeline developed by BiomeHub (Florianópolis, Santa Catarina, Brazil). Briefly, adapters were removed, and reads were trimmed by size = 150. Reads were mapped to the reference SARS-CoV-2 genome (GenBank accession number NC_045512.2) using Bowtie v2.4.2 (end-to-end and very-sensitive parameters) [42]. Mapping coverage and depth were retrieved using samtools v1.11 [43] (minimum base quality per base (Q) $\geq$ 30). Consensus sequences were generated using bcftools mpileup (Q $\geq$ 30; depth (d) $\leq$ 1000) combined with bcftools filter (DP > 50) and bcftools consensus v1.11 [44]. Coverage values for each genome were plotted using the karyoploteR v1.12.4 R package [45]. Finally, we assessed the consensus sequences quality using Nextclade v0.14.2 (https://clades.nextstrain.org/; accessed on 4 May 2021).

4.4. Mutation Analysis

SNPs and insertions/deletions in each sample were identified using snippy variant calling and core genome alignment pipeline v4.6.0 (https://github.com/tseemann/snippy; accessed on 4 May 2021), which uses FreeBayes v1.3.2 [46] to call variants and snpEff v5.0 [47] to annotate and predict their effects on genes and proteins. Genome map and

SNP histogram were generated after running MAFFT v7.475 [48] alignment using the msastats.py script, and plotAlignment and plotSNPHist functions [49]. Sequence positions refer to GenBank RefSeq sequence (NC_045512.2), isolated and sequenced from an early case from Wuhan (China) in 2019.

We identified global virus lineages using the dynamic nomenclature implemented in Pangolin v2.3.8 [50] (https://github.com/cov-lineages/pangolin; accessed on 4 May 2021) and global clades and mutations using Nextclade v0.14.2 (https://clades.nextstrain.org/; accessed on 4 May 2021). We also used Pathogenwatch (https://pathogen.watch/; accessed on 4 May 2021) and Microreact [51] to explore mutations and lineages across time and geography initially.

4.5. Maximum Likelihood Phylogenomic Analysis

All available SARS-CoV-2 genomes (1,048,519 sequences) were obtained from GISAID on April 26, 2021 and combined with our 56 sequences to obtain a global representative dataset. These sequences were subjected to analysis inside the NextStrain ncov pipeline [52] (https://github.com/nextstrain/ncov; accessed on 4 May 2021). In this workflow, sequences were aligned using Nextalign v0.1.6 (https://github.com/neherlab/nextalign; accessed on 4 May 2021). In the initial filtering step, short and low-quality sequences or those with incomplete sampling dates were excluded. Uninformative sites and ends (100 positions in the beginning and 50 in the end) were also masked from the alignment. Genetically closely related genomes to our focal subset were selected, prioritizing sequences geographically closer to Brazil's RS state. The maximum likelihood (ML) phylogenetic tree was built using IQ-TREE v2.1.2 [53], employing the general time-reversible (GTR) model with unequal rates and base frequencies [54]. The tree's root was placed between lineage A and B (Wuhan/WH01/2019 and Wuhan/Hu-1/2019 representatives), and sequences that deviate more than four interquartile ranges from the root-to-tip regression of genetic distances against sampling dates were removed from the analysis. A time-scaled ML tree was generated with TreeTime v0.8.1 [55] under a strict clock and a skyline coalescent prior with a mean rate of 8×10^{-4} substitutions per site per year. Finally, clades and mutations were assigned and geographic movements inferred. The results were exported to JSON format to enable interactive visualization through Auspice.

Additionally, as P.1 sequences mostly represent our dataset, we downloaded all complete and high-quality global genomes assigned to P.1 PANGO lineage (4499 sequences) submitted until 26 April 2021. These sequences were aligned using MAFFT v7.475, the ends of the alignment (300 in the beginning and 500 in the end) were masked, and the ML tree was built with IQ-TREE v2.0.3 using the GTR + F + R3 nucleotide substitution model as selected by the ModelFinder [56]. Branch support was calculated using the Shimodaira–Hasegawa approximate likelihood ratio test (SH-aLRT) [57] with 1000 replicates.

Local sequences were classified according to the following scheme: monophyletic clades composed by one local genome were classified as "isolated", while clades composed by 2 < genomes < 4 were considered "clusters", and, if $\geq$ 4 local genomes were represented, we assigned a "clade" designation.

ML trees were inspected in TempEst v1.5.3 [58] to investigate the temporal signal through regression of root-to-tip genetic divergence against sampling dates. For the P.1 ML tree, samples with missing days of the collection were filled with the 15th day of the month. ML and time-resolved trees were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/; accessed on 4 May 2021) and ggtree R package v2.0.4 [59].

4.6. Discrete Bayesian Phylogeographic and Phylodynamic Analysis

Considering the four identified clades composed of $\geq$4 sequences from this study, we extracted the clade members using the caper R package v1.0.1 [60]. Clade-specific ML trees and root-to-tip regression assignments were generated as described above. Evolutionary parameter estimates and spatial diffusion were assessed separately for each

clade using a Bayesian Markov Chain Monte-Carlo (MCMC) approach implemented in BEAST v10.4 [61]. The BEAGLE library [62] was used to enhance computational time. Time-stamped Bayesian trees were generated using the HKY + Γ nucleotide model [63], a strict molecular clock model with a Continuous Time Markov Chain (CTMC) rate reference prior [64] (mean rate = 8×10^{-4}) and a non-parametric skygrid tree prior [65] with grid points defined by the approximate number of weeks spanned by the duration of the phylogeny.

The MCMC chains were run in duplicates for at least 50 million generations, and convergence was checked using Tracer v1.7.1 [66]. Log and tree files were combined using LogCombiner v1.10.4 to ensure stationarity and good mixing after removing 10% as burn-in. Maximum clade credibility (MCC) was generated using TreeAnnotator v1.10.4 [61]. Viral migrations were reconstructed using a reversible discrete asymmetric phylogeographic model [67] to infer the locations of the internal nodes of the tree. A discretization scheme with a resolution of different Brazilian states and other countries was applied. Location diffusion rates were estimated using the Bayesian stochastic search variable selection (BSSVS) [67] procedure employing Bayes factors to identify well-supported rates.

4.7. Geoplotting

Geographical maps and general plots were generated using R v3.6.1 [68], and the ggplot2 v3.3.2 [69], geobr v.1.4 [70], and sf v0.9.8 [71] packages. For the discrete phylogeographic analysis, SpreaD3 v0.9.7.1 software [72] was used to map spatiotemporal information embedded in MCC trees.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/pathogens10080988/s1. Figure S1. Spatiotemporal distribution of the 56 sequenced samples from RS. Figure S2. Neighbouring countries, Brazilian divisions and RS intermediate regions. Figure S3. Proportion of the 10 most frequent lineages of SARS-CoV-2 across time in four Brazilian regions. Figure S4. ML tree augmenting the B.1.1.28 clade where the local sequence (RS-HBM-39491) was placed. Figure S5. ML tree augmenting the P.2 clade where the local sequence (RS-HBM-39486) was placed. Figure S6. Coverage depth plots for each genome sequenced. Figure S7. Expanded ML tree built using 4499 P.1 sequences deposited in GISAID until 26 April 2021. Tips represented by sequences sequenced in this study are augmented to enable visualization of the different introductions, clusters, and clades reported. Table S1. Collection of all mutations (n = 175) found in the 56 sequenced genomes, including effects on SARS-CoV-2 genes and proteins. The table is ordered by genomic position. Table S2. GISAID acknowledgment table of the 8635 global sequences used in the Nextstrain workflow. Table S3. GISAID acknowledgment table of the 4499 P.1 sequences analyzed. Supplementary File S1. Data and code used to reproduce the results presented.

Author Contributions: Conceptualization, R.A.Z. and C.E.T.; Data curation, V.B.F., C.P. (Christiano Perin), A.H. and C.P. (Camila Peter); Formal analysis, V.B.F., G.D.C., G.B.C. and C.E.T.; Investigation, V.B.F., G.D.C., C.P. (Christiano Perin), A.H., G.B.C., R.A.Z. and C.E.T.; Methodology, V.B.F., G.D.C., C.P. (Camila Peter), G.B.C., R.A.Z. and C.E.T.; Project administration, R.A.Z. and C.E.T.; Resources, C.P. (Christiano Perin), A.H., C.P. (Camila Peter), L.N.R., F.A.C., R.A.Z. and C.E.T.; Software, V.B.F.; Supervision, G.B.C. and C.E.T.; Validation, V.B.F., G.D.C. and G.B.C.; Visualization, V.B.F. and G.D.C.; Writing—original draft, V.B.F., G.D.C., R.A.Z. and C.E.T.; Writing—review and editing, V.B.F., G.D.C., C.P. (Christiano Perin), A.H., G.B.C., P.A.G.F., L.N.R., F.A.C., R.A.Z. and C.E.T. All authors have read and agreed to the published version of the manuscript.

Funding: The sequencing was supported by donations from Beppler & Puppi Advogados, Smellbox Produtos de Higiene Ltd.a., and Dr. Leonardo Mestre Negri. Scholarships and Fellowships were supplied by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001 and Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA).

Institutional Review Board Statement: Ethical approval was obtained from the Comitê de Ética em Pesquisa em Seres Humanos da Universidade Federal de Ciências da Saúde de Porto Alegre (CEP-UFCSPA) under process number CAAE 35083220.2.0000.5345. The study was performed in accordance with the Declaration of Helsinki. All samples belonging to the Hospital da Brigada Militar patients that yielded positive RT-qPCR had their laboratory electronic records reviewed to compile metadata such as date of collection, sex, age, symptoms, exposure history, and clinical status, when available. Samples were anonymized before being received by the study investigators, following Brazilian and international ethical standards.

Informed Consent Statement: This study obtained a waiver of informed consent approved by Comitê de Ética em Pesquisa em Seres Humanos da Universidade Federal de Ciências da Saúde de Porto Alegre (CEP-UFCSPA) under process number CAAE 35083220.2.0000.5345.

Data Availability Statement: Full tables acknowledging the authors and corresponding labs submitting sequencing data used in this study can be found in Files S3 and S4. Consensus genomes generated in this study were deposited in the GISAID database under Accession IDs: EPI_ISL_2139494 to EPI_ISL_2139549. Data and code used to reproduce the results presented are available in Supplementary Materials.

Acknowledgments: We thank the administrators of the GISAID database and research groups across the world for supporting the rapid and transparent sharing of genomic data during the COVID-19 pandemic. We also thank the staff of Hospital da Brigada Militar, Laboratório Exame, Beppler & Puppi Advogados, Smellbox Produtos de Higiene Ltd.a., Leonardo Mestre Negri, and BiomeHub Pesquisa e Desenvolvimento who directly contributed to this study.

Conflicts of Interest: Dr. Cadegiani has served as a clinical director for Applied Biology, Inc. The other authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. World Health Organization. WHO Director-General's Opening Remarks at the Media Briefing on COVID-19—11 March 2020. Available online: 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 on 27 May 2021).
2. Candido, D.; Claro, I.M.; de Jesus, J.G.; Souza, W.M.; Moreira, F.R.R.; Dellicour, S.; Mellan, T.A.; du Plessis, L.; Pereira, R.H.M.; Sales, F.C.S.; et al. Evolution and Epidemic Spread of SARS-CoV-2 in Brazil. *Science* **2020**, *369*, 1255–1260. [CrossRef] [PubMed]
3. IBGE (Brazilian Institute of Geography and Statistics) Rio Grande do Sul—Cidades e Estados. Available online: https://www.ibge.gov.br/cidades-e-estados/rs.html (accessed on 17 May 2021).
4. IBGE (Brazilian Institute of Geography and Statistics) Regiões Geográficas. Available online: https://www.ibge.gov.br/apps/regioes_geograficas/ (accessed on 17 May 2021).
5. Rio Grande do Sul Department of Health—SES-RS Confirmado o Primeiro Caso de Novo Coronavírus no Rio Grande do Sul. Available online: https://saude.rs.gov.br/confirmado-o-primeiro-caso-de-novo-coronavirus-no-rio-grande-do-sul (accessed on 24 November 2020).
6. Secretaria de Planejamento, Governança e Gestão—Governo do Estado do Rio Grande do Sul Cogestão Regional—Distanciamento Controlado. Available online: https://distanciamentocontrolado.rs.gov.br/ (accessed on 17 May 2021).
7. Mullen, J.L.; Tsueng, G.; Latif, A.A.; Alkuzweny, M.; Cano, M.; Haag, E.; Zhou, J.; Zeller, M.; Matteson, N.; Andersen, K.G.; et al. Outbreak.Info. Available online: https://outbreak.info (accessed on 19 July 2021).
8. Rambaut, A.; Loman, N.; Pybus, O.; Barclay, W.; Barrett, J.; Carabelli, A.; Connor, T.; Peacock, T.; Robertson, D.; Volz, E.; et al. Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations. Available online: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed on 4 January 2021).
9. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Detection of a SARS-CoV-2 Variant of Concern in South Africa. *Nature* **2021**, *592*, 438–443. [CrossRef]
10. Faria, N.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.D.S.; Mishra, S.; Crispim, M.A.E.; Sales, F.C.S.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [CrossRef]
11. Naveca, F.; Nascimento, V.; Souza, V.; Corado, A.; Nascimento, F.; Silva, G.; Costa, Á.; Duarte, D.; Pessoa, K.; Mejía, M.; et al. COVID-19 Epidemic in the Brazilian State of Amazonas Was Driven by Long-Term Persistence of Endemic SARS-CoV-2 Lineages and the Recent Emergence of the New Variant of Concern P.1. Available online: https://www.researchsquare.com/article/rs-275494/v1 (accessed on 1 March 2021).

12. Dhar, M.S.; Marwal, R.; Radhakrishnan, V.S.; Ponnusamy, K.; Jolly, B.; Bhoyar, R.C.; Sardana, V.; Naushin, S.; Rophina, M.; Mellan, T.A.; et al. Genomic Characterization and Epidemiology of an Emerging SARS-CoV-2 Variant in Delhi, India. *medRxiv* **2021**, *6*, 21258076. [CrossRef]
13. Peacock, T.P.; Sheppard, C.M.; Brown, J.C.; Goonawardane, N.; Zhou, J.; Whiteley, M.; Consortium, P.V.; de Silva, T.I.; Barclay, W.S. The SARS-CoV-2 Variants Associated with Infections in India, B.1.617, Show Enhanced Spike Cleavage by Furin. *bioRxiv* **2021**, *5*, 446163. [CrossRef]
14. Buss, L.F.; Prete, C.A.; Abrahim, C.M.M.; Mendrone, A.; Salomon, T.; de Almeida-Neto, C.; França, R.F.O.; Belotti, M.C.; Carvalho, M.P.S.S.; Costa, A.G.; et al. Three-Quarters Attack Rate of SARS-CoV-2 in the Brazilian Amazon during a Largely Unmitigated Epidemic. *Science* **2021**, *371*, 288–292. [CrossRef]
15. Sabino, E.C.; Buss, L.F.; Carvalho, M.P.S.; Prete, C.A.; Crispim, M.A.E.; Fraiji, N.A.; Pereira, R.H.M.; Parag, K.V.; da Silva Peixoto, P.; Kraemer, M.U.; et al. Resurgence of COVID-19 in Manaus, Brazil, despite High Seroprevalence. *Lancet* **2021**, *397*, 452–455. [CrossRef]
16. Brazilian Ministry of Health Painel Coronavírus Brasil. Available online: https://covid.saude.gov.br/ (accessed on 17 May 2021).
17. Franceschi, V.B.; Ferrareze, P.A.G.; Zimerman, R.A.; Cybis, G.B.; Thompson, C.E. Mutation Hotspots, Geographical and Temporal Distribution of SARS-CoV-2 Lineages in Brazil, February 2020 to February 2021: Insights and Limitations from Uneven Sequencing Efforts. *medRxiv* **2021**, *8*, 21253152. [CrossRef]
18. Lamarca, A.P.; de Almeida, L.G.P.; Francisco, R. da S.; Lima, L.F.A.; Scortecci, K.C.; Perez, V.P.; Brustolini, O.J.; Sousa, E.S.S.; Secco, D.A.; Santos, A.M.G.; et al. Genomic Surveillance of SARS-CoV-2 Tracks Early Interstate Transmission of P.1 Lineage and Diversification within P.2 Clade in Brazil. *medRxiv* **2021**, *3*, 21253418. [CrossRef]
19. Rio Grande do Sul Health Surveillance Center Genomic Bulletin 5 (16/04/2021). Available online: https://coronavirus.rs.gov.br/upload/arquivos/202104/16173629-vigilancia-genomica-rs-boletim05-compactado.pdf (accessed on 20 April 2021).
20. Shu, Y.; McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data—From Vision to Reality. *Eurosurveillance* **2017**, *22*, 30494. [CrossRef] [PubMed]
21. De Almeira, L.G.; Lamarca, A.P.; Francisco Junior, R.d.S.; Cavalcante, L.; Gerber, A.L.; Guimarães, A.P.d.C.; Machado, D.T.; Alves, C.; Mariani, D.; Cruz, T.F.; et al. Genomic Surveillance of SARS-CoV-2 in the State of Rio de Janeiro, Brazil: Technical Briefing—SARS-CoV-2 Coronavirus/NCoV-2019 Genomic Epidemiology. Available online: https://virological.org/t/genomic-surveillance-of-sars-cov-2-in-the-state-of-rio-de-janeiro-brazil-technical-briefing/683 (accessed on 4 May 2021).
22. Voloch, C.M.; Francisco, R.D.S.; de Almeida, L.G.P.; Cardoso, C.C.; Brustolini, O.J.; Gerber, A.L.; Guimarães, A.P.D.C.; Mariani, D.; da Costa, R.M.; Ferreira, O.C.; et al. Genomic Characterization of a Novel SARS-CoV-2 Lineage from Rio de Janeiro, Brazil. *J. Virol.* **2021**, *95*. [CrossRef] [PubMed]
23. 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.; et al. Genomic Epidemiology of SARS-CoV-2 in Esteio, Rio Grande Do Sul, Brazil. *BMC Genom.* **2021**, *22*, 371. [CrossRef]
24. Francisco, R.D.S., Jr.; Benites, L.F.; Lamarca, A.P.; de Almeida, L.G.P.; Hansen, A.W.; Gularte, J.S.; Demoliner, M.; Gerber, A.L.; Guimarães, A.P.D.C.; Antunes, A.K.E.; et al. Pervasive Transmission of E484K and Emergence of VUI-NP13L with Evidence of SARS-CoV-2 Co-Infection Events by Two Different Lineages in Rio Grande Do Sul, Brazil. *Virus Res.* **2021**, *296*, 198345. [CrossRef] [PubMed]
25. Salvato, R.S.; Gregianini, T.S.; Campos, A.A.S.; Crescente, L.V.; Vallandro, M.J.; Ranieri, T.M.S.; Vizeu, S.; Martins, L.G.; da Silva, E.V.; Pedroso, E.R.; et al. Epidemiological Investigation Reveals Local Transmission of SARS-CoV-2 Lineage P.1 in Southern Brazil. *Rev. Epidemiol. E Controle Infecção* **2021**, *1*, 1–6. [CrossRef]
26. Soares da Silva, M.; Demoliner, M.; Hansen, A.; Gularte, J.; Silveira, F.; Heldt, F.; Filippi, M.; Pereira da Silva, F.; Mallmann, L.; Fink, P.; et al. Early Detection of SARS-CoV-2 P.1 Variant in Southern Brazil and Reinfection of the Same Patient by P.2. Available online: https://www.researchsquare.com (accessed on 14 May 2021).
27. Kubik, S.; Marques, A.C.; Xing, X.; Silvery, J.; Bertelli, C.; De Maio, F.; Pournaras, S.; Burr, T.; Duffourd, Y.; Siemens, H.; et al. Recommendations for Accurate Genotyping of SARS-CoV-2 Using Amplicon-Based Sequencing of Clinical Samples. *Clin. Microbiol. Infect.* **2021**, *27*, 1036. [CrossRef] [PubMed]
28. Walls, A.C.; Park, Y.-J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292. [CrossRef] [PubMed]
29. Deng, X.; Garcia-Knight, M.A.; Khalid, M.M.; Servellita, V.; Wang, C.; Morris, M.K.; Sotomayor-González, A.; Glasner, D.R.; Reyes, K.R.; Gliwa, A.S.; et al. Transmission, Infectivity, and Neutralization of a Spike L452R SARS-CoV-2 Variant. *Cell* **2021**, *184*, 3426–3437. [CrossRef]
30. Martin, D.P.; Weaver, S.; Tegally, H.; San, E.J.; Shank, S.D.; Wilkinson, E.; Giandhari, J.; Naidoo, S.; Pillay, Y.; Singh, L.; et al. The Emergence and Ongoing Convergent Evolution of the N501Y Lineages Coincides with a Major Global Shift in the SARS-CoV-2 Selective Landscape. *medRxiv* **2021**, *2*, 21252268. [CrossRef]
31. Resende, P.C.; Delatorre, E.; Gräf, T.; Mir, D.; Motta, F.C.; Appolinario, L.R.; da Paixão, A.C.D.; Mendonça, A.C.D.F.; Ogrzewalska, M.; Caetano, B.; et al. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. *Front. Microbiol.* **2021**, *11*, 615280. [CrossRef]
32. Mir, D.; Rego, N.; Resende, P.C.; López-Tort, F.; Fernandez-Calero, T.; Noya, V.; Brandes, M.; Possi, T.; Arleo, M.; Reyes, N.; et al. Recurrent Dissemination of SARS-CoV-2 through the Uruguayan-Brazilian Border. *medRxiv* **2021**, *1*, 20249026. [CrossRef]

33. Naveca, F.; Nascimento, V.; Souza, V.; Corado, A.; Nascimento, F.; Silva, G.; Costa, Á.; Duarte, D.; Pessoa, K.; Gonçalves, L.; et al. Phylogenetic Relationship of SARS-CoV-2 Sequences from Amazonas with Emerging Brazilian Variants Harboring Mutations E484K and N501Y in the Spike Protein. Available online: https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585 (accessed on 24 February 2021).
34. Choi, B.; Choudhary, M.C.; Regan, J.; Sparks, J.A.; Padera, R.F.; Qiu, X.; Solomon, I.H.; Kuo, H.-H.; Boucau, J.; Bowman, K.; et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **2020**, *383*, 2291–2293. [CrossRef]
35. Kemp, S.A.; Collier, D.A.; Datir, R.P.; Ferreira, I.A.T.M.; Gayed, S.; Jahun, A.; Hosmillo, M.; Rees-Spear, C.; Mlcochova, P.; Lumb, I.U.; et al. SARS-CoV-2 Evolution during Treatment of Chronic Infection. *Nature* **2021**, *592*, 277–282. [CrossRef] [PubMed]
36. Gräf, T.; Bello, G.; Venas, T.M.M.; Pereira, E.C.; Paixão, A.C.D.; Appolinario, L.R.; Lopes, R.S.; Mendonça, A.C.d.F.; da Rocha, A.S.B.; Motta, F.C.; et al. Identification of SARS-CoV-2 P.1-Related Lineages in Brazil Provides New Insights about the Mechanisms of Emergence of Variants of Concern—SARS-CoV-2 Coronavirus / NCoV-2019 Genomic Epidemiology. Available online: https://virological.org/t/identification-of-sars-cov-2-p-1-related-lineages-in-brazil-provides-new-insights-about-the-mechanisms-of-emergence-of-variants-of-concern/694/1 (accessed on 17 May 2021).
37. Plante, J.A.; Mitchell, B.M.; Plante, K.S.; Debbink, K.; Weaver, S.C.; Menachery, V.D. The Variant Gambit: COVID-19's next Move. *Cell Host Microbe* **2021**, *29*, 508–515. [CrossRef] [PubMed]
38. Gupta, S.; Mallick, D.; Banerjee, K.; Sarkar, S.; Lee, S.T.M.; Basuchowdhuri, P.; Jana, S.S. D155Y Substitution of SARS-CoV-2 ORF3a Weakens Binding with Caveolin-1: An in Silico Study. *bioRxiv* **2021**, *3*, 437194. [CrossRef]
39. Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M.L.; et al. Detection of 2019 Novel Coronavirus (2019-NCoV) by Real-Time RT-PCR. *Eurosurveillance* **2020**, *25*, 2000045. [CrossRef]
40. World Health Organization COVID-19 Clinical Management: Living Guidance. Available online: https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-clinical-2021-1 (accessed on 1 May 2021).
41. Eden, J.-S. SARS-CoV-2 Genome Sequencing Using Long Pooled Amplicons on Illumina Platforms. *bioRxiv* **2020**. [CrossRef]
42. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef]
43. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
44. Li, H. A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef]
45. Gel, B.; Serra, E. KaryoploteR: An R/Bioconductor Package to Plot Customizable Genomes Displaying Arbitrary Data. *Bioinforma. Oxf. Engl.* **2017**, *33*, 3088–3090. [CrossRef]
46. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv* **2012**, arXiv:1207.3907. Available online: http://arxiv.org/abs/1207.3907 (accessed on 4 May 2021).
47. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [CrossRef] [PubMed]
48. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
49. Du Plessis, L. Laduplessis/SARS-CoV-2_Guangdong_Genomic_Epidemiology; Initial Release; Zenodo. Available online: https://zenodo.org/record/3922606 (accessed on 4 May 2021).
50. Rambaut, A.; Holmes, E.C.; O'Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403–1407. [CrossRef] [PubMed]
51. Argimón, S.; Abudahab, K.; Goater, R.J.E.; Fedosejev, A.; Bhai, J.; Glasner, C.; Feil, E.J.; Holden, M.T.G.; Yeats, C.A.; Grundmann, H.; et al. Microreact: Visualizing and Sharing Data for Genomic Epidemiology and Phylogeography. *Microb. Genom.* **2016**, *2*, e000093. [CrossRef]
52. Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* **2018**, *34*, 4121–4123. [CrossRef]
53. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
54. Tavare, S. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lect. Math. Life Sci.* **1986**, *17*, 57–86.
55. Sagulenko, P.; Puller, V.; Neher, R.A. TreeTime: Maximum-Likelihood Phylodynamic Analysis. *Virus Evol.* **2018**, *4*. [CrossRef]
56. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef]
57. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef] [PubMed]
58. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen). *Virus Evol.* **2016**, *2*, vew007. [CrossRef]
59. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T.-Y. Ggtree: An r Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol. Evol.* **2017**, *8*, 28–36. [CrossRef]

60. Orme, D.; Freckleton, R.; Thomas, G.; Petzoldt, T.; Fritz, S.; Isaac, N.; Pearse, W. Caper: Comparative Analyses of Phylogenetics and Evolution in R (R Package Version 1.0.1). Available online: https://www.scienceopen.com/document?vid=d750bff2-a400-41dd-aa1e-728bb7aaf4d5 (accessed on 10 May 2021).
61. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10. *Virus Evol.* **2018**, *4*, vey016. [CrossRef]
62. Ayres, D.L.; Darling, A.; Zwickl, D.J.; Beerli, P.; Holder, M.T.; Lewis, P.O.; Huelsenbeck, J.P.; Ronquist, F.; Swofford, D.L.; Cummings, M.P.; et al. BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Syst. Biol.* **2012**, *61*, 170–173. [CrossRef] [PubMed]
63. Hasegawa, M.; Kishino, H.; Yano, T. Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA. *J. Mol. Evol.* **1985**, *22*, 160–174. [CrossRef]
64. Ferreira, M.A.R.; Suchard, M.A. Bayesian Analysis of Elapsed Times in Continuous-Time Markov Chains. *Can. J. Stat.* **2008**, *36*, 355–368. [CrossRef]
65. Gill, M.S.; Lemey, P.; Faria, N.R.; Rambaut, A.; Shapiro, B.; Suchard, M.A. Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci. *Mol. Biol. Evol.* **2013**, *30*, 713–724. [CrossRef]
66. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef] [PubMed]
67. Lemey, P.; Rambaut, A.; Drummond, A.J.; Suchard, M.A. Bayesian Phylogeography Finds Its Roots. *PLoS Comput. Biol.* **2009**, *5*, e1000520. [CrossRef] [PubMed]
68. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
69. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009; ISBN 978-0-387-98141-3.
70. Pereira, R.; Gonçalves, C.; De Araujo, P.; Carvalho, G.; De Arruda, R.; Nascimento, I.; Da Costa, B.; Cavedo, W.; Andrade, P.; Da Silva, A.; et al. Geobr: Loads Shapefiles of Official Spatial Data Sets of Brazil. Available online: https://github.com/ipeaGIT/geobr (accessed on 4 May 2021).
71. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R. J.* **2018**, *10*, 439–446. [CrossRef]
72. Bielejec, F.; Baele, G.; Vrancken, B.; Suchard, M.A.; Rambaut, A.; Lemey, P. SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. *Mol. Biol. Evol.* **2016**, *33*, 2167–2169. [CrossRef] [PubMed]

APÊNDICE B Artigo 2 - Epidemiological profile of COVID-19 in Brazil: a data review

Revista

Reviews in Medical Virology

Artigo em andamento como primeiro autor, editado até o dia 8 de abril de 2022. No trabalho, será ainda incluída análise estatística com o alvo de publicar na revista *Reviews in Medical Virology* (Qualis A1; JCR 6,989).

Título

Epidemiological profile of COVID-19 in Brazil: a data review

Autores

Gabriel Dickin Caldana[1], Patrícia Aline Gröhs Ferrareze[1], Claudia Elizabeth Thompson[1,2]*

[1] Graduate Program in Health Sciences, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil.
[2] Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil.
* Corresponding author

Epidemiological profile of COVID-19 in Brazil: a data review

Address for correspondence:

Claudia Elizabeth Thompson

Department of Pharmacosciences, Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), 245/200C Sarmento Leite St, Porto Alegre, RS, Brazil. ZIP code: 90050-170. Phone: +55 (51) 3303 8889.

E-mail: cthompson@ufcspa.edu.br, thompson.ufcspa@gmail.com

Running title: Epidemiological profile of COVID-19 in Brazil

Keywords: COVID-19, epidemiology, SARS-CoV-2, pandemic, Brazil

Introduction

After its initial emergence in Wuhan, China, in late 2019, SARS-CoV-2 quickly spread around the world leading to the officially recognized COVID-19 pandemic in March 2020 (WHO). SARS-CoV-2 infection can manifest a wide variety of symptoms, ranging from asymptomatic cases and mild clinical manifestations to severe and critical cases. As of March 9, more than 450 million cases and more than 6 million deaths have been confirmed worldwide. In Brazil, the third most affected country by COVID-19, 29.2 million cases and 653 thousand deaths were reported (Dong, 2020; Brazilian Health Ministry). From a virological point of view, this may be related to the continental magnitude of Brazil, allowing multiple viral introductions and the emergence of new Variants of Concern (VOCs) with increased infectious potential (Candido, 2020). The COVID-19's hotspots changed during 2020, starting in Asia, followed by Europe and the Americas. After months of gradual evolution, new VOCs (e.g. Alpha, Beta, Gamma, Delta, and Omicron) carrying signature amino acid substitutions in the spike protein have emerged (Mullen, 2020). These lineages independently

arose in the United Kingdom, South Africa, Brazil, India, and once more in South Africa (Rambaut, 2021; Tegally, 2021; Faria, 2021; Naveca, 2021; Cherian, 2021; Callaway, 2021), leading to secondary outbreaks, regardless the extremely high seroprevalence estimates (Buss, 2021).

An alarming number of cases and deaths in Brazil by recurrent Gamma (P.1 and sublineages) became evident in December 2020. The variant harbors mutations of potential biological significance in the spike's receptor-binding domain (RBD) such as E484K, K417T, and N501Y (Mullen, 2020). These amino acid substitutions may lead to immunity evasion and increased transmissibility (Sabino, 2021). After this outbreak, a large increase of cases, hospitalizations, and deaths resulted in the second wave of COVID-19, similarly experienced in the first wave (Brazilian Health Ministry). In Brazil, the second wave started in Manaus, northern Brazil, where the P.1 lineage was first described, causing a large number of deaths in a panorama of lack of oxygen and ICU beds (Sabino, 2021).

The Delta variant (lineage B.1.617.2 and AY sublineages), firstly detected in India at the end of January 2021, harbors L452R and P681R. Among others, these are the most relevant mutations (Cherian, 2021). The first one grants a higher affinity to the ACE2 receptor (Starr, 2021) and might allow the evasion of CD8 T cells antiviral activity (Website - Koshy, 2021). The second substitution would facilitate the membrane fusion and integration of the virus to the host cell (Website - Haseltine, 2021; Bertram, 2013). The combination of the aforementioned and co-occurring mutations may be responsible for Delta higher transmissibility (Shiehzadegan, 2021)

On 26 November, the WHO named the new lineage spotted in South Africa, BA.1, as Omicron variant, joining Delta (lineage B.1.617.2 and AY lineages), Alpha (lineage B.1.1.7), Beta (lineage B.1.351), and Gamma (P.1 and sublineages) on the current list of VOCs. This variant contains more than 30 mutations in the spike protein, many of which have already been found in other VOCs. Amino acid substitutions at sites 417, 484, and 493 have been associated with the escape of antibody neutralization (Han, 2022). These changes are linked to elevated infectivity and immune evasion (Callaway, 2021; Torjesen, 2021).

Concurrently to the increase of the vaccination rates, there was observed a reduction of the number of cases and deaths in the second half of 2021, in Brazil (aTaylor, 2022). The emergency of Omicron increased the number of positive cases, doubling from one week to another in January 2022. However, the number of deaths is not even close to the rates that Brazil reached in the years 2020 and 2021 in the first and second waves. Concomitantly, high vaccination rates, represented by almost 70% of the Brazilian population with a complete vaccination schedule by January 2022, might have been preventing more serious cases of COVID-19 (Reuters, 2022; Maisa, 2022; Maslo, 2022; UK Health Security Agency, 2022).

Due to the viral genotypic diversity, it is important to study the dynamics of SARS-CoV-2. Precious information can be obtained about the proteins and viral genes through molecular techniques and bioinformatics. This data might support the development of nonpharmaceutical interventions, vaccines, drugs, and other possible solutions to protect and treat the population. International viral sequencing efforts have allowed the submission of millions of genomes in the Global Initiative on Sharing All Influenza Data (GISAID) (Shu and McCauley, 2020). This study aims to investigate some epidemiological aspects in Brazil, analyzing the available genome sequences to date, to better understand the dynamics of SARS-CoV-2 in the country regarding several consensus statistics information.

METHODOLOGY

Brazilian sequences and metadata

On December 7th, 2021, complete genomes and metadata were downloaded from the GISAID platform for each Brazilian state after the exclusion of sequences with low coverage. The metadata provide information about the lineage, sex, age, collection date, and location of the Brazilian sequences.

Consensus statistics, cases and deaths

In order to analyze the epidemiological information available, we extracted data from the Brazilian Institute of Geography and Statistics (IBGE) website (https://sidra.ibge.gov.br/tabela/6706) and the Brasil.IO project (https://brasil.io/dataset/covid19/caso_full/ - accessed on December 10th, 2021). These data allow the analysis of the Brazilian population in the different states and regions in relation to population estimates, age, sex, and COVID-19 cases, and deaths.

Mean, median, interquartile range, maximum and minimum values were calculated based on all age information from each Brazilian state. This resulted in a table with five columns - for each region - and several rows - for each age entry. Splitting the genome counts by months in the columns and regions in the rows allowed the visualization of the sequencing counts and proportions in the whole period. Figure 2 was built from this splitted data inserting a stacked column graph, resulting in figures for the absolute counts per month and the contribution of each region.

Information about cases and deaths extracted from the Brasil.IO project were summarized into pivot tables for each Brazilian region, where the columns represent the cases and deaths proportions - calculated based on the population estimates taken from the IBGE repository - and the rows represent months. These tables provide the insertion of combined charts with columns portraying the cases per 100 thousand inhabitants in the primary Y axis, the line the deaths per 100 thousand inhabitants in the secondary Y axis and the time in months in the X axis.

Cases and deaths per 100 thousand inhabitants were calculated with the formula: (total number of cases or deaths/population of the region or the state)x100,000. Case-fatality rates (%) were calculated considering the ratio of the deaths per 100k inhabitants and cases per 100k inhabitants: (deaths per 100,000/cases per 100,000)x100. The correlation between the proportion of cases and the case-fatality ratio was achieved by using the CORREL function and selecting the matrixes of the analyzed entries.

Lineages and Variants frequencies

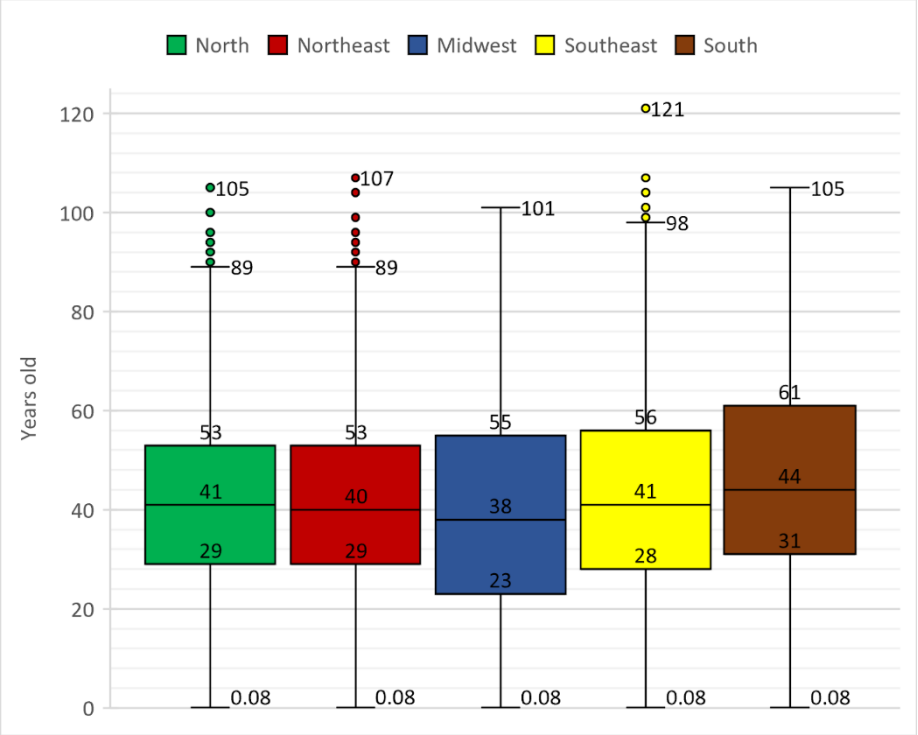
The metadata downloaded from the GISAID platform also contains the lineage for each genome, as well as the date and state of the samples collection. The lineages proportion charts were made by summarizing the aforementioned information in a pivot table, where the columns represent the months and the rows represent the designated main lineages and variants. This step also divides the proportions for every Brazilian region, yielding a chart for each one.

Results

Brazil

A total of 76,413 sequences were downloaded from the GISAID platform ranging from February 2020 to November 2021 (22 months). Of these, 22,405 (29.32%) were from male patients, 25,022 (32.75%) from females, and 28,986 (37.93%) were unknown. A count of 46,652 sequences had age information with a mean age of 42.54 years, median of 41 years, an interquartile range (IQR) of 27, minimum age of one month (0.08 year), and maximum age of 121 years - the numbers for each region can be seen in Figure 1. City information was only found for 35,919 sequences. The main lineages and the variants proportions for the whole period is shown in Table 1. Moreover, 114 samples were labeled as “None” lineage.

Figure 1. Box plot chart showing the patient age distribution for the five Brazilian regions according to the SARS-CoV-2 sequenced genomes from the GISAID database.



In 2020, Brazil sequenced 5,337 SARS-CoV-2 genomes with a mean of 485.15 genomes per month, while in 2021 - until December 7th - 71,498 with a mean of 6,499.81 per month. The whole studied period yields the mean of 724 sequenced samples per month with an upward tendency towards mid-2021 (Figure 2). The Southeast region is responsible for the majority of the sequencing in the country, followed by the Northeast, South, North, and Midwest regions (Table 1).

Figure 2. Number of sequenced SARS-CoV-2 genomes for the five Brazilian regions according to the GISAID database.

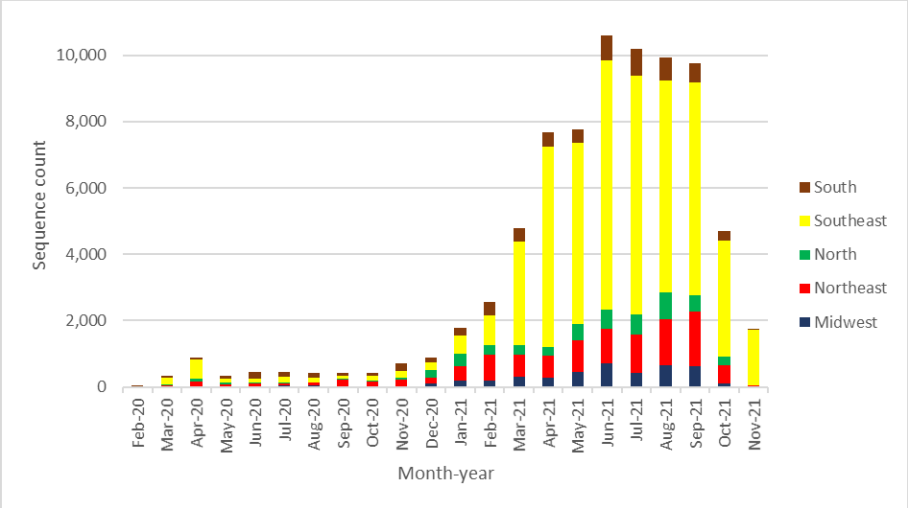


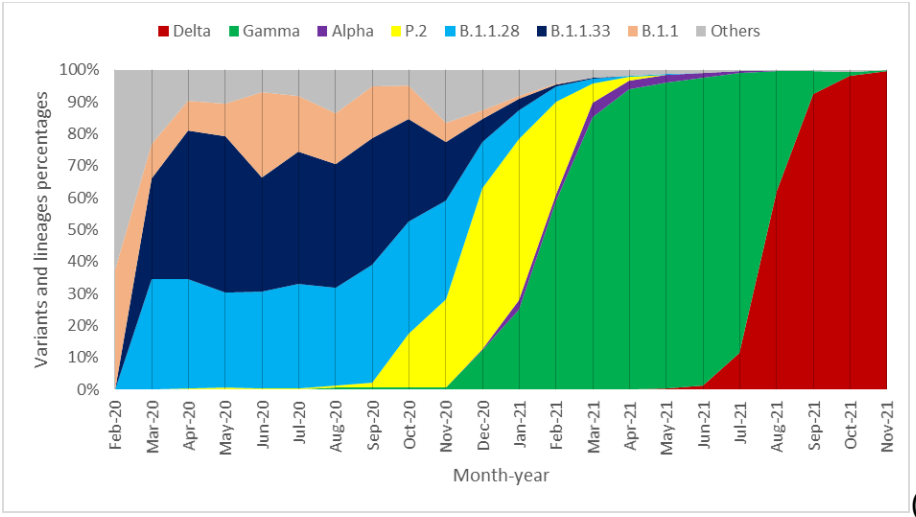
Table 1. Proportions regarding population, sequencing, cases, and deaths in the studied period for each Brazilian region

	Population (2021)	Sequences	Cumulative cases	Cumulative deaths	Sequenced cases
North	9%	7%	9%	8%	0.026%
Northeast	27%	14%	22%	19%	0.018%
Midwest	8%	5%	11%	10%	0.023%
Southeast	42%	66%	39%	48%	0.057%
South	14%	8%	19%	16%	0.021%
Brazil	**100%**	**100%**	**100%**	**100%**	**0.036%**

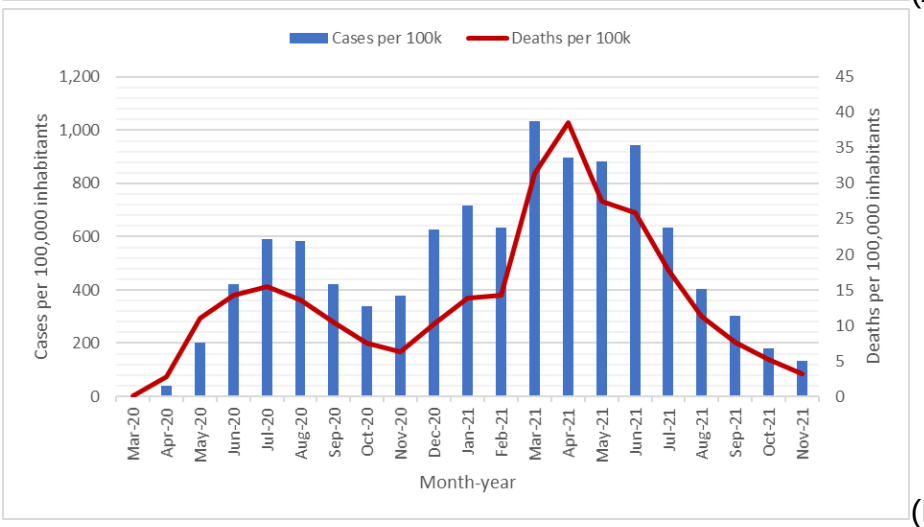
As shown in Figure 3A, until September 2020, B.1.1.33 was the most prevalent lineage followed by B.1.1.28 and B.1.1. After that, the three lineages started to decline. Alpha variant had a discrete presence in Brazil and reached a maximum of 4.56% of the studied genomes in March 2020. Beta had a minor prevalence with only ten sequences in 2021 - nine in March and April in the Southeast and one in August in the Northeast. P.2 emerged in the country in early 2020 and significantly increased in percentage between December 2020 and January 2021, being quickly replaced by Gamma in the following months. Between March and July 2021, the variant yielded the highest proportion, corresponding to more than 90% of the sequenced genomes between April and June 2021. In addition, this variant represented 57.8% of all Brazilian sequenced

genomes in the period (Table 2). From July 2021 towards the end of the year, Delta overthrew Gamma in proportion, reaching more than 90% of the sequences in September 2021. Furthermore, Supplementary Figure 4 highlights the distribution of genomes among the five different Brazilian regions considering age ranges.

Figure 3. SARS-CoV-2 lineage frequencies and case-fatality ratio for the Brazilian territory up to December 7, 2021. (A) Lineage frequencies according to the SARS-CoV-2 sequenced genomes available on the GISAID database. (B) Proportion of cases and deaths for 100,000 inhabitants according to the Brasil.IO project database.



(A)



(B)

Table 2. Overall proportions of the main lineages and variants on each Brazilian region in the studied period.

	North	Northeast	Midwest	Southeast	South	Brazil
Alpha	0.02%	0.56%	1.20%	1.64%	0.43%	**1.26%**
Gamma	74.73%	53.28%	69.46%	57.64%	46.01%	**57.80%**
Delta	11.76%	28.00%	18.40%	33.31%	26.48%	**29.86%**
B.1.1	0.42%	3.20%	0.21%	0.18%	2.83%	**0.83%**
B.1.1.28	3.38%	3.56%	1.81%	1.90%	6.16%	**2.57%**
B.1.1.33	2.12%	3.18%	2.09%	1.65%	6.21%	**2.29%**
P.2	4.60%	5.72%	5.78%	2.28%	7.75%	**3.53%**
Others	2.98%	2.51%	1.05%	1.39%	4.11%	**1.86%**

Table 3 displays the proportions of cases and deaths, and the case-fatality rate of COVID-19 in the studied period. Until December 7th, the accumulated number of 22,163,231 cases and 616,293 deaths were recorded. These indicate an overall measure of 10,390 cases and 289 deaths per 100 thousand inhabitants with an overall case-fatality of 2.78% - considering the estimated population of 213,317,639 inhabitants (IBGE). Interestingly, we found a moderate negative correlation (r = -0.56) between the proportion of cases and the case-fatality ratio. For example, Roraima state (RR - Northern Brazil) yielded the highest proportion of cases (19,717 per 100k) but had a remarkably low mortality rate of 1.6%. In contrast, Rio de Janeiro state (RJ - Southeastern Brazil) had a relatively low proportion of cases (7,720 per 100k) but presented the highest case-fatality ratio of all states.

Table 3. Overall cases and deaths per 100 thousand inhabitants and case-fatality rate on each Brazilian state and their respective regions in the studied period.

Region	State	Cases per 100k		Deaths per 100k		Case-fatality	
		Region	State	Region	State	Region	State
North	AC	10,048	9,730	250	204	2.49%	2.10%
	AM		10,089		323		3.21%
	AP		14,252		228		1.60%
	PA		6,987		193		2.77%
	RO		15,419		367		2.38%
	RR		19,717		315		1.60%
	TO		14,546		244		1.68%
Northeast	AL	8,538	7,183	207	189	2.42%	2.63%
	BA		8,432		183		2.17%
	CE		10,308		267		2.59%
	MA		5,124		144		2.82%
	PB		11,379		235		2.07%
	PE		6,636		210		3.16%
	PI		10,127		219		2.17%
	RN		10,776		211		1.96%
	SE		11,902		259		2.17%
Midwest	DF	14,326	16,747	355	357	2.48%	2.13%
	GO		13,071		341		2.61%
	MS		13,363		342		2.56%
	MT		15,527		393		2.53%
Southeast	ES	9,628	15,168	327	322	3.40%	2.12%
	MG		10,333		263		2.55%
	RJ		7,720		396		5.13%
	SP		9,531		331		3.48%
South	PR	14,199	13,660	319	352	2.25%	2.58%
	RS		13,051		316		2.42%
	SC		16,842		273		1.62%
BRAZIL		**10,390**		**289**		**2.78%**	

July 2020 was the worst month in relation to the number of cases and deaths in the first wave of COVID-19 pandemics in Brazil. In this period, there was a count of 1.25 million cases (589.6 cases per 100 thousand inhabitants) and 32,936 deaths. October and November mark the valley between waves. Respectively, over 720.3 thousand cases (337.7 cases per 100 thousand inhabitants) and 13,296 deaths were registered these months. The utmost peak of cases and deaths in the country was in March 2021, with over 2.2 million cases (1,033.8 cases per 100 thousand inhabitants), and in April 2021, with over 82.2 thousand deaths (Figure 3B). This time spam represents the apex of the second

wave of COVID-19 in the country and coincides with the very steep ascending proportion of the Gamma variant introduced a few months earlier. Interestingly, in February 2021, the incidence of SARS-CoV-2 decreased in comparison to the previous month. Given the continental magnitude of Brazil, each region presented different counts, rates, and proportions of cases and deaths

Regional Data

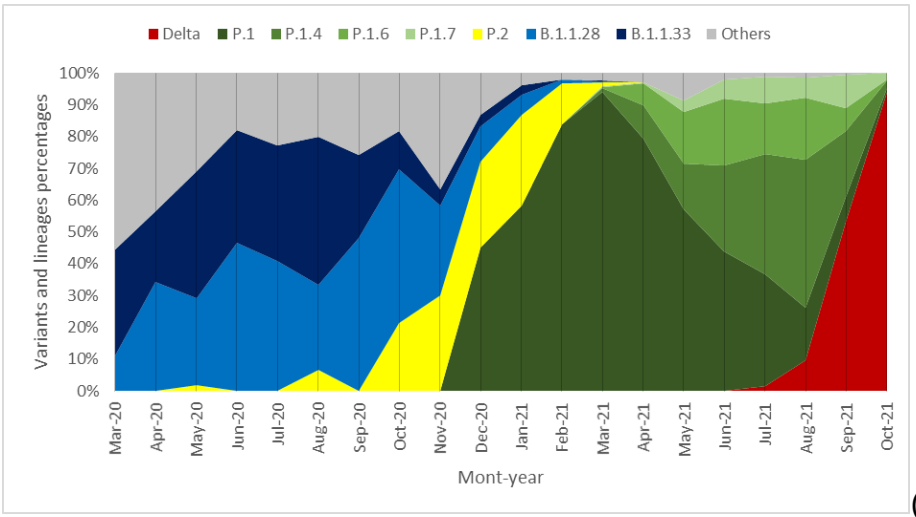
Northern Brazil

A total of 5,002 sequences were downloaded from the GISAID platform ranging from March 2020 to October 2021 (20 months). A total of 2,341 (46.80%) sequences were from male patients, 2,327 (46.52%) from female patients, and 334 (6.68%) were unknown. Of these, 4,754 sequences had associated age information with a mean age of 42.18 years (median = 41, IQR = 24, maximum of one month and minimum of 105 shown in Figure 1). City information was found for 4,629 sequences. The North region represents 9% of Brazilian inhabitants and includes seven states: Acre, Amazonas, Amapá, Pará, Rondônia, Roraima, and Tocantins. Furthermore, the region has the lowest population density (4.9 people per km^2) and exhibits a younger population - with a higher proportion between 0 and 29 years old (Supplementary table 3). The distribution of main lineages and variants in the studied period can be seen in Table 2.

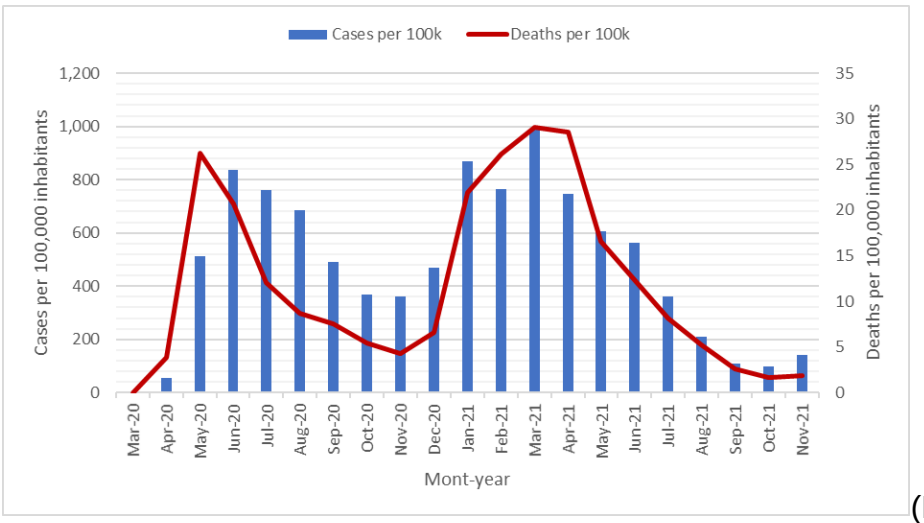
The first wave of COVID-19 had an abrupt growth in May 2020, which represents the third month with more COVID-19 deaths (4,962) and proportion of cases (836.1 per 100 thousand inhabitants). Between the first and second waves, November was the bottom valley with 817 deaths and 362.1 cases per 100 thousand inhabitants. In December 2020, the region had the largest amount of sequencing concomitantly to a major introduction of P.1 near the end of 2020 (Figure 4A). The rise was very sharp and reached more than 90% of the sequences in March 2021. In this month, North had the highest proportion of cases per 100 thousand inhabitants (992.6) and the highest number of people dying from COVID-19 per month (5,505). Moreover, the diversification of P.1 in different sublineages, in particular P.1.4, P.1.6, and P.1.7, started around

February to March 2021 (Figure 4A). Gamma represented over 74% of the sequences, which is the highest percentage of this variant among the regions (Table 2). Unlike other Brazilian territories, Delta had a delayed escalation, reaching 50% only in September (Figure 4A). It also showed the lowest proportion in comparison to other regions (11.76%) considering the whole studied period (Table 2).

Figure 4. SARS-CoV-2 lineage frequencies and case-fatality ratio for Northern Brazil up to December 7, 2021. (A) Lineage frequencies according to the SARS-CoV-2 sequenced genomes from the GISAID database. (B) Proportion of cases and deaths for 100,000 inhabitants according to the Brasil.IO project database.



(A)



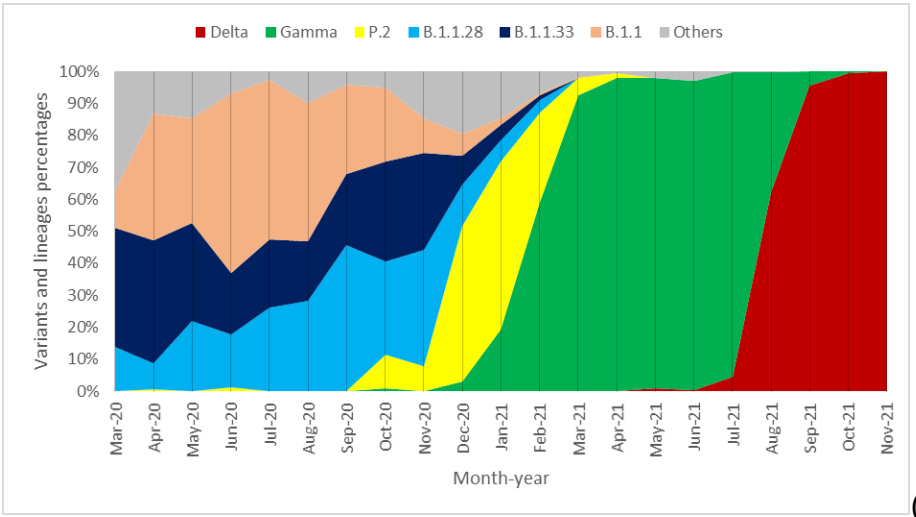
(B)

Northeastern Brazil

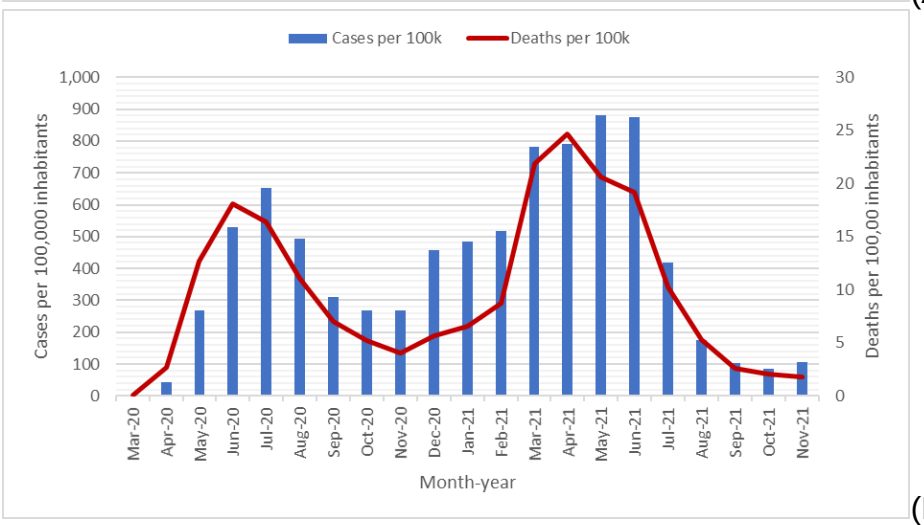
A number of 10,569 sequences were downloaded from the GISAID platform ranging from February 2020 to November 2021 (22 months). Of these, 4,478 (42.37%) were from male patients, 5,056 (47.84%) from female patients and 1,035 (9.79%) were unknown. A total of 9,325 sequences had age information, with a mean age of 41.66 years (median = 40, IQR = 24, maximum of 107 and minimum of one month, Figure 1). City information was available for 8,632 sequences. Furthermore, the region was responsible for sequencing 14% of the Brazilian genomes - the second-highest percentage. The Northeast contains 27% of the Brazilian inhabitants, with a population density of 37.2 people per km^2 and nine states: Alagoas, Bahia, Ceará, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte, and Sergipe (Tables 2 and 6). People between 10 and 19 years old are the majority among other age ranges (Supplementary table 3). The main lineages and variants proportions in the studied period can be seen in Table 2.

In the first wave, in mid-2020, the highest proportion of cases was 652.7 per 100 thousand inhabitants in July and the largest number of deaths was in June when 4,962 people died due to COVID-19. These months had a larger proportion of B.1.1 reaching more than half of the genomes in the period. Like other regions, B.1.1.28 and B.1.1.33 were present in the first year of pandemics, but in lower percentages. November was the period between waves, in which the region had 268 cases per 100 thousand people and 2,321 deaths. With the rise of Gamma, which is responsible for 53.28% of all genomes in the region, and P.2 at the end of 2020 and beginning of 2021, the first three mentioned lineages already started to drop. After that, Gamma reached more than 95% of the total sequences in April 2021, which is the month that yields the largest number of deaths (14,212). In the following two months, the proportion of cases was 880.6 per 100 thousand inhabitants. In addition, this period still had more than 95% of Gamma, but the number of deaths started to sink. Delta appeared in June 2021, reaching a proportion around 90% already in September and 100% in November. This variant represented 28% of the genomes in the Northeast (Figure 5 and Table 5b).

Figure 5. SARS-CoV-2 lineage frequencies and case-fatality ratio for Northeastern Brazil up to December 7, 2021. (A) Lineage frequencies according to the SARS-CoV-2 sequenced genomes from the GISAID database. (B) Proportion of cases and deaths for 100,000 inhabitants according to the Brasil.IO project database.



(A)



(B)

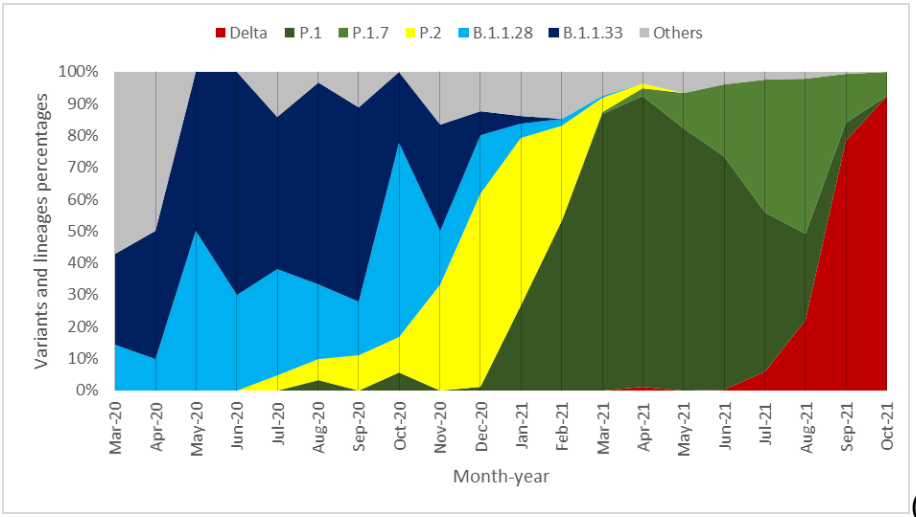
Midwestern Brazil

A total of 3,821 sequences were downloaded from the GISAID platform ranging from March 2020 to October 2021 (20 months). Of these, 1,808 (47.32%) were from male patients, 1,880 (49.20%) from female patients and 133 (3.48%) were unknown. A total of 3,473 sequences had associated age information. The mean and median ages were the lowest among other regions - respectively 39.66 and 38 years (IQR = 32, maximum of 101 and minimum of one month, Figure 1). City information was available for 3,417 sequences. The Midwest region is

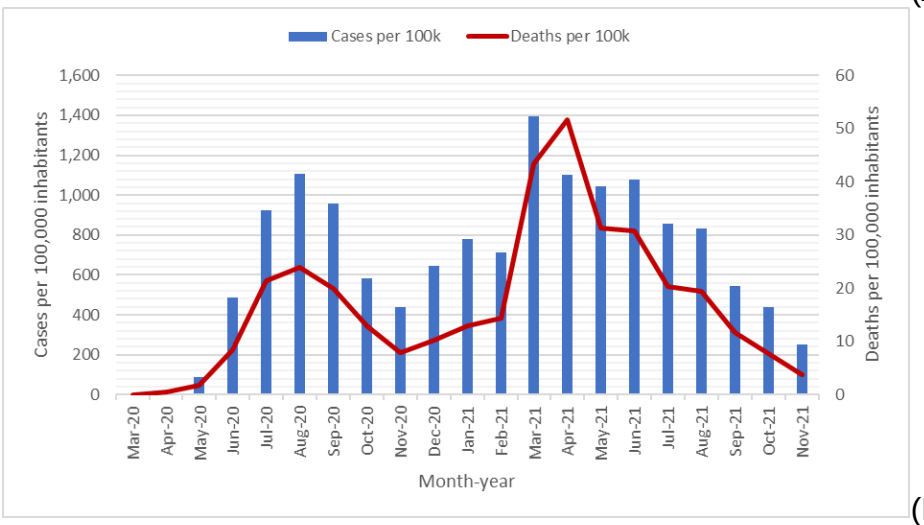
composed of four states: Distrito Federal, where the Brazilian capital city Brasília is located, Goiás, Mato Grosso do Sul, and Mato Grosso. Despite the low population density of the region (10.4 people per km^2), Distrito Federal has 537.1 people per km^2, the largest density of all Brazilian states. The Midwest region represents the smallest portion of the country's population and contributed with the lowest percentage of genome sequencing (Table 2). The region yields the highest overall cases per 100 thousand people (14,325.7) and the highest overall deaths per 100 thousand people (355.2), which is even larger than Southeast's that has a much greater population density. Moreover, it is possible to observe that Midwest had younger patients (Supplementary table 4) - the bars between 1 and 20 years old are over the average line and over all other regions. This agrees with the fact that the region has a younger population (Supplementary table 3). The main lineages and variants proportions in the studied period can be seen in Table 2.

August 2020, in the first wave, yields the highest number of cases per 100 thousand inhabitants and deaths - respectively 1,0104.7 and 4,003. Moreover, between May and September 2020, the B.1.1.33 was the most prevalent lineage, ranging from about 50% to 70% in the period. November was the month between waves with 645.5 cases per 100 thousand people and 1,706 deaths. From this month forward, with the escalation of P.2 and Gamma around November and December 2020, the second wave started to rise. P.2 reached almost 60% in December and Gamma more than 93% in April 2021. The number of cases reached a maximum in March 2021 (1,395.1 cases per 100 thousand people) and the deaths peaked in the following month (8,638 deaths) (Figure 6). In addition, Gamma still represented more than 90% of the sequenced genomes until July, but the deaths started to drop. The Delta variant started to rise in July 2021 and steeply reached a proportion of 92% in October. Among the genomes sequenced in the Midwest region, Gamma represented the vast majority with 69.46% and Delta with just 18.40%. Furthermore, a noteworthy diversification of P.1 was also apparent.

Figure 6. SARS-CoV-2 lineage frequencies and case-fatality ratio for Midwestern Brazil up to December 7, 2021. (A) Lineage frequencies according to the SARS-CoV-2 sequenced genomes from the GISAID database. (B) Proportion of cases and deaths for 100,000 inhabitants according to the Brasil.IO project database.



(A)



(B)

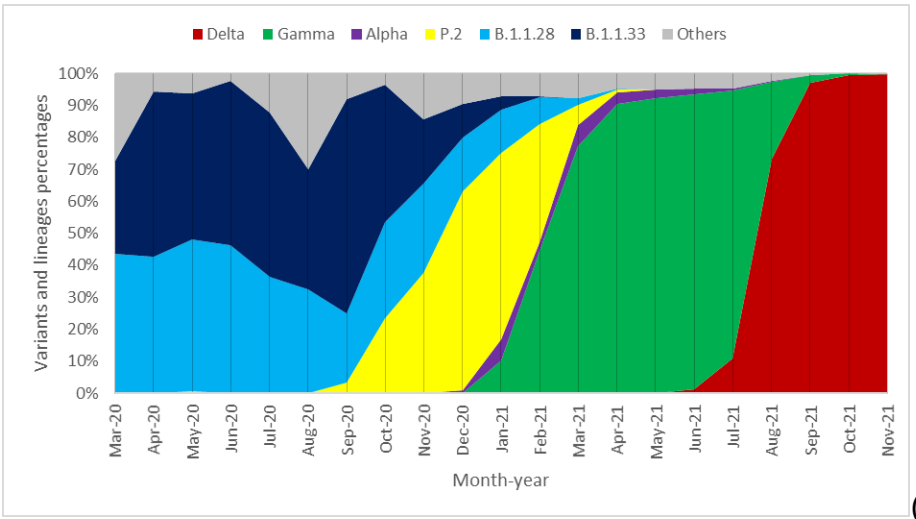
Southeastern Brazil

A total of 50,775 sequences were downloaded from the GISAID platform ranging from February 2020 to November 2021 (22 months). Of these, 11,237 (22.13%) were from male patients, 13,048 (25.70%) from female patients and 26,490 (52.17%) were unknown. A number of 23,622 sequences had age information with a mean of 42.55 years (median = 41, IQR = 28, maximum of 121 and minimum of one month, Figure 1). The region is composed of slightly older people in comparison to the aforementioned regions (Supplementary table 3). City information was available for 15,227 sequences. The Southeast region is

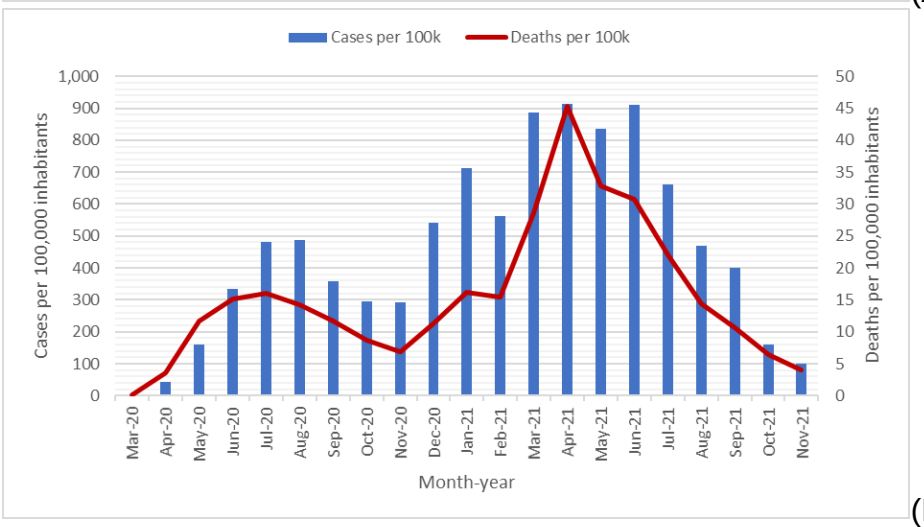
composed of four states (Espírito Santo, Minas Gerais, Rio de Janeiro, and São Paulo) and has a population density of 96.9 people per km^2, the largest of all regions (Table 6). Furthermore, the population density of Rio de Janeiro and São Paulo are 399.2 and 187.9, the second and third highest of all Brazilian states, respectively (Supplementary table 1). The region represents the bulkiest portion of the country's population (42%), hence contributing with the highest sequencing rate (66%) of the Brazilian genomes. Rio de Janeiro and São Paulo Southeast hold the first and second place among all Brazilian states in overall case-fatality rates, 5.13% and 3.48%, respectively. These numbers of the Southeast region are the highest among all territories (3.398%). However, the territory yields an overall proportion of 9.627,8 cases per 100 thousand inhabitants, the second-lowest (Table 1). The main lineages and variants proportions in the studied period can be seen in Table 2.

Characterizing the apex of the first wave of COVID-19, July and August 2020 had a similar proportion of cases - 482.3 and 488.2 cases per 100 thousand people, respectively - and July had a total of 14,332 deaths. At this period, the most prevalent lineages were B.1.1.28 and B.1.1.33 fluctuating between 70% and 90%. The valley between waves was in November 2020, when 291.8 cases per 100 thousand people and 6,224 deaths were documented. In this month, P.2 was the most prevalent lineage, followed by B.1.1.28 and B.1.1.33. In December 2020 and January 2021, P.2 represented around 60% of the sequenced genomes. At the same time, Gamma started to steeply escalate and reached more than 92% in May and June 2021. The Alpha variant had a minor prevalence throughout the first half of 2021 and did not exceed 6.5% in any month. The second wave can be characterized by two peaks of cases, in April and June 2021 - 912.6 and 910.1 cases per 100 thousand people, respectively - and a sharp rise of deaths in April - 40,609. In addition, Gamma represented 90% and 92% of the sequenced genomes in April and June. The first sequences of Delta started to emerge around the middle of 2021 and abruptly rose in proportion, reaching more than 90% in September. Of all genomes sequenced in the Southeast region, Gamma was identified in 57.64% and Delta, the highest proportion among regions, in 33.31% of them (Figure 7).

Figure 7. SARS-CoV-2 lineage frequencies and case-fatality ratio for Southeastern Brazil up to December 7, 2021. (A) Lineage frequencies according to the SARS-CoV-2 sequenced genomes from the GISAID database. (B) Proportion of cases and deaths for 100,000 inhabitants according to the Brasil.IO project database.



(A)



(B)

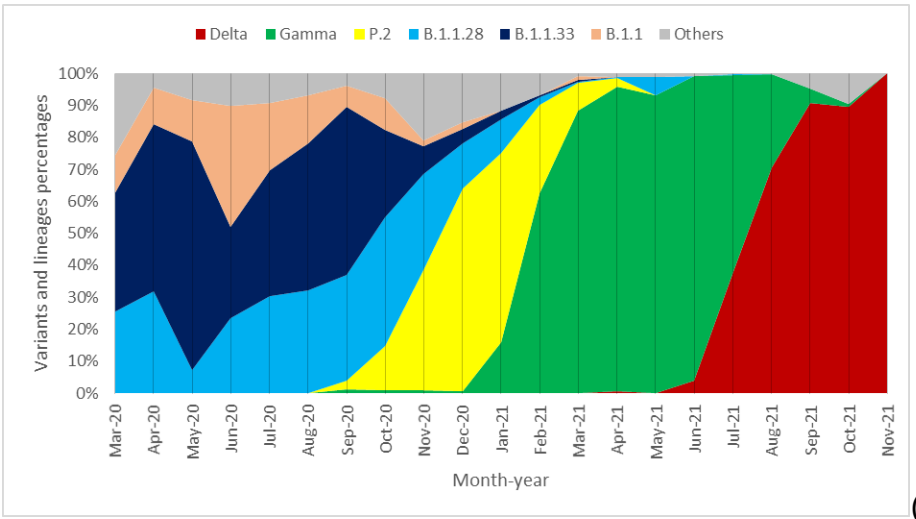
Southern Brazil

A total of 6,246 sequences were downloaded from the GISAID platform ranging from March 2020 to November 2021 (21 months). Of these, 2,711 (43.40%) sequences were male patients, 2,541 (40.68%) female patients and 994 (15.91%) were unknown. A total of 5,507 genomes had associated age information with a mean of 46.10 years (median = 41, IQR = 28, maximum of 121 and minimum of one month, Figure 1), the highest of all five Brazilian regions. Additionally, the South region is composed of a slightly older population, with a higher percentage of inhabitants between 50 and 79 years old in comparison to the other regions (Supplementary table 3). City information was

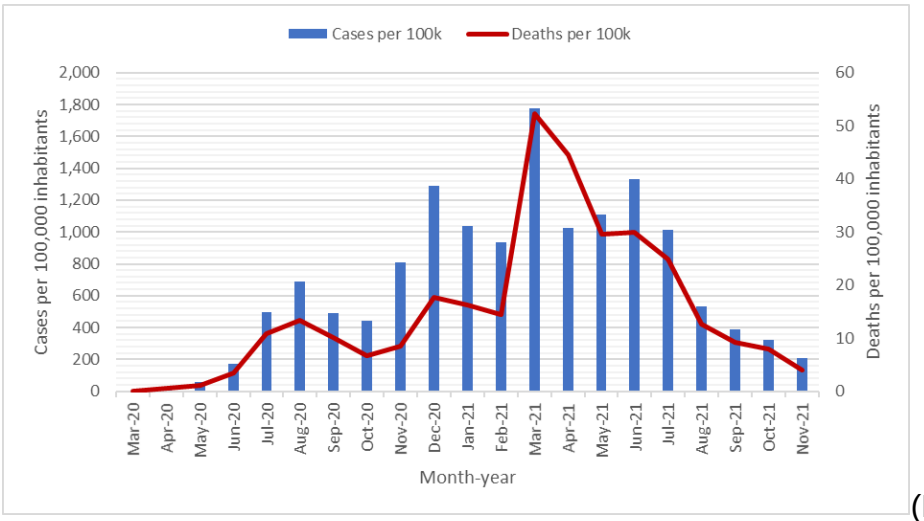
available for 4,014 sequences. The region is divided into three states (Paraná, Santa Catarina, and Rio Grande do Sul), representing 14% of Brazilians, and has a population density of 52.7 people per km^2 (Tables 2 and 6 and Supplementary table 1). The overall case-fatality is the lowest (2.25%) and the overall proportion of cases is the second-highest among territories (14,199 per 100 thousand inhabitants). The main lineages and variants distributions in the studied period can be seen in Table 2. Furthermore, B.1.1 corresponded to a significant percentage of sequences in the first year of pandemics - similarly to the Northeast.

The waves of COVID-19 in this region appear to be characterized by a different pattern (Figure 8). There are four peaks of cases (August and December 2020, and March and June 2021) and three valleys (October 2020 and February and April 2021). Moreover, the apex in the studied period was in March 2021, when 1,775 cases per thousand people (the largest proportion among all regions in all months) and 15,919 deaths were registered. The most prevalent lineage/variant in each peak was B.1.1.33 in August 2020 (45%), P.2 in December 2020 (63%), and Gamma in both March (88%) and June (95%) 2021. Different from other regions, the Gamma variant represented only 46% of the sequenced genomes, whereas B.1.1.28, B.1.1.33, and P.2 yield the largest proportions in comparison to other territories - respectively, 6.16%, 6.21%, and 7.75%. Meanwhile, Delta emerged in April 2021 reaching 70% of the sequences in August (Figure 8). In addition, this variant corresponded to 26.48% of the sequencing efforts in the region.

Figure 8. SARS-CoV-2 lineage frequencies and case-fatality ratio for Southern Brazil up to December 7, 2021. (A) Lineage frequencies according to the SARS-CoV-2 sequenced genomes from the GISAID database. (B) Proportion of cases and deaths for 100,000 inhabitants according to the Brasil.IO project database.



(A)



(B)

4 Discussion

This work is a retrospective study of the epidemiological course of SARS-CoV-2 in Brazil involving every territory, which is truly pertinent to the understanding of viral behavior. The country sequenced a total of 76,413 samples until December 7th, 2021. Unfortunately, it was an insufficient sequencing effort considering the number of cases in the same period (>22.140.000) representing around 0.34%. Apart from the North, all regions had the majority of their sequences from female patients. The Southeast yields the largest difference, with 7.2% fewer male patients (Supplementary Figure 6). Furthermore, the Southeast

and South regions have an older population, with a higher proportion of patients over 60 years old and fewer under 30 years old (Supplementary Figures 3 and 4).

The first pandemic wave occurred in different months in each region, ranging from May to August 2020. At this time, B.1.1.28, B.1.1.33, and B.1.1 were the most prevalent lineages. Figure 1A shows that until November 2020, there were few sequenced samples and the efforts on sequencing SARS-CoV-2 increased only after December 2020. This period was the onset of the second wave, which devastated Brazil between March and June 2021, reaching the astonishing mark of 82,392 deaths in April 2021. Furthermore, the higher number of cases and deaths - considering the whole studied period and all regions - belong to the South in March 2021. In this month, the region reached a total of 539,855 cases and 15,919 deaths, which represented 1,775.7 cases per 100 thousand inhabitants and 52.36 deaths per 100 thousand inhabitants (Supplementary Tables 3 and 4). Porto Alegre, the capital city of Rio Grande do Sul state, was considered a major epicenter of the collapse of the Brazilian healthcare system in the second wave. In March 2021, the state reached a death proportion of 64 deaths per 100 thousand inhabitants, similar to what Amazonas experienced in January and February (Supplementary Tables 5 and 6), with growth of COVID-19 cases in the state of Amazonas and the escalation of the Gamma variant. There were even higher proportions of deaths in the apex of the second wave in comparison to the first wave, characterizing the second wave as deadlier (Supplementary Table 6).

P.1 emerged in Manaus, firstly detected in travelers returning to Japan from Amazonas on January 2nd, 2021 (Fujino, 2021), and soon was recognized as a VOC. The lineage harbors 21 lineage-defining mutations, ten of them in the Spike protein. The emergence of this VOC was attributed as a cause of the second wave of COVID-19 in Brazil (Sabino, 2021). However, precise elucidation is difficult due to the low sequencing rate in the country, especially before December 2021 (Naveca, 2021). In Manaus (Amazonas state capital city, in the North region), a study indicated that possibly around 76% of the population had already been infected with SARS-CoV-2 by October 2020. (Buss, 2020; Fontanet, 2020) The seroprevalence in Manaus would be above the theoretical

herd immunity threshold of 67%. Within these circumstances, the abrupt increase in hospital admissions due to COVID-19 in Manaus during January 2021 was unforeseen and critical (Faria, 2021). After a sharp peak in late April 2020, COVID-19 hospitalizations in Manaus remained stable from May to November 2021, despite the relaxation of sanitary protocols (Sabino, 2021).

The endemic transmission of SARS-CoV-2 in Amazonas in the apex of the second wave was influenced by the spread of hitherto new local P.1 sublineages that are more transmissible (Naveca, 2021 preprint). These were named as “P.1 plus” because of their independently acquired new substitutions in the Spike protein (for instance P.1+679K, P.1+681H, P.1+681R, and P.1+NTDdels). The main sublineages that rose in proportion around March 2021 contain mutations such as S:N679K (P.1.4) and S:P681H (P.1.6 and P.1.7). Furthermore, the first substitution is currently known to be present in Omicron and the second one in Omicron and Alpha. After March 2021, P.1 already started to decline in proportion - overthrown by its sublineages. In August 2021, P.1 plus was responsible for more than 72% of the sequenced genomes in the North region (Figure 4). A study indicated that the endemic transmission of SARS-CoV-2 in Amazonas during the second wave was linked with the evolution of Gamma through the acquisition of NTD deletions and the aforementioned furin cleavage site mutations (P681H and N679K) (Naveca, 2021 preprint). The study also observed that P.1.4 and P.1.6 were associated with a 6-fold increase in the RNA load, suggesting that P.1 plus infected individuals might be more infectious than those harboring only the parental P.1 mutations. The endemic transmission may have selected more infectious lineages over lineages that are fitter to evade the immune system. In the Midwest region, P.1.7 emerged in March 2021 and reached almost half of the sequenced genomes in August. Apparently, this lineage arose nearly at the same time as in the North and, more discreetly, in the Northeast. This similarity concurs with the geographic proximity of the regions. Four states from the North (Rondônia, Amazonas, Pará, and Tocantins) border two states from the Midwest (Mato Grosso, and Goiás); two states of the North (Pará and Tocantins) border three states of the Northeast (Maranhão, Piauí, and Bahia); one state of the Midwest (Goiás) border one state of the Northeast (Bahia).

With the drop of the incidence of COVID-19 around October and November 2020, many states and municipalities relaxed the pandemic protocols. But the emergence of P.1 and the end of the year celebrations concurred in the following months, which marks the ascend of the second wave. In February 2021, the prevalence of SARS-CoV-2 decreased in comparison to the previous month. With exception of the Northeast, all regions experienced a lowering in the number of cases and a steep rise in March, and some concurring circumstances might be considered. A few months earlier, P.1 was first introduced in the North and quickly spread around the country in the following months (Fujino, 2021). At the same time, two other highly transmissible VOCs, Alpha and Beta, were emerging in other countries (Rambaut, 2021; Tegally, 2021).

The vaccination in the Brazilian territory started officially on January 17th, 2021 in São Paulo, the most populated state. Health workers, people over 75 years old, people over 60 years old who live in homes for the elderly, and the indigenous population were prioritized (Empresa Brasil de Comunicação, 2021). In August, half of the Brazilian population was vaccinated with at least one dose (Ritchie, 2020). Brazil's vaccination campaign converged with a greatly reduced number of hospital admissions and deaths in the second half of 2021 reaching 67,5% of the Brazilian population fully vaccinated by the end of 2021 (Taylor, 2022; Ritchie, 2020). As of Dec 9th, more than 315 million COVID-19 vaccine doses were applied in the Brazilian population (Brazilian Health Ministry, vaccination data). Over 139.4 million people were fully immunized (second dose or single dose) and over 15.4 million booster doses were applied. After cases fall in mid 2021, a sharp rise in the number of Omicron cases was already noticeable in the first weeks of January but with proportionately fewer deaths (Taylor, 2022). According to the GISAID database, the first Omicron sequences in the country are from late November 2021. Globally, the number of new COVID-19 infections recorded from late December 2021 to the first days of January 2022 increased by 71% from the previous week. Around 9.5 million cases were registered worldwide - the highest number to date and a quick turnabout of the steady decline in infections that the world has experienced since October 2021. Weekly infections rose by 65%, 78%, and 100% in Europe, Southeast Asia, in the Americas, respectively (Taylor, 2022).

The insufficient and different sequencing efforts can be considered a limitation. Despite the regions having different numbers of inhabitants, the percentage of sequenced genomes is not proportional to the population sizes. The dissimilarity is larger in the Southeast, Northeast, and South regions. They represent 42%, 27%, and 14% of the Brazilian population, respectively, but contributed with the contrasting sequencing rates of 66%, 14%, and 8%. Furthermore, the Northeast yields a disparately higher number of genomes and, nevertheless, has a sequencing rate of 0.018%, the lowest among other territories (Table 1). The Southeast has the largest percentage of sequenced genomes from its own cases. This area is known for having not only some of the most populated states but also for housing the epicenter of Brazilian academic research. Research institutes and facilities in Biology, Health Sciences, and Public Health like Instituto Oswaldo Cruz, Butantan, Instituto Adolfo Lutz, and large universities are localized mainly in the States of São Paulo and Rio de Janeiro.

FINAL REMARKS

The review of COVID-19 data in Brazil unravels the course of the pandemic in the country during the studied period. This kind of scrutiny made it possible to identify the waves and proportions of cases and deaths in each Brazilian region. Furthermore, it was possible to observe the rise and decline of the main lineages and variants analyzed and their incidence on the population. Gamma is presumed to be the most lethal variant in the history of the pandemic so far in the country. In all regions, it was the apparent cause of the second wave and the alarming numbers of cases and deaths. The discrepancy in the counts and ratios of the variants among the different Brazilian territories is also evident. Brazil has continental magnitude and this allows the introduction of new lineages in different regions and at different times. In addition, the country has a large genetic, cultural and geographic diversity, contributing to the heterogeneity of the unveiled results. A fact that draws attention - and may be understood as a limitation of the analysis - is the particularly low number of sequenced genomes in relation to the number of confirmed cases in the country. Thus, the presented results might not reflect the exact epidemiology of the pandemics in the country.

However, this study is still a remarkably rich source of information and may help guide the development of non-pharmaceutical interventions and possible solutions to protect the population and prevent new pandemics related to infectious diseases.

REFERENCES

World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 on 11 March 2020. Available at: 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 on May 17th, 2021

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. **Lancet Infectious Diseases**. 2020 May;20(5):533-534.

Brazilian Health Ministry. Available at: https://covid.saude.gov.br/ - Accessed on March 18th, 2021

Candido, D.; Claro, I.M.; de Jesus, J.G.; Souza,W.M.; Moreira, F.R.R.; Dellicour, S.; Mellan, T.A.; du Plessis, L.; Pereira, R.H.M.; Sales, F.C.S.; et al. Evolution and Epidemic Spread of SARS-CoV-2 in Brazil. **Science.** 2020, 369, 1255–1260.

Mullen, J.L.; Tsueng, G.; Latif, A.A.; Alkuzweny, M.; Cano, M.; Haag, E.; Zhou, J.; Zeller, M.; Matteson, N.; Andersen, K.G.; *et al*. 2020. Available online: https://outbreak.info - Accessed on July 19th, 2021.

Andrew Rambaut, Nick Loman, Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, Tom Connor, Tom Peacock, David L Robertson, Erik Volz. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Available at https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 - Accessed on November 15th, 2021.

Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenço J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T. Detection of a SARS-CoV-2 variant of concern in South Africa. **Nature**. 2021 Apr;592(7854):438-443.

Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, Crispim MAE, Sales FCS, Hawryluk I, McCrone JT, Hulswit RJG, Franco LAM, Ramundo MS, de Jesus JG, Andrade PS, Coletti TM, Ferreira GM, Silva CAM, Manuli ER, Pereira RHM, Peixoto PS, Kraemer MUG, Gaburo N Jr, Camilo CDC, Hoeltgebaum H, Souza WM, Rocha EC, de Souza LM, de Pinho MC, Araujo LJT, Malta FSV, de Lima AB, Silva JDP, Zauli DAG, Ferreira ACS, Schnekenberg RP, Laydon DJ, Walker PGT, Schlüter HM, Dos Santos ALP, Vidal MS, Del Caro VS, Filho RMF, Dos Santos HM, Aguiar RS, Proença-Modena JL, Nelson B, Hay JA, Monod M, Miscouridou X, Coupland H, Sonabend R, Vollmer M, Gandy A, Prete CA Jr, Nascimento VH, Suchard MA, Bowden TA, Pond SLK, Wu CH, Ratmann O, Ferguson NM, Dye C, Loman NJ, Lemey P, Rambaut A, Fraiji NA, Carvalho MDPSS, Pybus OG, Flaxman S, Bhatt S, Sabino EC. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. **Science**. 2021 May;372(6544):815-821.

Naveca FG, Nascimento V, de Souza VC, Corado AL, Nascimento F, Silva G, Costa Á, Duarte D, Pessoa K, Mejía M, Brandão MJ, Jesus M, Gonçalves L, da Costa CF, Sampaio V, Barros D, Silva M, Mattos T, Pontes G, Abdalla L, Santos JH, Arantes I, Dezordi FZ, Siqueira MM, Wallau GL, Resende PC, Delatorre E, Gräf T, Bello G. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. **Nat Med.** 2021 Jul;27(7):1230-1238.

Cherian S, Potdar V, Jadhav S, Yadav P, Gupta N, Das M, Rakshit P, Singh S, Abraham P, Panda S, Team N. SARS-CoV-2 Spike Mutations, L452R, T478K,

E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. **Microorganisms**. 2021 Jul 20;9(7):1542.

Shu Y. McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. **Euro Surveillance**. Vol 22(13):30494. Mar 2020

Buss LF, Prete CA Jr, Abrahim CMM, Mendrone A Jr, Salomon T, de Almeida-Neto C, França RFO, Belotti MC, Carvalho MPSS, Costa AG, Crispim MAE, Ferreira SC, Fraiji NA, Gurzenda S, Whittaker C, Kamaura LT, Takecian PL, da Silva Peixoto P, Oikawa MK, Nishiya AS, Rocha V, Salles NA, de Souza Santos AA, da Silva MA, Custer B, Parag KV, Barral-Netto M, Kraemer MUG, Pereira RHM, Pybus OG, Busch MP, Castro MC, Dye C, Nascimento VH, Faria NR, Sabino EC. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. **Science**. 2021 Jan 15;371(6526):288-292.

Starr, T.N.; Greaney, A.J.; Dingens, A.S.; Bloom, J.D. Complete Map of SARS-CoV-2 RBD Mutations That Escape the Monoclonal Antibody LY-CoV555 and Its Cocktail with LY-CoV016. **Cell Rep. Med**. 2021, 2, 100255.

Koshy, J. Coronavirus|Indian 'Double Mutant' Strain Named B.1.617. The Hindu. 2021. Available online: https://www.thehindu.com/news/national/indian-double-mutant-strain-named-b1617/article34274663.ece - Accessed on October 9th, 2021.

Haseltine, W. An Indian SARS-CoV-2 Variant Lands in California. More Danger Ahead? Forbes, 12 April 2021. Available online: https://www.forbes.com/sites/williamhaseltine/2021/04/12/an-indian-sars-cov-2-variant-lands-in-california-more-danger-ahead/?sh=6f92503b3b29 - Accessed on October 9th, 2021).

Bertram, S.; Dijkman, R.; Habjan, M.; Heurich, A.; Gierer, S.; Glowacka, I.; Welsch, K.; Winkler, M.; Schneider, H.; Hofmann-Winkler, H.; et al. TMPRSS2 Activates the Human Coronavirus 229E for Cathepsin-Independent Host Cell Entry and Is Expressed in Viral Target Cells in the Respiratory Epithelium. **J. Virol**. 2013, 87, 6150–6160.

Shiehzadegan S, Alaghemand N, Fox M, Venketaraman V. Analysis of the Delta Variant B.1.617.2 COVID-19. **Clin Pract**. 2021 Oct 21;11(4):778-784.

Han P, Li L, Liu S, Wang Q, Zhang D, Xu Z, Han P, Li X, Peng Q, Su C, Huang B, Li D, Zhang R, Tian M, Fu L, Gao Y, Zhao X, Liu K, Qi J, Gao GF, Wang P. Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. **Cell**. 2022 Feb. 185(4):630-640.e10.

Callaway E. Heavily mutated Omicron variant puts scientists on alert. **Nature**. 2021 Dec;600(7887):21.

Torjesen I. Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear. **BMJ**. 2021 Nov;375:n2943.

Fontanet A, Cauchemez S. COVID-19 herd immunity: where are we? **Nat Rev Immunol.** 2020; 20: 583–84.

Sabino EC, Buss LF, Carvalho MPS, Prete CA Jr, Crispim MAE, Fraiji NA, Pereira RHM, Parag KV, da Silva Peixoto P, Kraemer MUG, Oikawa MK, Salomon T, Cucunuba ZM, Castro MC, de Souza Santos AA, Nascimento VH, Pereira HS, Ferguson NM, Pybus OG, Kucharski A, Busch MP, Dye C, Faria NR. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. **Lancet**. 2021 Feb 6;397(10273):452-455.

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

Fujino, T. et al. Novel SARS-CoV-2 variant identified in travelers from Brazil to Japan. **Emerg. Infect. Dis**. 2021

Naveca FG, Nascimento V, de Souza VC, Corado AL, Nascimento F, Silva G, Costa Á, Duarte D, Pessoa K, Mejía M, Brandão MJ, Jesus M, Gonçalves L, da Costa CF, Sampaio V, Barros D, Silva M, Mattos T, Pontes G, Abdalla L, Santos JH, Arantes I, Dezordi FZ, Siqueira MM, Wallau GL, Resende PC, Delatorre E, Gräf T, Bello G. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. **Nat Med.** 2021 Jul;27(7):1230-1238.

Empresa Brasil de Comunicação, Agência Brasil. Vaccination against COVID-19 starts nationwide. Jan, 2021. Available at: https://agenciabrasil.ebc.com.br/saude/noticia/2021-01/vacinacao-contra-covid-19-come%C3%A7a-em-todo-o-pais - Accessed on April 5th, 2022.

Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Available at: https://ourworldindata.org/covid-vaccinations?country=BRA - Accessed on February 7th, 2022.

aTaylor, L. Covid-19: Brazil sees omicron cases soar but data blackout obscures true impact. **The British Medical Journal**. 2022; 376 :o133

Brazilian Health Ministry, vaccination data. Available at: https://www.gov.br/saude/pt-br/vacinacao/ - Accessed on December 9th, 2021.

bTaylor, L. Covid-19: Omicron drives weekly record high in global infections. **The British Medical Journal**. 2022; 376 :o66

Reuters, 2022. Available at:

https://www.reuters.com/world/americas/brazil-reels-omicron-spreads-weighing-hospitals-economy-2022-01-14/ - Accessed on February 7th, 2022.

Maisa A, Spaccaferri G, Fournier L, Schaeffer J, Deniau J, Rolland P, Coignard B; regional COVID-19 investigation team; EMERGEN consortium. First cases of Omicron in France are exhibiting mild symptoms, November 2021-January 2022. **Infect Dis Now**. 2022 Feb 12:S2666-9919(22)00036-7.

Maslo C, Friedland R, Toubkin M, Laubscher A, Akaloo T, Kama B. Characteristics and Outcomes of Hospitalized Patients in South Africa During the COVID-19 Omicron Wave Compared With Previous Waves. **JAMA**. 2022 Feb 8;327(6):583-584.

United Kingdom Health Security Agency. SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 34. Jan 2022. Ref: UKHSA publications gateway number GOV-10924. Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1050236/technical-briefing-34-14-january-2022.pdf - Accessed on April 1st, 2022

Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. **J Clin Med**. 2020 Apr 24;9(4):1225

SUPPLEMENTARY

Supplementary Table 1. Population estimates, area and population density of each Brazilian state and their respective regions.

Region	State	State Acronym	Population		Area (km²)		Population density (people/km²)	
			Region	State	Region	State	Region	State
North	Acre	AC	18,906,962	906,876	3,850,509.94	164,123.96	4.9	5.5
	Amazonas	AM		4,269,995		1,559,167.89		2.7
	Amapá	AP		877,613		142,470.76		6.2
	Pará	PA		8,777,124		1,245,870.80		7.0
	Rondônia	RO		1,815,278		237,765.24		7.6
	Roraima	RR		652,713		223,644.53		2.9
	Tocantins	TO		1,607,363		277,466.76		5.8
Northeast	Alagoas	AL	57,667,842	3,365,351	1,552,167.01	27,843.30	37.2	120.9
	Bahia	BA		14,985,284		564,760.43		26.5
	Ceará	CE		9,240,580		148,894.44		62.1
	Maranhão	MA		7,153,262		329,642.18		21.7
	Paraíba	PB		4,059,905		56,467.24		71.9
	Pernambuco	PE		9,674,793		98,067.88		98.7
	Piauí	PI		3,289,290		251,756.52		13.1
	Rio Grande do Norte	RN		3,560,903		52,809.60		67.4
	Sergipe	SE		2,338,474		21,925.42		106.7
Midwest	Distrito Federal	DF	16,707,336	3,094,325	1,606,316.67	5,760.78	10.4	537.1
	Goiás	GO		7,206,589		340,203.33		21.2
	Mato Grosso do Sul	MS		2,839,188		357,145.53		7.9
	Mato Grosso	MT		3,567,234		903,207.02		3.9
Southeast	Espírito Santo	ES	89,632,912	4,108,508	924,565.48	46,074.45	96.9	89.2
	Minas Gerais	MG		21,411,923		586,521.12		36.5
	Rio de Janeiro	RJ		17,463,349		43,750.43		399.2
	São Paulo	SP		46,649,132		248,219.48		187.9
South	Paraná	PR	30,402,587	11,597,484	576,736.82	199,298.98	52.7	58.2
	Rio Grande do Sul	RS		11,466,630		281,707.16		40.7
	Santa Catarina	SC		7,338,473		95,730.68		76.7
BRAZIL			**213,317,639**		**8,510,295.91**		**25.1**	

Supplementary Table 2. Proportions of the main lineages and variants per month in the whole country.

	Delta	Gamma	Alpha	P.2	B.1.1.28	B.1.1.33	B.1.1	Others
Feb-20							37.50%	62.50%
Mar-20					34.41%	31.76%	10.59%	23.24%
Apr-20				0.23%	34.44%	46.41%	9.35%	9.58%
May-20				0.60%	29.61%	48.94%	10.27%	10.57%
Jun-20				0.23%	30.45%	35.68%	26.59%	7.05%
Jul-20				0.23%	32.72%	41.42%	17.62%	8.01%
Aug-20		0.49%		0.73%	30.56%	38.88%	15.89%	13.45%
Sep-20		0.49%		1.72%	36.86%	39.56%	16.22%	5.16%
Oct-20		0.72%		16.67%	35.02%	32.37%	10.39%	4.83%
Nov-20		0.71%		27.52%	31.06%	18.30%	5.82%	16.60%
Dec-20		12.33%	0.23%	50.57%	14.27%	7.19%	2.74%	12.67%
Jan-21		24.89%	3.07%	50.34%	8.92%	3.81%	0.91%	8.07%
Feb-21		59.02%	2.32%	28.82%	4.72%	0.55%	0.39%	4.17%
Mar-21	0.02%	85.33%	4.56%	5.83%	1.58%	0.15%	0.19%	2.34%
Apr-21	0.08%	93.77%	2.95%	0.96%	0.34%		0.08%	1.83%
May-21	0.17%	96.01%	2.25%	0.07%	0.30%		0.01%	1.20%
Jun-21	1.17%	96.31%	1.59%	0.01%	0.02%		0.02%	0.89%
Jul-21	11.45%	87.57%	0.51%	0.03%	0.03%		0.01%	0.40%
Aug-21	62.15%	37.44%	0.17%		0.01%			0.22%
Sep-21	92.43%	7.11%						0.46%
Oct-21	98.23%	1.15%						0.62%
Nov-21	99.77%	0.06%						0.17%

Supplementary Table 3. Cases per 100 thousand inhabitants of the regions per month.

Month-Year	North	Northeast	Midwest	Southeast	South
Mar-20	1.6	1.6	2.8	3.8	2.4
Apr-20	56.2	42.5	13.7	43.8	15.2
May-20	513.3	268.4	88.6	161.3	58.5
Jun-20	836.1	529.3	488.1	333.2	174.5
Jul-20	762.2	652.7	923.2	482.3	495.7
Aug-20	684.7	492.7	1104.7	488.2	686.2
Sep-20	492.6	310.1	956.8	358.4	491.6
Oct-20	368.6	268.0	582.8	294.2	444.2
Nov-20	362.1	268.2	437.1	291.8	810.7
Dec-20	469.2	458.5	645.5	541.9	1290.5
Jan-21	867.7	485.1	779.4	712.7	1034.8
Feb-21	763.7	516.9	713.8	560.9	937.8
Mar-21	992.6	780.8	1395.1	886.3	1775.7
Apr-21	747.6	791.3	1100.2	912.6	1026.8
May-21	606.0	880.6	1043.3	836.1	1107.3
Jun-21	562.0	876.0	1075.9	910.1	1333.8
Jul-21	359.5	417.6	857.3	660.4	1012.7
Aug-21	208.2	174.6	835.1	469.7	534.5
Sep-21	107.4	104.0	544.9	399.6	387.2
Oct-21	99.5	86.2	439.4	158.7	322.4
Nov-21	139.7	107.1	252.4	101.4	211.4
Dec-21	47.6	25.6	45.8	20.2	44.5

Supplementary Table 4. Deaths per 100 thousand inhabitants of the regions per month

Month-Year	North	Northeast	Midwaset	Southeast	South
Mar-20	0.02	0.04	0.0	0.2	0.03
Apr-20	3.83	2.69	0.4	3.6	0.61
May-20	26.24	12.72	1.8	11.6	1.18
Jun-20	20.69	18.08	8.5	15.2	3.45
Jul-20	12.10	16.36	21.5	16.0	10.83
Aug-20	8.66	11.06	24.0	14.3	13.44
Sep-20	7.54	7.03	20.0	11.6	10.13
Oct-20	5.45	5.17	13.0	8.7	6.79
Nov-20	4.32	4.02	8.0	6.9	8.56
Dec-20	6.59	5.65	10.2	11.4	17.66
Jan-21	21.90	6.59	12.9	16.2	16.25
Feb-21	26.12	8.72	14.4	15.4	14.41
Mar-21	29.12	21.88	43.4	28.6	52.36
Apr-21	28.50	24.64	51.7	45.3	44.55
May-21	16.58	20.65	31.3	32.8	29.63
Jun-21	12.35	19.21	30.7	30.8	29.87
Jul-21	8.20	10.27	20.3	22.0	24.92
Aug-21	5.25	5.31	19.5	14.4	12.69
Sep-21	2.61	2.58	11.6	10.6	9.31
Oct-21	1.62	2.07	7.8	6.5	8.01
Nov-21	1.83	1.82	3.7	4.1	4.08
Dec-21	0.50	0.39	0.5	0.8	0.62

Supplementary Table 5. Cases per 100 thousand inhabitants of the states per month

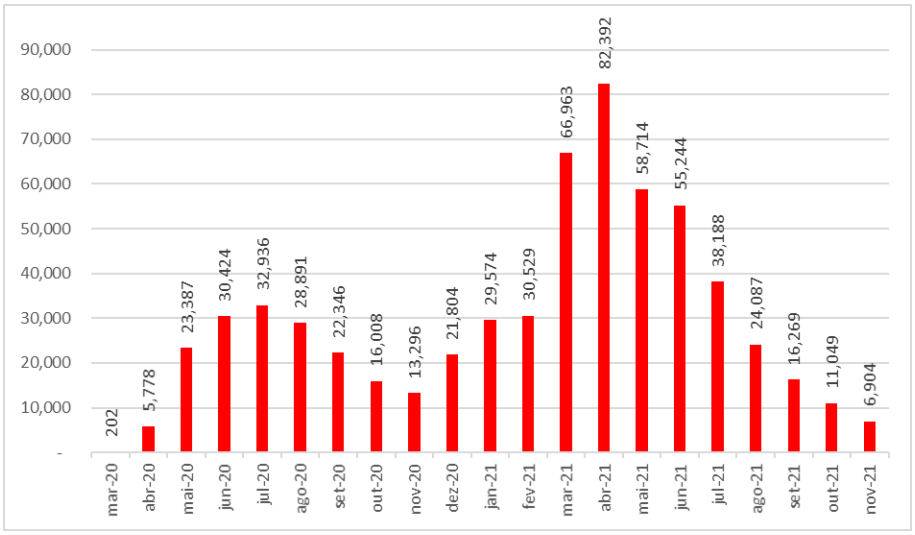
		Mar 20	Apr 20	May 20	Jun 20	Jul 20	Aug 20	Sep 20	Oct 20	Nov 20	Dec 20	Jan 21	Feb 21	Mar 21	Apr2 1	May 21	Jun 21	Jul 21	Aug 21	Sep 21	Oct 21	Nov 21
North	AC	5	40	641	776	703	554	394	284	602	591	755	1000	1337	898	521	334	175	75	11	14	18
	AM	4	119	846	690	705	453	446	524	387	534	1555	1140	774	507	366	367	342	177	51	31	50
	AP	1	122	971	2152	909	766	577	447	793	1033	1007	755	1581	936	726	577	459	138	47	90	116
	PA	0	34	398	774	572	512	350	243	207	257	407	409	626	599	513	418	203	135	82	78	125
	RO	0	27	245	898	977	890	593	317	489	837	1594	1328	2121	1403	927	994	564	296	147	247	386
	RR	2	77	486	1756	2692	1683	1068	1030	945	787	851	1205	1184	1030	1112	1349	1146	593	342	179	156
	TO	1	9	250	438	879	1635	1056	437	395	533	735	754	1708	1153	1166	1179	759	548	342	254	253
Northeast	AL	1	31	275	763	706	567	240	116	125	292	385	415	646	599	576	703	394	173	72	57	41
	BA	1	18	104	366	620	604	359	284	333	603	632	640	798	648	743	753	458	181	87	83	94
	CE	4	81	440	671	695	436	284	357	293	376	407	572	1223	1463	1375	920	351	134	96	35	66
	MA	1	48	444	631	562	433	307	171	103	109	92	165	323	343	327	380	275	167	100	68	56
	PB	0	22	301	832	883	566	382	293	298	523	618	727	937	825	943	1613	631	278	206	107	359
	PE	1	70	285	252	374	316	224	163	202	411	404	395	514	593	781	704	419	173	136	115	81
	PI	1	18	137	515	924	772	577	508	430	471	490	447	992	1041	1000	688	409	208	82	195	176
	RN	2	34	188	673	554	295	218	324	409	643	637	759	818	760	1258	2069	461	152	99	140	233
	SE	1	19	280	787	1424	591	210	294	249	956	1065	598	992	1162	1502	1138	460	137	28	17	13
Midwest	DF	11	35	270	1275	1844	1802	976	679	514	729	821	633	1541	1112	876	794	637	656	801	636	91
	GO	1	10	41	294	615	896	1075	616	336	405	578	627	1256	917	807	928	905	1065	628	561	432
	MS	2	7	46	250	601	848	719	436	618	1219	938	726	1250	1112	1537	1529	718	438	148	120	84
	MT	1	8	61	387	1003	1127	891	549	430	602	1025	949	1667	1451	1273	1258	1063	841	471	276	164
Southeast	ES	2	63	268	808	886	681	492	590	834	1418	1117	783	1367	1319	1106	873	609	487	562	513	322
	MG	1	7	40	161	383	418	367	298	268	591	895	674	1145	1099	994	1082	760	466	346	210	107
	RJ	4	50	252	339	303	333	236	259	254	460	494	355	371	543	703	532	422	552	902	208	125
	SP	5	57	174	368	559	562	389	280	269	473	675	566	918	930	790	976	709	439	222	86	71
South	PR	2	11	29	156	459	481	405	298	574	1177	1145	842	1717	876	1257	1650	812	681	462	377	204
	RS	3	11	68	154	347	516	565	497	650	1113	851	817	1783	1130	964	1134	1329	357	245	261	212
	SC	3	29	91	236	787	1277	514	592	1436	1747	1148	1278	1858	1104	1094	1148	835	581	491	331	221

.

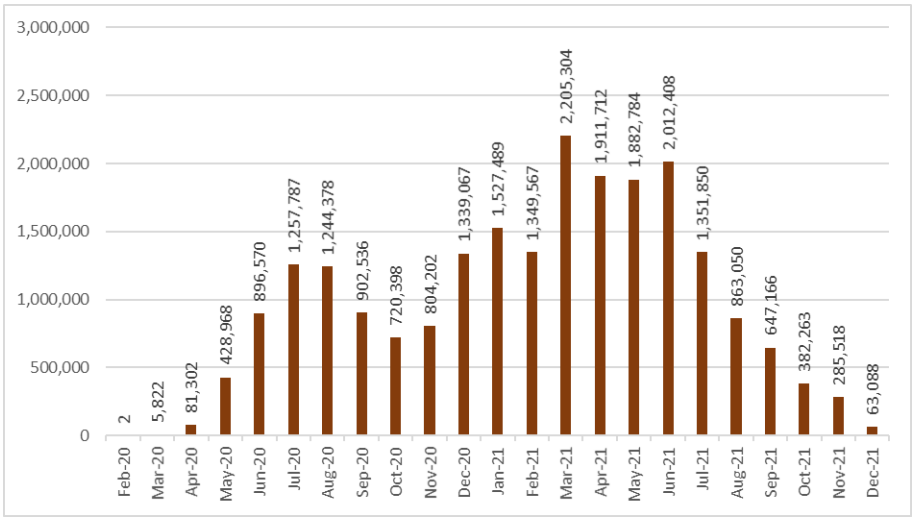
Supplementary Table 6. Deaths per 100 thousand inhabitants of the states per month.

		Mar 20	Apr 20	May 20	Jun 20	Jul 20	Aug 20	Sep 20	Oct 20	Nov 20	Dec 20	Jan 21	Feb 21	Mar 21	Apr 21	May 21	Jun 21	Jul 21	Aug 21	Sep 21	Oct 21	Nov 21
North	AC	0.0	2.1	14.2	23.9	18.3	8.9	5.2	3.7	3.3	7.9	7.9	14.4	29.1	29.4	14.7	8.4	6.7	1.7	2.6	0.8	0.3
	AM	0.1	9.9	38.1	18.1	10.4	8.9	9.2	11.7	8.2	9.2	66.3	64.2	27.0	14.3	8.4	7.7	5.0	3.7	0.8	1.1	0.8
	AP	0.0	3.9	21.4	22.2	16.9	10.9	5.5	4.4	6.7	13.4	15.3	9.2	18.5	27.5	17.4	15.8	8.2	5.2	3.2	1.1	1.4
	PA	0.0	2.6	30.8	23.2	9.0	4.9	4.6	1.9	1.9	3.2	5.2	11.3	21.2	29.4	16.4	11.0	6.4	4.6	2.3	1.0	1.9
	RO	0.1	0.8	7.7	19.9	19.5	15.2	11.5	5.5	5.7	14.1	23.5	33.4	71.2	56.2	32.8	20.8	12.8	6.1	2.8	2.3	3.9
	RR	0.0	1.1	16.7	30.3	30.5	12.3	9.3	5.8	5.5	9.0	10.6	37.4	37.7	25.6	19.3	16.5	16.9	13.9	8.1	4.3	3.4
	TO	0.0	0.2	4.4	8.1	11.6	18.5	16.2	9.3	4.4	4.3	9.3	9.1	32.7	31.2	20.5	21.0	18.4	10.3	6.3	5.5	2.5
Northeast	AL	0.0	1.4	11.8	18.1	15.3	9.5	5.5	4.9	3.0	4.5	7.6	7.5	16.5	19.8	15.8	17.5	13.9	7.8	4.2	2.3	1.8
	BA	0.0	0.7	3.7	7.9	10.7	12.9	9.0	5.9	4.3	5.7	6.5	11.5	23.4	21.0	18.4	18.5	11.6	4.9	2.5	1.5	1.5
	CE	0.1	5.2	27.2	33.9	16.7	7.8	6.3	3.9	3.0	3.9	5.2	8.8	31.0	38.6	29.9	23.0	9.9	5.5	2.1	3.0	1.8
	MA	0.0	2.8	10.8	15.0	13.5	6.1	4.3	4.2	3.4	2.9	2.8	5.0	14.2	16.9	12.1	12.6	8.3	5.5	2.0	0.9	0.8
	PB	0.0	1.6	7.2	15.2	20.5	15.7	9.2	6.8	4.8	9.3	9.5	10.8	30.7	25.9	21.6	23.0	9.4	4.9	3.1	2.7	2.6
	PE	0.1	5.8	23.2	20.9	17.9	10.7	6.8	3.9	4.2	6.4	7.2	6.7	12.2	19.3	18.9	18.8	11.4	6.2	3.7	2.9	2.3
	PI	0.1	0.6	4.4	15.7	20.3	14.9	8.6	8.4	7.0	6.3	6.8	8.7	24.0	29.9	25.2	20.4	6.5	3.3	2.0	2.7	2.8
	RN	0.0	1.5	7.5	21.0	22.8	10.9	3.7	5.1	3.4	8.5	8.4	9.2	25.7	26.9	18.1	18.2	8.4	4.8	2.0	1.7	2.8
	SE	0.0	0.6	6.2	22.2	32.4	18.1	7.7	7.3	4.4	7.5	12.6	7.8	23.1	33.1	34.9	26.0	9.1	3.5	0.8	0.9	0.6
Midwest	DF	0.1	0.9	4.5	13.5	28.5	34.0	23.7	13.8	8.0	10.6	9.5	9.2	38.5	57.2	28.2	18.8	11.9	14.0	13.3	13.3	5.1
	GO	0.0	0.4	1.3	5.1	16.5	20.6	21.5	14.6	8.6	6.0	9.4	14.3	44.3	46.4	28.9	29.3	21.5	23.5	14.1	9.5	4.8
	MS	0.0	0.3	0.4	2.3	10.7	17.6	14.9	10.2	6.3	20.0	20.2	14.4	35.4	50.2	39.0	48.0	25.5	15.0	6.4	2.9	1.4
	MT	0.0	0.3	1.5	15.9	34.0	27.2	18.1	11.1	8.1	10.6	17.1	19.0	52.4	59.0	32.7	30.1	21.1	19.7	9.2	3.5	2.0
Southeast	ES	0.0	2.3	12.4	25.4	21.8	14.9	9.4	7.5	10.3	19.5	18.9	13.3	26.7	49.3	31.0	16.6	10.0	8.9	7.4	8.8	6.4
	MG	0.0	0.4	0.9	3.2	8.4	12.0	9.5	7.7	4.8	8.7	14.7	16.4	26.9	43.7	31.7	26.8	19.7	11.8	7.3	4.8	2.9
	RJ	0.1	4.8	25.7	27.1	19.5	14.8	13.9	12.1	11.4	16.8	24.5	18.7	20.9	44.0	35.4	28.0	21.8	18.3	21.1	12.7	4.0
	SP	0.3	4.8	11.2	15.3	17.7	15.0	12.0	7.9	6.0	9.9	13.5	13.8	32.5	46.2	32.5	35.0	24.3	14.7	8.5	4.7	4.4
South	PR	0.0	0.7	0.9	4.0	11.0	11.8	10.3	6.3	8.1	15.7	17.7	14.1	43.5	49.0	34.7	37.0	39.0	19.5	13.8	12.2	2.5
	RS	0.0	0.5	1.4	3.4	11.0	13.6	11.7	8.9	8.9	18.0	15.7	15.0	64.0	45.5	28.3	28.0	16.9	7.5	5.8	5.5	5.4
	SC	0.0	0.6	1.3	2.7	10.4	15.8	7.3	4.3	8.8	20.3	14.8	13.9	48.1	36.1	23.7	21.6	15.2	9.9	7.8	5.3	4.5

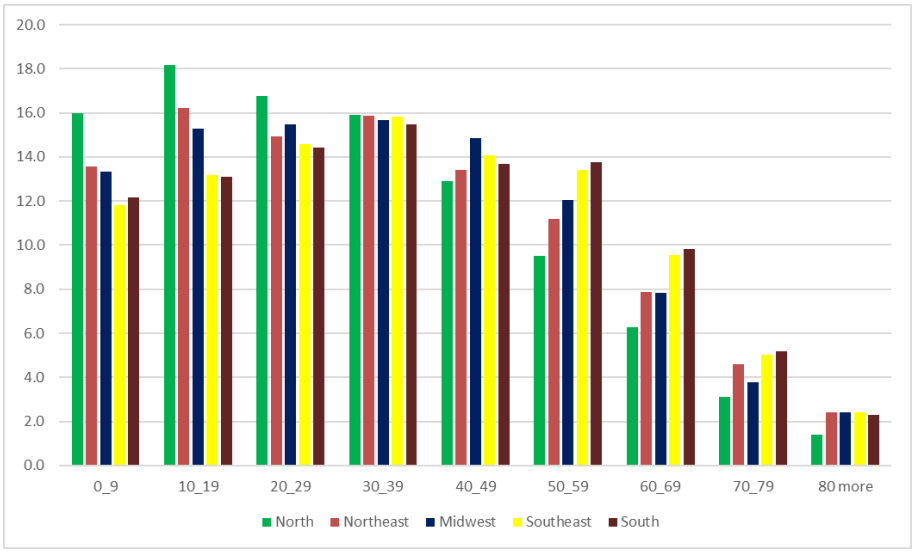
Supplementary Figure 1. Total deaths in each month in the country.



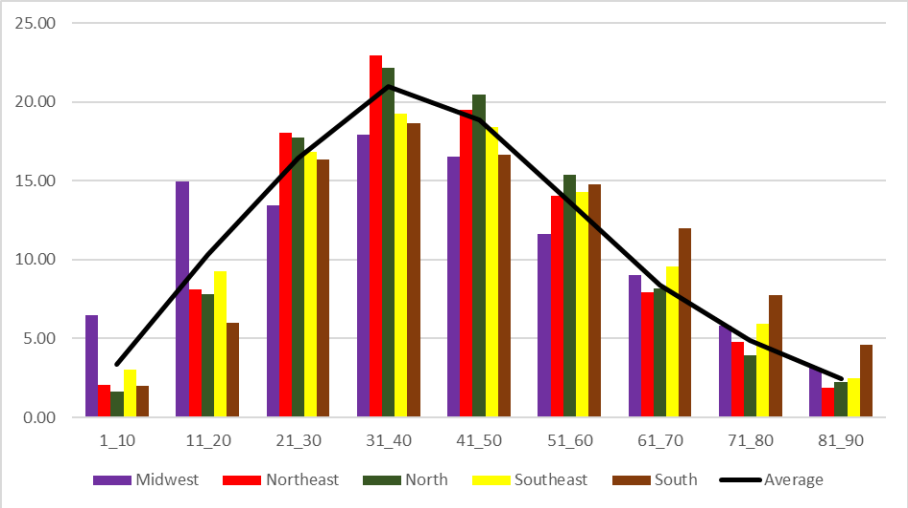
Supplementary Figure 2. Total of cases in each month in the country.



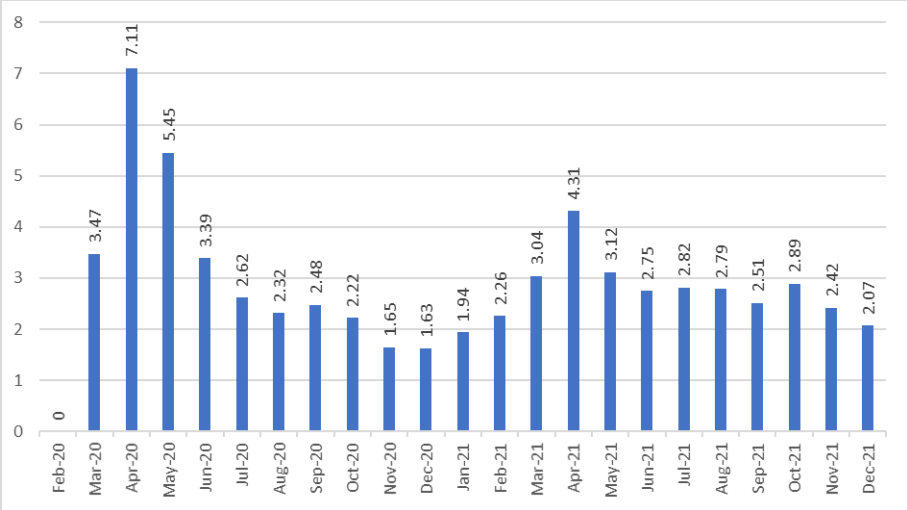
Supplementary Figure 3. Proportion of age ranges in every Brazilian region.



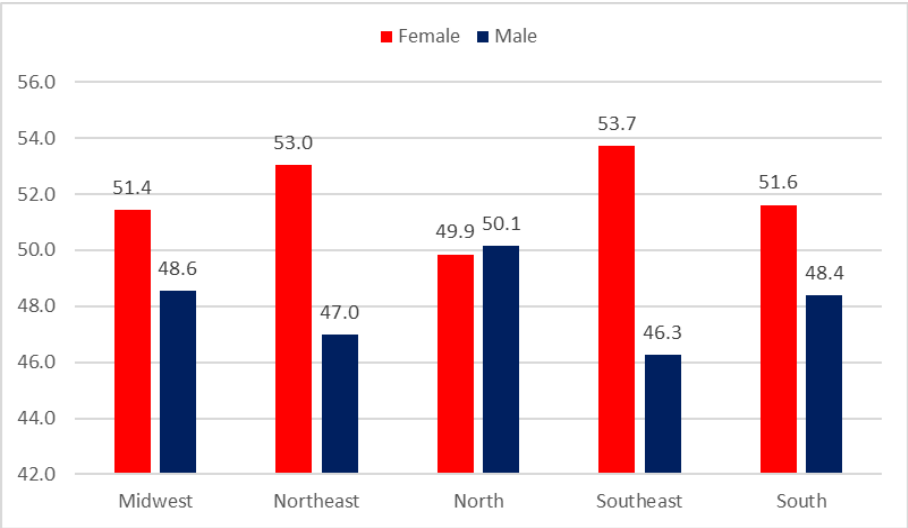
Supplementary Figure 4. Percentage of sequenced genomes in each age range in every Brazilian region.



Supplementary Figure 5. Case-fatality rate on every month of the studied period.



Supplementary Figure 6. Proportion of male and female patients in the Brazilian regions.



APÊNDICE C Artigo 3 - Population-based prevalence surveys during the Covid-19 pandemic: A systematic review

*Artigo publicado em coautoria.

Revista

Reviews in Medical Virology (JCR 6,989 e Qualis A1). 2020. DOI: https://doi.org/10.1002/rmv.2200

Título

Population-based prevalence surveys during the Covid-19 pandemic: A systematic review

Abstract

Population-based prevalence surveys of Covid-19 contribute to establish the burden of infection, the role of asymptomatic and mild infections in transmission, and allow more precise decisions about reopen policies. We performed a systematic review to evaluate qualitative aspects of these studies, assessing their reliability and compiling practices that can influence the methodological quality. We searched MEDLINE, EMBASE, bioRxiv and medRxiv, and included cross-sectional studies using molecular and/or serological tests to estimate the prevalence of Covid-19 in the general population. Survey quality was assessed using the Joanna Briggs Institute Critical Appraisal Checklist for Prevalence Studies. A correspondence analysis correlated methodological parameters of each study to identify patterns related to higher, intermediate and lower risks of bias. The available data described 37 surveys from 19 countries. The majority were from Europe and America, used antibody testing, and reached highly heterogeneous sample sizes and prevalence estimates. Minority communities were disproportionately affected by Covid-19. Important risk of bias was detected in four domains: sample size, data analysis with suficiente coverage,

measurements in standard way and response rate. The correspondence analysis showed few consistent patterns for high risk of bias. Intermediate risk of bias was related to American and European studies, municipal and regional initiatives, blood samples and prevalence >1%. Low risk of bias was related to Asian studies, nationwide initiatives, reverse-transcriptase polymerase chain reaction tests and prevalence <1%. We identified methodological standards applied worldwide in Covid-19 prevalence surveys, which may assist researchers with the planning, execution and reporting of future population-based surveys.

APÊNDICE D Artigo 4 - Genomic epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil

*Artigo publicado em coautoria.

Revista

BMC Genomics (JCR 3,969 e Qualis A2). 2021. DOI: https://doi.org/10.1186/s12864-021-07708-w

Título

Genomic epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil

Abstract

Background: Brazil is the third country most affected by Coronavirus disease-2019 (COVID-19), but viral evolution in municipality resolution is still poorly understood in Brazil and it is crucial to understand the epidemiology of viral spread. We aimed to track molecular evolution and spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Esteio (Southern Brazil) using phylogenetics and phylodynamics inferences from 21 new genomes in global and regional context. Importantly, the case fatality rate (CFR) in Esteio (3.26%) is slightly higher compared to the Rio Grande do Sul (RS) state (2.56%) and the entire Brazil (2.74%).

Results: We provided a comprehensive view of mutations from a representative sampling from May to October 2020, highlighting two frequent mutations in spike glycoprotein (D614G and V1176F), an emergent mutation (E484K) in spike Receptor Binding Domain (RBD) characteristic of the B.1.351 and P.1 lineages, and the adjacent replacement of 2 amino acids in Nucleocapsid phosphoprotein (R203K and G204R). E484K was found in two genomes from mid-October, which is the earliest description of this mutation in Southern Brazil. Lineages containing

this substitution must be subject of intense surveillance due to its association with immune evasion. We also found two epidemiologically related clusters, including one from patients of the same neighborhood. Phylogenetics and phylodynamics analysis demonstrates multiple introductions of the Brazilian most prevalent lineages (B.1.1.33 and B.1.1.248) and the establishment of Brazilian lineages ignited from the Southeast to other Brazilian regions.

Conclusions: Our data show the value of correlating clinical, epidemiological and genomic information for the understanding of viral evolution and its spatial distribution over time. This is of paramount importance to better inform policy making strategies to fight COVID-19.

APÊNDICE E Artigo 5 - E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil

*Artigo publicado em coautoria.

Revista

Infection, Genetics and Evolution (JCR 3,342 e Qualis A2). 2021. DOI: https://doi.org/10.1016/j.meegid.2021.104941

Título

E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil

Abstract

The COVID-19 pandemic caused by SARS-CoV-2 has affected millions of people since its beginning in 2019. The propagation of new lineages and the discovery of key mechanisms adopted by the virus to overlap the immune system are central topics for the entire public health policies, research and disease management. Since the second semester of 2020, the mutation E484K has been progressively found in the Brazilian territory, composing different lineages over time. It brought multiple concerns related to the risk of reinfection and the effectiveness of new preventive and treatment strategies due to the possibility of escaping from neutralizing antibodies. To better characterize the current scenario we performed genomic and phylogenetic analyses of the E484K mutated genomes sequenced from Brazilian samples in 2020. From October 2020, more than 40% of the sequenced genomes present the E484K mutation, which was identified in three different lineages (P.1, P.2 and B.1.1.33 - posteriorly renamed as N.9) in four Brazilian regions. We also evaluated the presence of E484K associated mutations and identified selective pressures acting on the spike

protein, leading us to some insights about adaptive and purifying selection driving the virus evolution.