



**UNIVERSIDADE FEDERAL DO CEARÁ CENTRO DE CIÊNCIAS
COORDENAÇÃO DO CURSO DE BACHARELADO EM
BIOTECNOLOGIA**

PEDRO MIGUEL CARNEIRO JERONIMO

**TRABALHO DE CONCLUSÃO DE CURSO
INVESTIGAÇÃO DA HISTÓRIA EVOLUTIVA DA VARIANTE DE
PREOCUPAÇÃO GAMA DE SARS-COV-2 NO ESTADO DO CEARÁ**

FORTALEZA

2022

PEDRO MIGUEL CARNEIRO JERONIMO

INVESTIGAÇÃO DA HISTÓRIA EVOLUTIVA DA VARIANTE DE
PREOCUPAÇÃO GAMA DE SARS-COV-2 NO ESTADO DO CEARÁ

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Biotecnologia do Centro de Ciências da
Universidade Federal do Ceará, como
requisito parcial à obtenção do grau de
bacharel em Biotecnologia.

Orientador: Prof. Dr. Fábio Miyajima

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- J54i Jeronimo, Pedro Miguel Carneiro.
 Investigação da história evolutiva da variante de preocupação Gama de SARS-CoV-2 no estado do Ceará /
 Pedro Miguel Carneiro Jeronimo. – 2022.
 70 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências,
 Curso de Biotecnologia, Fortaleza, 2022.
 Orientação: Prof. Dr. Fábio Miyajima.
1. SARS-CoV-2. 2. Monitoramento. 3. Variante Gama. 4. Vigilância. 5. Epidemiologia molecular. I.
 Título.

CDD 661

PEDRO MIGUEL CARNEIRO JERONIMO

INVESTIGAÇÃO DA HISTÓRIA EVOLUTIVA DA VARIANTE DE
PREOCUPAÇÃO GAMA DE SARS-COV-2 NO ESTADO DO CEARÁ

Monografia apresentada ao curso de
Graduação em Biotecnologia da
Universidade Federal do Ceará, como
parte dos requisitos para obtenção do
título de Bacharel em Biotecnologia.

Aprovada em: _04 / _02 / _2022_.

BANCA EXAMINADORA

Prof. Dr. Fábio Miyajima (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Nicholas Costa Barroso Lima
Universidade Federal do Ceará (UFC)

Me. Filipe Zimmer Dezordi
Instituto Aggeu Magalhães

AGRADECIMENTOS

À fundação para o desenvolvimento científico e tecnológico em saúde – FIOTEC, pelo apoio financeiro com a manutenção da bolsa de auxílio.

À minha mãe, Ivone Carneiro Jeronimo, por me apoiar em cada decisão e por sempre acreditar em mim e no meu futuro.

Ao Prof. Dr. Fábio Miyajima, pela excelente orientação e dedicação.

Aos professores e pesquisadores participantes da banca examinadora Nicholas Costa Barroso lima e Filipe Zimmer Dezordi pelo tempo, pelas valiosas colaborações e sugestões.

Aos meus amigos que fiz durante a graduação. A presença de vocês fez toda a diferença na minha vida acadêmica e pessoal.

"Be kind to yourself so that you can be
happy enough to be kind to the world"

(Misha Collins)

RESUMO

Desde o surgimento do novo coronavírus (SARS-CoV-2) em Wuhan em dezembro de 2019 até a data deste trabalho, muito foi descoberto e documentado a respeito, inclusive acerca de seu processo evolutivo. Contudo, a falta de estudos epidemiológicos de vigilância laboratorial em diversas regiões do Brasil, como o estado do Ceará, constituem lacunas de conhecimento ainda a serem preenchidas. Investigações prospectivas e regionalizadas das Variantes de Preocupação (VOCs) permitem uma melhor compreensão da dinâmica viral e sua disseminação, possibilitando a formulação de estratégias direcionadas de prevenção e combate ao agente etiológico. Tendo esta necessidade em vista, o presente estudo tem por objetivo apresentar hipóteses e prospectos sobre a introdução, a ascensão e o declínio da VOC Gama (linhagem P.1) de SARS-CoV-2 no estado do Ceará durante o período de dezembro de 2020 a novembro de 2021. Para tanto, a metodologia empregada engloba análises por Máxima Verossimilhança Temporal, por inferência Bayesiana para a determinação do período de introdução da variante, bem como do seu Número Reprodutivo e de sua dispersão geográfica, além do uso de algoritmos internos para avaliar a variação do perfil mutacional do vírus ao longo do tempo. Foram elaborados diferentes *datasets* usando genomas da plataforma de referência internacional *GISAI*D e dados disponibilizados pelo Laboratório de Biologia Molecular e Inovação Diagnóstica (BIOMIND) do Centro de Hematologia e Hemoterapia do Ceará. As análises indicam que a primeira introdução da VOC Gama no Ceará ocorreu ainda em dezembro de 2020, com introduções posteriores provenientes principalmente da região Norte e Nordeste do País. Ademais, foi inferido que a cidade de Fortaleza foi o epicentro de exportação da linhagem P.1 e de suas sublinhagens para os demais municípios do estado. Por fim, o atual trabalho indica: i) ter havido um provável atraso na identificação do primeiro caso de VOC Gama no estado; ii) que a provável direção principal de disseminação da capital para o interior; e iii) os vários eventos de exportação do vírus de regiões brasileiras para o Ceará. Em suma, este trabalho reforça a importância da implementação de estratégias de vigilância molecular de vírus emergentes e monitoramento contínuo de suas linhagens e variantes para o direcionamento de medidas prevenção e contenção.

Palavras-chave: SARS-CoV-2, monitoramento, Variante Gama, Vigilância, Epidemiologia Molecular.

ABSTRACT

Since the emergence of the new coronavirus (SARS-CoV-2) in Wuhan in December 2019 until now, much has been uncovered and documented, including its evolutionary process. However, the lack of epidemiological studies of laboratory surveillance in several regions of Brazil, such as the state of Ceara, represents a knowledge gap yet to be filled. Prospective regionalized investigations of Variants of Concern (VOCs) allow a better understanding of the viral dynamics and dissemination, enabling the design of targeted strategies for tackling the etiological agent. This study aims to present hypotheses and prospects on the introduction, expansion and decline of the VOC Gamma (P.1 lineage) of SARS-CoV-2 in the state of Ceará during the period of December 2020 to November 2021. For this purpose, the methodology employed made use of temporal maximum-likelihood Estimation and Bayesian inference to determine the probable period of introduction of the variant, as well as its reproductive number, geographical dispersions, along with the use of in-house algorithms to assess the variation of the virus mutation profile over time. Different datasets were created using genomes from the international reference database GISAID and data made available by the Laboratory of Molecular Biology and Diagnostic Innovation (BIOMIND) at the Hematology and Hemotherapy Center of Ceara. The analyses indicate that the first introduction of VOC Gamma in Ceará occurred most likely in December 2020, with subsequent introductions coming particularly from North and Northeast great regions of the country. Furthermore, it was inferred that the State's capital, Fortaleza, was the epicenter for the introduction and spread of P.1-like lineages to the other municipalities of the state. Lastly, this work suggests that: i) it has probably occurred a delay in identifying the first VOC Gamma in the state; ii) the likely key direction of dissemination and spread was from the capital to the countryside; and iii) various concomitant introductions subsequently took place coming from various regions of the country. In summary, this work reinforces the relevance of the implementation of molecular surveillance strategies of emerging viruses and the continuous monitoring of their lineages and variants for the tailoring preventive and containment measures.

Keywords: SARS-CoV-2, Monitoring, VOC Gamma, Surveillance, Molecular Epidemiology.

SUMÁRIO

1. INTRODUÇÃO	10
2. REVISÃO BIBLIOGRÁFICA	11
2.1. Descrição molecular do SARS-CoV-2	11
2.1.1.NSP1	11
2.1.2.NSP2	12
2.1.3.NSP3	12
2.1.4.NSP4	12
2.1.5.NSP5	12
2.1.6.NSP6	13
2.1.7.NSP7, NSP8 e NSP12	13
2.1.8.NSP9	13
2.1.9.NSP10, NSP14 e NSP16	13
2.1.10.NSP11	14
2.1.11.NSP13	14
2.1.12.NSP15	14
2.1.13.ORF2 (<i>Spike</i>).....	14
2.1.14.ORF3a	15
2.1.15.ORF4 (Envelope).....	15
2.1.16.ORF5 (Membrana)	15
2.1.17.ORF6.....	15
2.1.18.ORF7a	16
2.1.19.ORF7b	16
2.1.20.ORF8.....	16
2.1.21.ORF9 (Nucleocapsídeo).....	16
2.1.22.ORF10	17
2.2. Transmissão do SARS-CoV-2	17
2.3. Medidas de prevenção não farmacológicas	18
2.4. Epidemiologia	19
2.5. Vacinas.....	20
2.6. Sequenciamento Genômico	21
2.6.1.Sanger	21
2.6.2.NGS - <i>Next-Generation Sequencing</i>	21
2.6.3.Controle de qualidade e <i>minor variants</i>	22
2.7. Árvores filogenéticas	26
3. OBJETIVOS	28
3.1. Objetivo geral	28
3.2. Objetivos específicos.....	28
4. METODOLOGIA	29
4.1. Obtenção e triagem das amostras.....	29
4.2. Sequenciamento.....	29
4.3. Montagem dos genomas	30

4.4. Automação de processos	30
4.5. Geração de mapas de profundidade	33
4.6. Número de óbitos	34
4.7. Prevalência relativa	34
4.8. Filogeografia e inferência Bayesiana.....	35
4.9. Estimação do número efetivo de reprodução (R0)	36
4.10. Análise de máxima verossimilhança temporal.....	37
4.11. Frequência de mutações e <i>Minor variants</i>	38
4.12. Dados associados	38
4.10.1 Número de voos por região	38
4.10.2. Índice de permanência domiciliar	39
4.10.3. Vacinação.....	39
5. RESULTADOS E DISCUSSÃO.....	40
5.1. Evolução de casos no Ceará.....	40
5.2. Surgimento da linhagem P.1 no Ceará.....	44
5.3. Distribuição geográfica	48
5.4. Sublinhagens da P.1	49
6. CONCLUSÃO.....	53
7. REFERÊNCIAS	54

1. INTRODUÇÃO

A síndrome respiratória aguda grave causada pelo coronavírus 2 (SARS-CoV-2) é uma doença que teve início na província de Wuhan na China e se espalhou rapidamente ao redor do mundo (YANG et al., 2020), apresentando um espectro clínico de pneumonia que varia de casos leves a graves. A capacidade do vírus de se disseminar através do contato físico e do ar, superior à de transmissão do *Middle East respiratory syndrome coronavirus* (MERS-CoV) e do *severe acute respiratory syndrome* (SARS) *coronavirus* (SARS-CoV) (SHENG, 2020), inclusive por assintomáticos, é fator que possibilitou o estabelecimento de uma pandemia. Com a disseminação a nível global, o primeiro caso da recém-descoberta SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) logo foi relatado no Brasil em 25 de fevereiro de 2020 (THE LANCET, 2020).

As novas variantes do vírus, como a Variante de Preocupação (em inglês, *variant of concern* = VOC) (OMS, 2021), se mostram hoje uma preocupação mundial, pela possibilidade de nova infecção mesmo dentre os pacientes já vacinados e por novos sintomas e quadros clínicos (ALTMANN; BOYTON; BEALE, 2021). O surgimento da linhagem Gama no início do mês de novembro de 2020, no estado do Amazonas, substituindo a linhagem B.1.1.28, indica sua alta transmissibilidade, até mesmo entre uma população já fortemente exposta a outras linhagens do vírus (NAVECA; SOUZA; COSTA, 2021). Além disso, a análise da heterogeneidade presente nas sequências, buscando por mutações definidoras de linhagens, como é o caso da Gama, torna possível a identificação de convergências evolutivas, e, portanto, facilita a formulação de medidas de combate ao vírus ao se notar o aumento da frequência de mutações relacionadas a linhagens mais transmissíveis.

Durante o ano de 2020, diversos trabalhos foram feitos com o intuito de descrever como o vírus se comportava e divergia nos estados e regiões do país (BOTELHO-SOUZA et al., 2021; NAVECA et al., 2021a; PERICO et al., 2021; SANTOS PAIVA et al., 2020). Contudo, este tipo de análise ainda não se encontra feita para o estado do Ceará, o qual sofreu com duas ondas de aumento de casos (KERR et al., 2021). Embora órgãos estatais tenham tomado providências com o fito de reduzir os impactos do vírus na vida das pessoas (CEARÁ, 2020), a descrição dos principais eventos associados a epidemia no estado continua uma incógnita. A identificação dos períodos e modos de introdução de SARS-CoV-2 podem ser usados não somente para futuros casos de vírus com transmissão semelhante, mas também dentro da

própria epidemia como modo de combater o surgimento e disseminação de novas Variantes de Preocupação.

Portanto, buscando preencher essa lacuna, o presente estudo tem por objetivo descrever como, de que modo e quando a Variante de Preocupação Gama se instalou no Ceará e por quais possíveis motivos apresentou queda na prevalência relativa frente a outras linhagens.

Tendo em vista o contexto apresentado, a vigilância molecular é de extrema importância para rastrear o percurso de transmissão do vírus (GIRE et al., 2014). Para isso, o sequenciamento extensivo do genoma viral de diferentes regiões do país fornece informações sobre a prevalência de cepas virais e diferenças regionais que podem levar a uma melhor compreensão dos padrões de transmissão e rastreamento de surtos (SOUZA et al., 2021).

2. REVISÃO BIBLIOGRÁFICA

2.1. Descrição molecular do SARS-CoV-2

O vírus conhecido atualmente por SARS-CoV-2 é um vírus envelopado, de RNA de fita simples, sentido positivo e com aproximadamente 30.000 bases nucleotídicas, codificando 9.860 aminoácidos (CHAN et al., 2020). O genoma viral é traduzido em 11 genes: ORF1ab, clivadas proteoliticamente para se formar 16 proteínas não estruturais; Spike (S); Envelope (E); Membrana (M); Nucleocapsídeo (N); ORF3a; ORF6; ORF7a; ORF7b; ORF8; e ORF10 (YOSHIMOTO, 2020).

A ORF1ab possui duas janelas de leitura diferentes, sendo separada em ORF1a e ORF1b (ZHANG et al., 2021). As proteínas derivadas dessa região não possuem função estrutural e são dividida em 16 (NSP1-16) (YOSHIMOTO, 2020).

2.1.1. NSP1

Também conhecida como proteína líder, é responsável por inibir a tradução de proteínas do hospedeiro, ao se ligar à subunidade ribossomal 40S ao inserir o sua porção C-terminal no canal de mRNA (SCHUBERT et al., 2020). Graças a essa inibição, a ação de parte do sistema imune que depende da tradução celular fica comprometida, dificultando a remoção da infecção (THOMS et al., 2020). Possivelmente por bloquear a própria expressão, o vírus pode modular a concentração desta proteína na célula hospedeira para manter a produção de suas proteínas enquanto a inibe para proteínas nativas (SCHUBERT et al., 2020).

2.1.2. NSP2

Esta proteína, em SARS-CoV-1, possui a capacidade de interagir com proteínas hospedeiras proibitina 1 e proibitina 2 podendo estar relacionado a perturbação da sinalização celular (CORNILLEZ-TY et al., 2009). Além disso, em SARS-CoV-2, está relacionada à interferência na produção e modificação de lipídeos e ao processamento de RNA entre o vírus e o hospedeiro (DÍAZ, 2020). Não obstante, algumas mutações específicas diferem na sequência de aminoácidos de SARS-CoV-1 e SARS-CoV-2, podendo conferir uma maior estabilidade para o vírus de 2019 (ANGELETTI et al., 2020).

2.1.3. NSP3

Trata-se de uma protease similar a papaína com cerca de 200 kDa (YOSHIMOTO, 2020) que é a responsável pela liberação de um complexo de replicação, em SARS-CoV, presente na ORF1ab (HARCOURT et al., 2004). Ademais, a NSP3, em SARS-CoV-2, também possui função de atenuar a resposta do sistema imune do hospedeiro via interferons, sendo potencialmente um alvo para o desenvolvimento de produtos terapêuticos (SHIN et al., 2020).

2.1.4. NSP4

A NSP4 interage com a NSP3, via aminoácidos H120 e F121 (SAKAI et al., 2017) com capacidade de remodelar a membrana plasmática da célula hospedeira como outros vírus de fita de sentido positivo (SNIJDER et al., 2006). Com o impedimento da interação entre a NSP4 e a NSP3, a replicação viral fica comprometida, indicando seu papel fundamental (SAKAI et al., 2017). Ademais, a participação da NSP4, individualmente, também é capaz de interferir na atividade mitocondrial (DÍAZ, 2020).

2.1.5. NSP5

Conhecida como proteinase similar a 3C, a NSP5 cliva a porção C-terminal da poliproteína codificada no gene ORF1ab em 11 locais (DAI et al., 2020). O reconhecimento da sequência Leu-Gln|(Ser, Ala, Gly) pela proteína induz o corte proteico (ZHANG et al., 2020). Além disso, também está associada a interferência no processo de replicação e regulação do DNA da célula hospedeira (DÍAZ, 2020).

2.1.6. NSP6

Assim como a NSP4, a NSP6 possui sua função atrelada a formação de vesículas membranosas (SAKAI et al., 2017). A função dessas membranas é de proteger o complexo de replicação dos lisossomos da célula hospedeira ao limitar o crescimento da vesícula (COTTAM; WHELBAND; WILEMAN, 2014). A formação dessas vesículas está associada à modificação da estrutura e função do retículo endoplasmático e do Complexo de Golgi (DÍAZ, 2020). A proteína NSP6, juntamente da NSP1, participa da supressão da sinalização de interferon do tipo I, desempenhando um papel também na evasão do sistema imune (XIA et al., 2020).

2.1.7. NSP7, NSP8 e NSP12

Ambas as proteínas NSP7 e NSP8 interagem formando um hexadecâmero, usando oito unidades de cada, agindo como uma primase para a replicação viral (DÍAZ, 2020), além de interagir com a NSP12 (RNA polimerase dependente de RNA), aumentando a sua processividade (SUBISSI et al., 2014).

A NSP12 equivale a RNA polimerase dependente de RNA, a qual se une com os cofatores codificados pelas NSP7 e NSP8, sendo a principal responsável pela replicação viral (GAO et al., 2020). A atividade da enzima sem seus cofatores é limitada (SUBISSI et al., 2014), logo, a formação do complexo NSP12-NSP7-NSP8 é de suma importância para a transcrição dos longos genomas presentes na família dos coronavírus (HILLEN et al., 2020).

2.1.8. NSP9

A NSP9 está envolvida na perturbação do tráfego de proteínas na célula, como o transporte nuclear (BANERJEE et al., 2020; DÍAZ, 2020). Essa modificação de processos celulares se dá pela ligação da NSP9 com o componente de RNA da partícula de reconhecimento de sinal (BANERJEE et al., 2020).

2.1.9. NSP10, NSP14 e NSP16

A proteína NSP10 forma um complexo tanto com a NSP14 quanto com a NSP16 (YOSHIMOTO, 2020). O complexo NSP10-NSP14 forma uma exorribonuclease necessária para manter a fidelidade da replicação viral, ao mesmo tempo em que também é responsável por excisar inibidores análogos a nucleotídeos incorporados ao RNA em construção (LIU et al., 2021). Em relação ao complexo NSP10-NSP16, este é essencial para a prevenir e reduzir a

resposta imunológica do hospedeiro durante a tradução do RNA viral (ROSAS-LEMUS et al., 2020).

2.1.10. NSP11

É uma proteína com apenas 13 aminoácido de comprimento e com função pouco elucidada (YOSHIMOTO, 2020). A troca de janela de leitura entre a ORF1a e ORF1b ocorre na região dessa proteína (GADHAVE et al., 2020). Embora para o Vírus da Hepatite Murina a proteína NSP11 seja indispensável para a replicação viral (DEMING et al., 2007), para o vírus da bronquite infecciosa do coronavírus aviário, trata-se de uma proteína não essencial (FANG et al., 2008).

2.1.11. NSP13

A proteína NSP13 de SARS-CoV possui 99,8% de identidade quando comparada a NSP13 de SARS-CoV-2 depositada no *GenBank* sob o código NC_045512.2 (YOSHIMOTO, 2020). Em SARS-CoV, essa proteína tem a função de helicase, sendo capaz de desenrolar tanto RNA, quanto DNA duplex (JANG et al., 2020). Além disso, quando essa helicase se liga a NSP12, sua atividade é aumentada (IVANOV et al., 2004). Em SARS-CoV-2, também é dito que a NSP13 possui a capacidade de antagonizar na produção de interferon do tipo I (XIA et al., 2020).

2.1.12. NSP15

A NSP15 possui sua estrutura conservada entre os coronavírus, sendo uma endoribonuclease específica para uridina e com modo de ação similar a RNase A (PILLON et al., 2021). Ademais, a NSP15, em SARS-CoV, consegue evitar a ativação de interferons ao controlar o tamanho da calda poli-U presente na extremidade 5' de RNA viral de sentido negativo (HACKBART; DENG; BAKER, 2020).

2.1.13. ORF2 (*Spike*)

A proteína *Spike* interage com o receptor celular de ECA-2 (enzima conversora de angiotensina 2) pela unidade de ancoramento S1, enquanto a unidade S2 é responsável pela fusão da membrana viral e da membrana celular (HOFFMANN et al., 2020). Para a efetiva entrada na célula, o sítio S1/S2 é clivado por uma protease da célula hospedeira chamada furina, sendo dividida entre uma unidade de ancoramento e uma unidade de fusão (WALLS et al.,

2020). A afinidade da unidade S1 de SARS-CoV-2 pela ECA-2 é maior do que a vista para SARS-CoV (WRAPP et al., 2020), potencialmente explicando sua maior transmissibilidade (WALLS et al., 2020). Por ser uma proteína envolvida na entrada do vírus na célula, trata-se de um alvo para a neutralização via anticorpos (WRAPP et al., 2020).

2.1.14. ORF3a

A proteína codificada por esse gene é primariamente um canal iônico proteico (YOSHIMOTO, 2020). Sua função pode estar relacionada com a liberação do vírus (REDONDO et al., 2021), com a ativação de inflamassomas, com a geração de tempestades de citocinas (SIU et al., 2019) e com o processo de apoptose (REN et al., 2020).

2.1.15. ORF4 (Envelope)

A proteína do Envelope (proteína E) de SARS-CoV-2 pode atuar no aumento na perturbação da barreira epitelial, o que pode levar a um aumento no processo de inflamação (DE MAIO et al., 2020). Ademais, a proteína E também está envolvida na realocação da proteína *Spike* para uma região específica do Complexo de Golgi ou para o Retículo Endoplasmático, processos necessários para o agrupamento e montagem de novas partículas virais (BOSON et al., 2021).

2.1.16. ORF5 (Membrana)

Assim como a proteína E, a proteína codificada pela ORF5, proteína da Membrana, está envolvida com a translocação da proteína *Spike* para o Complexo de Golgi ou para o Retículo Endoplasmático (BOSON et al., 2021). Para SARS-CoV, a proteína da Membrana está envolvida com o processo de indução de apoptose da célula hospedeira ao interferir com cascatas de sinalização (TSOI et al., 2014).

2.1.17. ORF6

A proteína acessória originada pela tradução do gene ORF6 é responsável pelo bloqueio do transporte de fatores regulatórios e de fatores de transcrição para o núcleo, diminuindo a produção de interferons (MIORIN et al., 2020; XIA et al., 2020).

2.1.18. ORF7a

Assim como a ORF6, a proteína ORF7a é capaz de antagonizar com a produção de interferons (XIA et al., 2020). Possui a capacidade de sequestrar o sistema de ubiquitinação da célula hospedeira para aumentar a sua função de inibir a fosforilação de fatores de transcrição necessários para a produção de interferons (CAO et al., 2021). Estruturalmente, é dividida entre um peptídeo sinal na região N-terminal, um ectodomínio composto por folhas beta e uma região transmembrana (REDONDO et al., 2021). Além disso, o ectodomínio é tido como capaz de interagir fortemente com o receptor de células CD14+, provocando uma superprodução de citocinas (ZHOU et al., 2021).

2.1.19. ORF7b

As proteínas codificadas pela ORF7b em SARS-CoV e SARS-CoV-2 possuem 85,4% de identidade (YOSHIMOTO, 2020). Em SARS-CoV, a proteína está presente no Complexo de Golgi e que não possui função essencial para o funcionamento do ciclo viral (LIU et al., 2014). Em SARS-CoV-2, contudo, há evidências que indicam função na inibição da fosforilação de fatores de transcrição necessários para a produção de interferons (XIA et al., 2020).

2.1.20. ORF8

Embora a ORF8 seja uma proteína de SARS-CoV-2 com pouquíssima identidade quanto comparada com as proteínas da ORF8a e ORF8b de SARS-CoV (YOSHIMOTO, 2020), em SARS-CoV, lhes são atribuídas funções relevantes para a infecção viral, como a induzir a apoptose em células hospedeiras (CHEN et al., 2007) e a antagonizar no processo de sinalização para a produção de interferons (WONG et al., 2018).

2.1.21. ORF9 (Nucleocapsídeo)

Possui 90,5% de similaridade quando comparada com a mesma proteína de SARS-CoV (YOSHIMOTO, 2020). Sua função básica consiste em se ligar ao RNA viral e prover estabilidade, além de se oligomerizar para formar o capsídeo viral (GRUNEWALD et al., 2018). Em SARS-CoV-2, também é lhe atribuída a função de antagonizar com RNAi antiviral (MU et al., 2020).

2.1.22. ORF10

Trata-se de uma proteína hipotética de 38 aminoácidos de comprimento e que não possui função essencial para a infecção de SARS-CoV-2 em humanos (PANCER et al., 2020). Além disso, deleções na região dessa proteína parecem não afetar a transmissibilidade ou a sua capacidade de replicação (REDONDO et al., 2021).

2.2. Transmissão do SARS-CoV-2

O período que um indivíduo pode transmitir o vírus SARS-CoV-2 para outros varia de um a dois dias, após o surgimento dos sintomas, e prolonga-se até uma semana (CHEN et al., 2021), onde a cavidade oral é a primeira região para a replicação viral, potencialmente servindo como fonte de transmissão (DANIELL et al., 2021). Contudo, estudos (ARONS et al., 2020; HE et al., 2020) apontam que indivíduos pré-sintomáticos podem contribuir para a disseminação do vírus causador da COVID-19, ainda durante o período de incubação, o qual é o tempo necessário para a multiplicação viral e manifestação dos primeiros sintomas de uma doença (GAO et al., 2021). O vírus pode ser transmitido pelo ar (BAZZAZPOUR et al., 2021), por contato direto e por superfícies (SANTARPIA et al., 2020), podendo permanecer ativo em suspensão por horas (DOREMALEN et al., 2020). Ademais, embora a transmissão por pessoas sintomáticas seja muito maior do que por pessoas assintomáticas (CEVIK et al., 2021), a liberação de partículas, que podem carregar o vírus, durante conversas, mesmo em pessoas assintomáticas, pode servir como modo de transmissão em cômodos confinados (STADNYTSKYI et al., 2020). Ainda, em ambientes hospitalares, a presença de partículas contendo RNA de SARS-CoV-2 pode estar presente em quantidade semelhante ou até mesmo superior em alas e UTIs não dedicadas para tratamento de COVID-19, em relação as alas e UTIs dedicadas (AMATO-LOURENÇO et al., 2021). Não obstante, SANTARPIA et al. (2020) observaram a viabilidade viral das amostras coletadas do ar de hospitais, demonstrando o potencial de replicação do vírus.

Aliado a esta facilidade de transmissão por diversas vias, está o surgimento de Variantes de Preocupação com cargas virais em média 1.000 vezes maior às observadas nas linhagens iniciais B (LI et al., 2021), e com capacidade de transmissão superiores a outras Variantes já descritas (BOLZE et al., 2021).

2.3. Medidas de prevenção não farmacológicas

No início da pandemia, a ausência de vacinas foi suprida pela implementação de medidas não farmacológicas para contenção do vírus (MACHADO et al., 2020). Atualmente, a utilização de vacinas eficazes já é uma realidade, embora necessite-se mais tempo para que a população mundial atinja uma cobertura vacinal satisfatória (CHOWELL et al., 2021). O uso de algumas medidas não farmacológicas podem reduzir em mais de 35% o valor estimado do número reprodutivo (R_0), número de pessoas que cada pessoa infectada é capaz de infectar, de SARS-CoV-2 (BRAUNER et al., 2021).

Ao se observar algumas medidas de isolamento específicas, é indicado que, tanto o fechamento de escolas e universidades como forma de conter a transmissão no início da pandemia, quanto reduzir a quantidade de pessoas para menos de 10 em encontros, tiveram um grande impacto (BRAUNER et al., 2021). Enquanto isso, de um modo geral, o efeito do *lockdown*, no período inicial da pandemia, mesmo em diferentes países europeus, foi de uma redução substancial no valor estimado do número reprodutivo para o vírus (FLAXMAN et al., 2020). Além das medidas de isolamento, a lavagem das mãos é uma estratégia já usada para conter outras doenças (HUANG; STEWARDSON; GRAYSON, 2014). Ademais, a higiene das mãos é uma ideia já estabelecida, estando associada com a redução na taxa de infecções nosocomiais e que tem o potencial de reduzir a transmissão de doenças, sendo responsável por uma diminuição de 24% neste aspecto (HUANG; STEWARDSON; GRAYSON, 2014; KANTOR, 2020). Portanto, a prática de lavar as mãos, seja com sabão ou com produtos à base de álcool são essenciais para evitar a colonização e a transmissão entre indivíduos (BERARDI et al., 2020). Não obstante, essa prática se relaciona com outras recomendações como de não tocar nas mucosas de olhos, nariz e boca, ou na parte frontal de máscaras (SANTOS et al., 2021). O uso de máscaras cirúrgicas para pacientes é tido como capaz de reduzir a transmissão de aerossóis, quando se analisa casos suspeitos ou com sintomas leves (OLIVERA; COAGLIO; ALDO, 2020). Já o uso de máscaras PFF2, KN95 ou similares por 100% da população é capaz de reduzir a circulação do vírus em até 90%, embora o encorajamento ao uso e distribuição destas máscaras, no Brasil, não tenha ocorrido por parte do Governo Federal (MACHADO et al., 2020). Embora o uso de apenas algumas estratégias não seja suficiente, o uso em conjunto das medidas de prevenção podem reduzir drasticamente a contaminação e disseminação de SARS-CoV-2 (CHOWELL et al., 2021).

2.4. Epidemiologia

Em dezembro de 2019, a província de Hubei, em Wuhan, tornou-se o epicentro de um surto de pneumonia de causa desconhecida (WANG et al., 2020). O primeiro caso conhecido é de uma vendedora de frutos do mar, a qual teve sintomas em 10 de dezembro de 2019 (WOROBEY, 2021). Embora houvesse um mecanismo chinês para notificação de pneumonia viral de etiologia desconhecida, o qual usa uma interface baseada em internet, este se mostrou insatisfatório para notificar os primeiros casos suspeitos de SARS-CoV-2 (WOROBEY, 2021). Aliado a isso, naquele momento, cerca de 93% dos casos não levavam a hospitalização, indicando que o caso índice é potencialmente ainda mais antigo, mas, ainda, associado ao mercado de Huanan (WOROBEY, 2021). Ademais, nos períodos iniciais, o período de incubação viral era de 5,2 dias, com um número reprodutivo estimado de 2,2 e dobrando o tamanho da epidemia a cada 7,4 dias (LI et al., 2020).

Em 7 de janeiro de 2020 o primeiro genoma de SARS-CoV-2 foi sequenciado (WANG et al., 2020). No final de janeiro já foram detectados casos no Japão, nos EUA, na Coreia do Sul, dentre outros países (WANG et al., 2020). O primeiro caso confirmado de SARS-CoV-2 no Brasil foi em 25 de fevereiro de 2020 (THE LANCET, 2020). Mesmo que houvesse políticas apropriadas à época, como isolamento rigoroso, que mesmo sem *lockdown* pode conferir uma grande proteção contra SARS-CoV-2 (BOON et al., 2021), devido a elevada quantidade de grupos em condições de instabilidade socioeconômica, o combate à COVID-19 seria árduo (THE LANCET, 2020).

As primeiras linhagens de SARS-CoV-2 detectadas no Brasil no início de 2020 foram as B.1.1, B.1.1.28, P.2, B.1.1.33, N.9 e N.10, sendo praticamente extintas após o surgimento da Variante Gama (linhagem P.1) (PERICO et al., 2021). A Gama, embora contida ao estado de Amazonas em 2020, com os eventos e viagens de fim de ano, rapidamente se espalhou para as outras regiões do país (NAVECA et al., 2021a). No início de 2021, na segunda onda da pandemia, a quantidade de mortos diários chegava a 3.500, totalizando até o final de 2021, mais de 600.000 mortes (LAMARCA et al., 2021). Em 26 de abril de 2021, o primeiro caso da Variante Delta (linhagem B.1.617.2) foi registrado no Brasil, tornando-se a linhagem dominante, embora, provavelmente devido à taxa de vacinação, uma nova onda de mortes não tenha acontecido (LAMARCA et al., 2021).

2.5. Vacinas

A resposta imunológica mediadas por células T CD4⁺ e os títulos de anticorpos anti-*Spike* diminuem com o tempo após uma infecção leve ou moderada por SARS-CoV-2, embora a quantidade de células B específicas para a proteína *Spike* se acumule (WHEATLEY et al., 2021). Embora esta última informação possa gerar uma falsa sensação de tranquilidade, não se considera o evento em pessoas com comorbidades, ao passo que, já a vacina BNT162b2 (Pfizer/BioNTech), por exemplo, oferece uma proteção de 95% após duas doses, independente de idade, sexo ou presença de condições de saúde específicas (POLACK et al., 2020). Por outro lado, a vacina ChAdOx1 (Oxford-AstraZeneca) oferece 70,4% de proteção (VOYSEY et al., 2021).

Vacinas foram produzidas com o intuito de reduzir a transmissão e a mortalidade por SARS-CoV-2, principalmente em pessoas com condições médicas específicas (WALSH et al., 2020). Além disso, a imunidade, a nível populacional, via vacina, também beneficia outros grupos, como aqueles que respondem menos (SHROTRI et al., 2021) ou não respondem à vacina, que não podem se vacinar ou que não tem acesso à vacina (VOYSEY et al., 2021).

Para a vacina BNT162b2, a título de ilustração, pode-se ver uma redução na proteção de longo prazo em relação a Variante de Preocupação Delta, num intervalo de apenas três meses (MIZRAHI et al., 2021). Uma redução significativa, para anticorpos anti-*Spike*, foi vista tanto para a vacina BNT16b2 quanto para ChAdOx1 no período de mais de 70 dias após a aplicação da segunda dose, apesar de não se saber ainda qual o limite mínimo para se conferir proteção contra SARS-CoV-2 (SHROTRI et al., 2021). O decréscimo no título de anticorpos, segundo outro estudo (LEVIN et al., 2021), após 3 meses permanece constante, contudo, a taxa de neutralização sofre um decréscimo menor do que a ocorrida para o título o que indica a presença de anticorpos mais eficientes.

Parte da estratégia para se aumentar a eficiência da imunização está no intervalo entre a aplicação das doses, onde um período entre 3 a 4 semanas é empregado (SHROTRI et al., 2021). Um estudo que analisou a extensão do período de aplicação entre a primeira e a segunda dose da vacina BNT162b2 indicou um aumento no pico de anticorpos após um intervalo de 12 semanas, em relação ao padrão de 3 semanas, enquanto que a resposta celular encontrada foi menor (PARRY et al., 2021).

A aplicação heteróloga também se configura como uma destas estratégias: no estudo (BARROS-MARTINS et al., 2021), o grupo que recebeu uma dose de ChAdOx1 seguida de BNT162b2 foi capaz de desenvolver anticorpos neutralizantes para as variantes de

preocupação, ao passo que o grupo que recebeu duas doses de ChAdOx1 não produziu anticorpos eficientes contra as linhagens P.1 (Variante Gama) e B.1.351 (Variante Beta). Embora isso possa ocorrer, é sabido que mesmo uma única dose é capaz, ainda de gerar proteção contra todas as 4 Variantes (NASREEN et al., 2021). Ademais, a aplicação de uma terceira dose é capaz de reduzir substancialmente os casos de COVID-19, em relação às pessoas que receberam somente duas doses (BAR-ON et al., 2021), além de aumentar a neutralização de variantes, como a Delta, mesmo em pacientes receptores de transplantes de órgãos sólidos, os quais costumam possuir respostas imunológicas atenuadas (KARABA et al., 2021).

2.6. Sequenciamento Genômico

2.6.1. Sanger

A tecnologia desenvolvida em 1977 por Frederick Sanger, o método de sequenciamento de DNA que utilizava dideoxynucleotídeos, foi empregada para, ao longo de vários anos, obter o primeiro rascunho do genoma humano (HUTCHISON, 2007). A técnica é baseada em análises de PCR que utiliza nucleotídeos comuns (deoxynucleotídeos) e nucleotídeos que não permitem a continuação da reação (dideoxynucleotídeos), formando diferentes comprimentos de fragmentos que podem ser separados por eletroforese e identificados (METZKER, 2005).

Atualmente, a tecnologia empregada para perfazer o sequenciamento Sanger difere do que foi inicialmente construído, utilizando-se técnicas modernas e avançadas de eletroforese capilar e detecção de fluoróforos (THERMOFISHER SCIENTIFIC, 2016).

2.6.2. NGS - *Next-Generation Sequencing*

Um dos maiores avanços da tecnologia de NGS, em relação ao sequenciamento Sanger, está na enorme quantidade de dados que se pode obter a um custo relativamente baixo, o que permite o sequenciamento de vários organismos e permite a execução de análises evolutivas (METZKER, 2010).

Algumas análises exigem, devido a pouca quantidade de amostra disponível para o sequenciamento, etapas de amplificação da amostra inicial (CHANG; LI, 2013). Embora muitas vezes necessária, essa etapa é suscetível a erros e vieses por conta do uso de iniciadores (*primers*) específicos, demonstrando a importância do controle de qualidade antes, durante e após a montagem dos genomas (CHEN et al., 2018).

De modo geral, para a construção de bibliotecas genômicas, a amostra é clivada, seja por enzimas ou por métodos físicos, em fragmentos menores (BRONNER; QUAIL, 2019). Os fragmentos de DNA são ligados a adaptadores, os quais são usados para a hibridização de

primers, permitindo a amplificação da região. Um passo comum entre algumas tecnologias de NGS: a biblioteca genômica é fixada em uma camada sólida, permitindo o sequenciamento simultâneo de várias amostras (METZKER, 2010).

Na tecnologia Illumina, a qual usa a estratégia de terminação cíclica reversível, após a adição de um nucleotídeo, a reação é parada para, logo em seguida, a clivagem do terminador permitir a continuação da reação em cadeia (METZKER, 2010). Com isso, o sequenciador é capaz de identificar, via fluorescência, qual base foi adicionada, sendo a identificação de um nucleotídeo em específico feita a partir do consenso da informação transmitida pelo cluster em um dado ciclo de amplificação (BRONNER; QUAIL, 2019).

O produto do sequenciamento, independente da tecnologia empregada, são os arquivos FastQ, os quais representam as sequências identificadas durante o sequenciamento, além de parâmetros de qualidade utilizados por algoritmos durante a montagem do genoma (COCK et al., 2009).

A montagem do genoma, por sua vez, pode ser realizada tanto por referência ou por *de novo*. A montagem por referência, a qual é realizada ao se mapear as leituras obtidas no sequenciamento dentro do arquivo FastQ contra uma sequência padrão (BEHJATI; TARPEY, 2013), embora possua limitações, é de interesse quando o objetivo da pesquisa está na comparação com um genoma de referência e a anotação de pequenas variações (METZKER, 2010).

2.6.3. Controle de qualidade e *minor variants*

Para a correta anotação dos nucleotídeos do genoma viral e para garantir confiança na designação de linhagens, o controle de qualidade e alguns pré-tratamentos nos arquivos oriundos dos sistemas de sequenciamento automatizado são de suma importância (CHEN et al., 2018). Programas como o fastp (CHEN et al., 2018) são capazes de retirar as sequências dos adaptadores do preparo das bibliotecas, remover as sequências dos primers usados na etapa de amplificação do RNA viral, correção de bases, filtragem de bases com baixa qualidade, além da geração de relatórios de qualidade. Outras etapas, as quais são desenhadas para funcionar em outros programas, como iVar (GRUBAUGH et al., 2019) e BWA (LI; DURBIN, 2010) também empregam escores de qualidade, como o Phred *score* e MAPQ. Estes parâmetros são, geralmente, calculados por modelos matemáticos integrados ao próprio sequenciador (ILLUMINA, 2011) ou por modelos presentes nos programas utilizados (LI, 2013).

Inicialmente, o Phred foi um algoritmo usado para avaliar a qualidade do sequenciamento Sanger (EWING; GREEN, 1998), interpretando, por exemplo, o formato dos picos dos eletroferogramas. Para os sequenciamentos de nova geração (NGS), parâmetros que variam conforme a química do processo são comparados com bancos de dados empíricos de acurácia conhecida, possibilitando a predição da qualidade da base para os sequenciamentos (ILLUMINA, 2011). De todo modo, o Phred *score* descreve a probabilidade de uma base em uma leitura estar errada, sendo calculado pela seguinte fórmula:

$$Q = -10 \log_{10} P$$

Onde P é a probabilidade de a base estar errada. Desse modo, por exemplo, caso uma base possua Phred *score* de 30 ou Q30, significa que a probabilidade desta base estar errada é de 1 em 1000, inferindo que a acurácia da base é de 99,9%.

Tabela 1 Relação entre valor de Phred *score*, probabilidade de a base estar errada e acurácia da base.

Phred score	Probabilidade de a base estar errada	Acurácia da base
10	1 em 10	0,9
20	1 em 100	0,99
30	1 em 1.000	0,999
40	1 em 10.000	0,9999
50	1 em 100.000	0,99999

Fonte: Elaborado pelo autor.

Outros conceitos utilizados para a avaliação da qualidade de um sequenciamento são a profundidade das leituras ao longo do genoma e a cobertura do genoma. A primeira é referente a quantidade de leituras com qualidade igual ou superior a especificada para determinada posição, e a segunda descreve a porcentagem do genoma montado em relação ao de referência. Estes conceitos podem ser integrados, assumindo que deva existir uma profundidade mínima para que uma posição possa ser contada como não ambígua (SIMS et al., 2014), onde por ambíguo se entende um nucleotídeo que representa mais de uma base nucleotídica comum (A, T, C ou G).

A relevância destes escores de qualidade permite a identificação de mutações que estão em frequência reduzida (*minor variants*), ou iSNVs (*intra-host single nucleotide variants*), quando comparada a um nucleotídeo majoritário, permitindo a designação de diferentes linhagens por programas como o *Pangolin* (O'TOOLE et al., 2021a).

Assim como a geração de dados se tornou muito maior devido a evolução da tecnologia de sequenciamento nos últimos anos, diversos *softwares* foram desenvolvidos, como *Freebayes* (GARRISON; MARTH, 2012), *GATK HaplotypeCaller* (PIROOZNIA et al., 2014), *Samtools* (LI, 2011) e *iVAR*, permitindo um maior aprofundamento da capacidade do sequenciamento de segunda geração de detectar tais variações (KOBOLDT, 2020).

Além disso, para se falar sobre a detecção de variantes e mutações minoritárias são empregados termos que refletem a capacidade do *software* de identificar verdadeiros positivos e verdadeiros negativos, respectivamente a sensibilidade e a especificidade. Contudo, como se deve esperar, é necessário atingir um balanço entre essas duas variáveis, já que, normalmente, para aumentar a sensibilidade, deve-se adotar padrões menos rígidos para a detecção, diminuindo a especificidade (MCCRONE; LAURING, 2016).

Este ponto leva a necessidade da validação da acurácia de um determinado programa, a qual pode ser feita a partir da comparação com outras tecnologias, como sequenciamento Sanger de variantes já conhecidas ou por genotipagem de polimorfismo de nucleotídeo único (SNP) (PIROOZNIA et al., 2014). Outras metodologias buscam disponibilizar bancos de dados já testados previamente que possuem variantes que são consenso entre diferentes programas (KOBOLDT, 2020), *pipelines* para que se possa validar diferentes abordagens e a utilização da elaboração artificial de populações virais para se testar a sensibilidade e especificidade de determinados programas (GRUBAUGH et al., 2019; MCCRONE; LAURING, 2016).

Ademais, há, também, além da dificuldade inerente ao processo de detecção dos nucleotídeos minoritários pelos programas, vieses que podem ser encontrados em diferentes etapas do preparo das amostras, desde a amplificação por reação em cadeia da polimerase (PCR) (GRUBAUGH et al., 2019), quanto em etapas posteriores no preparo das bibliotecas, o que torna bastante delicado o trabalho com esse tipo de análise (SCHIRMER et al., 2015).

Sendo esses desafios pertinentes para a correta anotação de *minor variants*, conjuntos de procedimentos são descritos como boas práticas antes de se tentar inferir algo, indicando a necessidade do pré-tratamento dos dados brutos, alinhamento seguido de filtro de qualidade para as bases, realinhamento (KOBOLDT, 2020; PIROOZNIA et al., 2014), além de, quando de suma importância para o trabalho, o uso de replicatas técnicas (GRUBAUGH et al., 2019).

A origem destas mutações pode ser devido, por exemplo, a ocorrência de coinfeções, quando o paciente é infectado por mais de uma linhagem ao mesmo tempo (PEDRO et al., 2021). De modo similar, como o vírus já possui escala global e grande transmissibilidade, é esperado que mutações ocorram, sendo elas benéficas ou maléficas ao ciclo de replicação do

vírus, inclusive em proteínas responsáveis pela multiplicação do material genético, potencialmente ocasionando um aumento da taxa de mutações (BOTELHO-SOUZA et al., 2021). Estas mutações, que também podem ser encontradas durante o processo de montagem do genoma, especialmente as com frequência crescente, podem apresentar pistas do surgimento de novas linhagens.

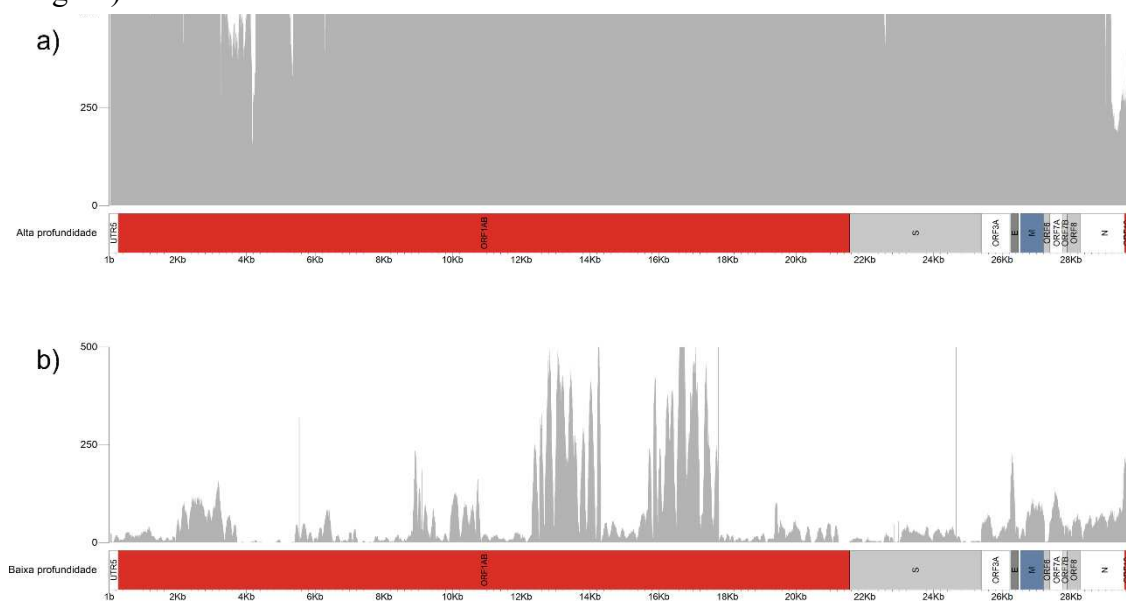
Ademais, o uso de gráficos representando a distribuição da profundidade de cada nucleotídeo para o genoma de SARS-CoV-2 também são empregados. Esta representação gráfica permite uma fácil inspeção visual em busca de regiões onde houve queda da profundidade em relação à média. Problemas como este podem estar relacionados a baixa qualidade da amostra usada para o sequenciamento (HEALEY et al., 2014), ao protocolo de amplificação do material genético ou a outros procedimentos no preparo da amostra para o sequenciamento (BRONNER; QUAIL, 2019). A implicação provocada por estas quedas (**figura 1**), no âmbito aqui explorado, é de dificuldades de detectar mutações em certas regiões do genoma e de capturar nucleotídeos minoritários presentes no *pool* da amostra clínica. A detecção de mutações específicas se faz bastante importante quando poucas mutações sinapomórficas são suficientes para separar sublinhagens (RAMBAUT et al., 2020) e permitir o acompanhamento da evolução viral detalhadamente. Além disso, como é esperado, dados os requisitos para a anotação dos nucleotídeos minoritários, descritos anteriormente, a profundidade média é um importante parâmetro a ser considerado. Estas estimativas corroboram com o que é descrito em (GRUBAUGH et al., 2019) onde é dito que a profundidade do sequenciamento é de suma relevância para a detecção de nucleotídeos alternativos dentro de

uma mesma amostra, além da frequência do genoma alternativo no *pool* da amostra e da quantidade inicial de cópias de cDNA antes do preparo das bibliotecas.

2.7. Árvores filogenéticas

Árvores filogenéticas são hipóteses da relação parental e de ancestralidade entre organismos e genes (GREGORY, 2008). São compostas por ramos e nós, onde os nós terminais consistem das unidades de estudo, formalmente chamadas de Unidades Taxonômicas Operacionais (UTO), e os ramos o grau de divergência entre as UTOs (FELTES et al., 2014). Devido ao conceito de que um organismo dá origem a outros, a relação entre os nós das árvores filogenéticas tende a representar esse parentesco. Dentro da árvore pode haver dicotomias (um nó dando origem a outros dois) e politomias (um nó interno dando origem a mais de dois nós). Normalmente, o uso de politomias em árvores filogenéticas representa uma maior incerteza em

Figura 1 - Mapas de profundidade: em a) uma amostra com perfil de alta qualidade (< 1% de bases ambíguas e >29Kb); em b) uma amostra com perfil de baixa qualidade (>10% de bases ambíguas).



Independentemente do modo de construção da árvore filogenética, quando se trata de filogenia molecular, são usados alguns parâmetros que irão reger os cálculos necessários para a formulação da hipótese de relação entre as UTOs (RODRIGUEZ; OLIVER; MARIN, 1990). Dentre esses parâmetros, pode-se destacar os modelos de substituição de nucleotídeos, a taxa de heterogeneidade entre sítios e a frequência de cada base. Os modelos de substituição de nucleotídeos buscam explicar e quantificar, como o próprio nome sugere, as taxas de alteração entre um dado nucleotídeo e outro (ALEXEI J. DRUMMOND; REMCO R. BOUCKAERT, 2015). As alterações de nucleotídeo são úteis para averiguar o quão distante duas sequências moleculares são. Contudo, uma análise simples como a contagem de diferenças pontuais entre sequências tende a subestimar a sua real diferença (ALEXEI J. DRUMMOND; REMCO R. BOUCKAERT, 2015). Modelos como K80 (KIMURA, 1980), HKY (HASEGAWA; KISHINO; YANO, 1985) e GTR (RODRIGUEZ; OLIVER; MARIN, 1990) utilizam-se de conceitos que vão desde a incorporação da frequência da base (A, T, G, C) até as taxas de transição e transversão. Não obstante, algumas regiões do genoma de organismos tendem a sofrer processos de substituição mais frequentemente do que outras, de modo a ser necessário incluir essa variabilidade no modelo de substituição de nucleotídeos (YANG, 1993).

Filogenias que envolvem cálculos de probabilidades ainda mais complexos e que incorporam o tempo como uma variável são capazes de estimar outras características como a data de surgimento do ancestral comum mais recente para um dado clado monofilético (ALEXEI J. DRUMMOND; REMCO R. BOUCKAERT, 2015), além de indicar a reconstrução geográfica da disseminação de um patógeno durante uma epidemia (ISHIKAWA et al., 2019; SUCHARD et al., 2018).

3. OBJETIVOS

3.1. Objetivo geral

Investigar o surgimento, expansão e declínio da Variante de Preocupação Gama do vírus SARS-CoV-2 no estado do Ceará no período de 2020/2021.

3.2. Objetivos específicos

- Identificar o período de introdução da Variante Gama do vírus SARS-CoV-2 no estado do Ceará;
- Ilustrar a sua expansão frente outras linhagens;
- Relatar casos de importação e exportação;
- Acompanhar o perfil mutacional da linhagem P.1 e de suas sublinhagens ao longo do tempo no estado do Ceará;
- Inferir a sua disseminação nos municípios do estado do Ceará;
- Identificar o período e possíveis causas para o declínio da Variante Gama no estado do Ceará.

4. METODOLOGIA

4.1. Obtenção e triagem das amostras

As amostras foram triadas e encaminhadas para o processo de sequenciamento foram coletadas em hospitais regionais do estado do Ceará:

- Hospital Leonardo da Vinci (HLV)
- Maternidade Escola (Meac)
- Hospital São José (HSJ)
- Hospital César Cals (HCC)
- Hospital de Messejana (HMS)
- Hospital Infantil Albert Sabin (HAS)
- Hospital Universitário Walter Cantídio (HUW)
- Hospital da Mulher de Fortaleza (HMF)
- Hospital da Polícia Militar (HPM)
- Hemoce (HMC)
- Banco de olhos do Ceará (BOC)

As amostras foram coletadas por *swab* de forma nasal ou oral, armazenadas a -4 °C, identificadas e encaminhadas para o Centro de Hematologia e Hemoterapia do Ceará (Hemoce)

A escolha das amostras que seguiram para a etapa de sequenciamento deve cumprir um pré-requisito em relação ao seu Ct (Cycle Threshold), ou seja, número de ciclos onde a amostra ultrapassa uma quantidade pré-determinada de fluorescência estabelecida pelo kit comercial utilizado. Para a etapa de RT-PCR utilizou-se o kit TaqPath COVID-19 CE-IVD RT-PCR Kit (ThermoFisher Scientific). Para serem consideradas como aptas para a etapa de sequenciamento, as amostras devem possuir valor de CT inferior a 25. Como o valor do Ct está relacionado com a carga viral da infecção (CEVIK et al., 2021), as amostras que foram sequenciadas apresentam esse viés.

4.2. Sequenciamento

A extração de RNA foi feita com o kit MagMAX viral/pathogen nucleic acid isolation, a conversão para cDNA, amplificação (geração dos *amplicons*) e preparo da biblioteca genômica foram realizados com o kit CovidSeq (Illumina). Utilizou-se o sequenciador de nova geração MiSeq (Illumina) localizado na Fundação Oswaldo Cruz Ceará para a geração de

leituras *pair-end* de 150 pb. Estas leituras, juntamente de valores de qualidade relacionados a cada nucleotídeo, formaram o arquivo de saída denominado FastQ.

Para este estudo foram utilizadas 3310 amostras positivas para SARS-CoV-2 com boa qualidade (<10% de bases ambíguas e profundidade média > 100) coletadas entre o período de janeiro a novembro de 2021.

4.3. Montagem dos genomas

Os arquivos FastQ necessários para a montagem dos genomas foram obtidos por meio do sequenciamento genômico de amostras preparadas segundo Protocolos Operacionais Padrão do Laboratório de Biologia Molecular e Inovação Diagnóstica (Biomind) do HEMOCE. Os genomas foram montados usando os programas fastp versão 0.20.1 (CHEN et al., 2018) para o pré-tratamento dos dados, BWA versão 0.7.17 (LI, 2013) para o mapeamento e alinhamento das leituras, samtools versão 1.9 (LI, 2011) e iVar 1.3.1 (GRUBAUGH et al., 2019) para a geração dos genomas consenso e do genoma minoritário, Pangolin versão 3.1.16 (O'TOOLE et al., 2021b) para a designação de cada linhagem, além de *scripts* para a integração dos programas. A *pipeline* utilizada está disponível no link <https://github.com/dezordi/ViralFlow> com o genoma de referência com número de acesso NC_045512.2 obtido na plataforma *National Center for Biotechnology Information* (NCBI).

A profundidade mínima para a designação de um nucleotídeo foi de 5 bases com qualidade igual ou superior à 30 no escore *Phred*.

Para a detecção das *minor variants* que compõe o genoma minoritário, foram estabelecidas três regras:

- i. O nucleotídeo deve estar presente nas duas leituras, sentido positivo e sentido negativo.
- ii. O nucleotídeo deve apresentar pelo menos 100 de profundidade com qualidade igual ou superior a 30 no escore *Phred*.
- iii. O nucleotídeo minoritário deve possuir frequência de pelo menos 5% do total de leituras para aquela posição.

4.4. Automação de processos

Foram desenvolvidos três *scripts* em *Python* e um em R para a automatização de alguns processos, bem como geração de dados e gráficos:

1. Auto.py: Automatiza o processo de entrada de dados no *pipeline* de montagem de genomas

Pseudocódigo 1 - Demonstração do funcionamento básico do script auto.py

Algoritmo 1

Entrada: Diretório das pastas com arquivos fastq, Sequência de referência, Profundidade mínima, Adaptadores

Início

Pastas \leftarrow DiretórioDasPastasComArquivosFastq.ListarPastas()

Para cada Pasta **em** Pastas **faça**

R1 \leftarrow Fastq1

R2 \leftarrow Fastq2

Executar comando da pipeline(SequênciaReferência, R1, R2, ProfundidadeMínima, Adaptadores)

Fim

Fim

Fonte: Elaborado pelo autor.

2. Freq.py: anota a frequência corrigida de cada posição no alinhamento múltiplo de sequências e retorna um gráfico com as frequências para o genoma completo.

Pseudocódigo 2 – Demonstração do funcionamento básico do script Freq.py

Algoritmo 2

Entrada: Diretório dos arquivos fasta alinhados com a referência separados por meses, frequência mínima

Início

ListaDeAlinhamentos \leftarrow DiretórioDosArquivosAlinhados.ListarArquivos()

Para cada Alinhamento **em** ListaDeAlinhamentos **faça**

Para cada posição **em** Alinhamento **faça**

 QuantidadeTotal \leftarrow posição.Contar(Tudo)

 QuantidadeN \leftarrow posição.Contar(N)

 QuantidadeA \leftarrow posição.Contar(A)

 QuantidadeT \leftarrow posição.Contar(T)

 QuantidadeC \leftarrow posição.Contar(C)

 QuantidadeG \leftarrow posição.Contar(G)

 FrequenciaA \leftarrow (QuantidadeA)/(QuantidadeTotal - QuantidadeN)

 FrequenciaT \leftarrow (QuantidadeT)/(QuantidadeTotal - QuantidadeN)

 FrequenciaC \leftarrow (QuantidadeC)/(QuantidadeTotal - QuantidadeN)

 FrequenciaG \leftarrow (QuantidadeG)/(QuantidadeTotal - QuantidadeN)

Se FrequenciaA \geq FrequenciaMínima **faça**

Escreva FrequenciaA, posição

Se FrequenciaT \geq FrequenciaMínima **faça**

Escreva FrequenciaT, posição

Se FrequenciaC \geq FrequenciaMínima **faça**

Escreva FrequenciaC, posição

Se FrequenciaG \geq FrequenciaMínima **faça**

Escreva FrequenciaG, posição

Fim

DesenharGráfico()

Fim

Fim

Fonte: Elaborado pelo autor

3. Arq.py: compila os resultados brutos do *pipeline*, tais como informações sobre cobertura, profundidade média, linhagem e *minor variants*

Pseudocódigo 3 – Demonstração do funcionamento básico do script Arq.py

Algoritmo 3

Entrada: Diretório das pastas com resultados

Início

Pastas ← DiretórioDasPastasComResultados.ListarPastas()

Cria arquivo CompiladoQualidade

Cria arquivo CompiladoPangolin

Cria arquivo CompiladoMinorVariants

Para cada Pasta em Pastas **faça**

CompiladoQualidade.Escreva(ArquivoQualidade)

CompiladoPangolin.Escreva(ArquivoPangolin)

CompiladoMinorVariants.Escreva(ArquivoMinorVariants)

Fim

Fim

Fonte: Elaborado pelo autor.

4. Depth_maps.R: cria os mapas de profundidade do genoma para todas as amostras.

Pseudocódigo 4 - Demonstração do funcionamento básico do script Depth_maps.py

Algoritmo 4

Entrada: Diretório das pastas com resultados

Início

Pastas ← DiretórioDasPastasComResultados.ListarPastas()

Para cada Pasta em Pastas **faça**

BAM ← ArquivoBam

BAI ← ArquivoBai

KaryoploteR(BAM, BAI)

Fim

Fim

Fonte: Elaborado pelo autor.

4.5. Geração de mapas de profundidade

Foram escolhidas duas amostras com o fito de demonstrar o **algoritmo 4**, dentro as disponíveis no banco de dados interno do Laboratório Biomind. Uma, nomeada “Alta profundidade” que apresentou um perfil de alta qualidade (<1% de bases ambíguas) e completos (>29Kb) e outra, denominada “Baixa profundidade” com perfil de baixa qualidade (>10% de bases ambíguas).

Os mapas de profundidade foram feitos com a biblioteca KaryoploteR, disponível no link https://bernatgel.github.io/karyoploter_tutorial/ para a linguagem de programação R, usando os arquivos de saída com extensão “.BAM” e “.BAI” do programa BWA e samtools

4.6. Número de óbitos

No dia 13 de novembro de 2021, o número de casos por Síndrome Respiratória Aguda Grave (SRAG) foi obtido na plataforma <https://opendatasus.saude.gov.br/> tanto para o ano de 2020 quanto para o ano de 2021. Os dados foram filtrados usando algoritmo interno para serem apresentados somente óbitos cuja causa foi definida como por Sars-Cov-2 e no Estado do Ceará.

4.7. Prevalência relativa

Foram disponibilizadas pelo serviço de Vigilância Genômica da Fiocruz Ceará as amostras sequenciadas no período de janeiro a novembro de 2021, totalizando 3310 amostras de boa qualidade (<10% de bases ambíguas e profundidade média > 100). Esse conjunto de amostras foi usado para formar o *dataset_1* (d1).

Os genomas analisados foram gerados a partir de amostras clínicas priorizadas para sequenciamento por critérios não randomizados, considerando-se a data de coleta do exame de rt-PCR. Logo, a seleção baseada em carga viral mínima (aferida por valor Ct), pelo resultado do teste molecular de inferência e/ou critérios epidemiológicos, como suspeita de importação, pode afetar a representatividade dos dados.

Tabela 1 - Diferentes conjuntos de dados usados para diferentes análises e suas respectivas características

Conjunto de dados	Características	Análises	Programas
	3310 amostras	Prevalência relativa	<i>scripts in-house</i>
Dataset_1 (d1)	< 10% de bases ambíguas	Frequência de mutações	
	204 amostras (P.1)	Filogenia Bayesiana	Beast v.1.10.4
Dataset_2 (d2)	< 1% de bases ambíguas	Número Reprodutivo	Beast v.2.6.5
	43 amostras (P.1*)	Filogenia Bayesiana	Beast v.1.10.4
Dataset_3 (d3)	< 1% de bases ambíguas	Número Reprodutivo	Beast v.2.6.5
	204 amostras (P.1 <i>Spike</i>)	Filogeografia	Beast v.1.10.4
Dataset_2_s (d2_s)	< 1% de bases ambíguas		
	43 amostras (P.1* <i>Spike</i>)	Filogeografia	Beast v.1.10.4
Dataset_3_s (d3_s)	< 1% de bases ambíguas		
	1322 amostras (P.1)	Filogeografia por Máxima	TreeTime
Dataset 4 (d4)	< 1% de bases ambíguas	Verossimilhança Temporal	PastML

Fonte: elaborado pelo autor.

4.8. Filogeografia e inferência Bayesiana

Todas as amostras P.1 e sublinhagens de P.1 de alta qualidade (<1% de bases ambíguas) e genomas completos (>29Kb) amostradas no estado do Ceará e que possuíam data de coleta completa (n = 1151) disponíveis na plataforma GISAID (<https://www.gisaid.org/>) no dia 05 de novembro de 2021, de janeiro até setembro de 2021 foram selecionadas. As amostras foram submetidas a uma etapa de agrupamento usando o programa CD-Hit (FU et al., 2012) com similaridade de 99,8% e cobertura de 100%, quando comparada a sequência representativa do grupo, e separadas por linhagem: P.1 (n = 229); sublinhagens P.1 (n = 47).

Os genomas foram alinhados usando o programa MAFFT v.7.475 (KATO; STANDLEY, 2013) e usados para a construção de uma árvore por Máxima Verossimilhança usando o programa IQ-TREE v.1.6.12 (NGUYEN et al., 2015). A árvore foi construída sob o modelo de substituição de nucleotídeos GTR (*general time-reversible*) com uma taxa de variação entre sítios seguindo a distribuição *Gamma* com quatro categorias (G4), uma frequência empírica de bases (+F), como realizado em (NAVECA et al., 2021a) e o suporte do ramo foi avaliado pelo teste de razão de verossimilhança aproximada baseado no procedimento do tipo Shimodaira-Hasegawa com 1.000 replicatas.

O sinal temporal das amostras foi avaliado segundo o programa TempEst v1.5.3 (RAMBAUT et al., 2016) usando as árvores filogenéticas anteriores e amostras que possuíam discrepância entre a data de coleta e distância genética foram consideradas como *outliers* (RAMBAUT et al., 2016) e excluídas, assim como aquelas que não possuíam município de coleta, totalizando 204 sequências P.1, compondo o *dataset_02* (d2) e 43 P.1+, compondo o *dataset_03* (d3).

Ambos os *datasets* foram utilizados para uma análise filogeográfica Bayesiana temporal usando a abordagem Bayesiana *Markov chain Monte Carlo* (MCMC) implementada no programa BEAST v.1.10.4 (SUCHARD et al., 2018) usando a biblioteca BEAGLE v.3 (SUCHARD; RAMBAUT, 2009) para diminuir o tempo computacional associado. A árvore Bayesiana foi construída usando o modelo GTR+F+I+G4 como em (NAVECA et al., 2021a), utilizando o modelo não paramétrico *Bayesian skyline* e um relógio molecular estrito com uma distribuição a priori uniforme e taxa de substituição de nucleotídeos por sítio por ano de $8-10 \times 10^{-4}$, sendo as cadeias MCMC executadas por 200 milhões de gerações.

Os *datasets* d2 e d3 foram usados para gerar outros dois, denominados d2_S e d3_S. Usando o programa GENEIOUS v.2021.2.2 (KEARSE et al., 2012), foi recuperado o gene S destas sequências, além de retirar do d1_S as sequências cujo município de coleta aparecia

somente uma vez, com o fito de tornar execrável a resolução da filogeografia. Foram usados os mesmos parâmetros para os novos *datasets*, com a exceção de que as cadeias MCMC foram executadas por 800 milhões e 100 milhões de gerações, respectivamente, além de um modelo filogeográfico discreto reversível com uma taxa de *continuous-time Markov chain* (CTMC) como referência a priori.

A convergência de ambas as análises foi avaliada pelo programa TRACER v.1.7.2 (RAMBAUT et al., 2018) garantindo um *effective sample size* (ESS) > 200 e que a corrida apresentou *mixing* apropriado. As árvores foram sumarizadas via TreeAnnotator v.1.10.4. e visualizada usando FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

4.9. Estimação do número efetivo de reprodução (R0)

Com a finalidade de estimar a trajetória da linhagem P.1 e de suas sublinhagens no estado do Ceará, foi usado o modelo de birth-death skyline (BDSKY) implementado no programa BEAST 2 v.2.6.5 (BOUCKAERT et al., 2019), tendo como arquivos de entrada os *datasets* d2 e d3. A taxa de amostragem e as distribuições a priori do modelo BDSKY foi ajustado conforme (NAVECA et al., 2021b). O número de intervalos para a linhagem P.1 foi de 7 meses, enquanto para as sublinhagens foi de 6 meses. As cadeias MCMC foram executadas por 100 milhões de gerações, apresentando *ESS* > para os parâmetros analisados.

4.10. Análise de máxima verossimilhança temporal

Para análise de importações e exportações, todas as amostras P.1 de alta qualidade (<1% de bases ambíguas) e completas (>29Kb) amostradas no Brasil (n = 28755) disponíveis na plataforma GISAID (<https://www.gisaid.org/>) de dezembro de 2020 até novembro de 2021 foram recolhidas, separadas por estado e submetidas a uma etapa de agrupamento usando o programa CD-Hit com similaridade de 99,8% e cobertura de 100%, quando comparada a sequência representativa do grupo. Posteriormente, de modo randomizado, uma amostra, por dia e por estado, foi escolhida para compor o *dataset_4* (d4) (n = 1322). A **tabela 3** indica a quantidade de amostras usadas em cada estado.

Tabela 3 - Quantidade de genomas usadas na árvore de Máxima Verossimilhança temporal por estado do Brasil.

Estado (Sigla)	Nº de amostras
AM	71
BA	54
GO	68
PA	41
RR	24
SC	67
AP	51
RS	70
PB	40
AC	5
SE	33
CE	64
MA	49
RN	43
RO	24
MT	30
TO	36
PR	61
ES	46
AL	77
MG	61
RJ	80
PI	26
MS	39
PE	47
SP	88
DF	27
Total	1322

Fonte: Elaborado pelo autor.

A filogeografia por máxima verossimilhança temporal foi realizada com os dados d4 usando a versão online do programa Treetime (SAGULENKO; PULLER; NEHER, 2018), usando uma taxa de substituição de nucleotídeos por sítio por ano de $8-10 \times 10^{-4}$. Ademais, utilizou-se a versão online do programa PASTML (ISHIKAWA et al., 2019) para se reconstruir o caráter ancestral das localizações da epidemia, usando as configurações de aproximação das probabilidades a posteriori marginais e o modelo F81. A dedução dos eventos de importação e exportação foi feita baseado na mudança de estados (cores) das regiões do Brasil para o Ceará

e do Ceará para o Brasil, respectivamente. Por fim, a árvore resultante foi então visualizada e colorida no programa Figtree.

4.11. Frequência de mutações e *Minor variants*

Para a análise da frequência mutacional da Variante de Preocupação Gama, o *dataset* d1 foi filtrado para compor somente as amostras cuja variante ficou definida como Gama (n = 1792), segundo a aplicação online Nextclade v.1.8.1, compondo os meses de janeiro a setembro de 2021.

Para compor o *dataset* de genomas minoritários, foram recolhidas as versões minoritárias dessas mesmas amostras filtradas do *dataset* d1. Como alguma das amostras não possuíam genoma minoritário, a quantidade desses genomas ficou levemente reduzida em relação ao majoritário (n = 1785).

Utilizando-se de um algoritmo interno (**pseudocódigo 2**), verificou-se a frequência (F) de cada base no genoma (**equação 01**) para cada mês, ajustando o valor da frequência ao se retirar a quantidade de bases ambíguas (N) do total para cada posição:

$$F = \frac{N^{\circ} \text{ de bases mutantes iguais na posição } i}{\text{Total de bases na posição } i - \text{Total de bases ambíguas para posição } i} \quad (01)$$

A análise foi feita tanto para as amostras comuns quanto para os genomas minoritários, tendo como frequência de corte 30%, a fim de evitar mutações oriundas de artefatos de sequenciamento.

A anotação das mutações é feita de duas formas, a depender se a mutação é sinônima ou não. Para mutações não sinônimas segue o padrão Gene:APOSB, onde: Gene representa o respectivo gene presente no genoma viral onde a mutação ocorreu, A o aminoácido original, POS a posição aminoacídica com o referencial estabelecido no início do gene, e B o aminoácido novo. Para mutações sinônimas segue o padrão Gene:XPOS'Y, onde: Gene representa o respectivo gene presente no genoma viral onde a mutação ocorreu, X o nucleotídeo original, POS' a posição nucleotídica com o referencial estabelecido no início do genoma, e Y o nucleotídeo novo.

4.12. Dados associados

4.10.1 Número de voos por região

O número de voos foi recolhido, no dia 23 de novembro de 2021, no portal da Agência Nacional de Aviação Civil (ANAC) (<https://sas.anac.gov.br/sas/bav/view/frmConsultaVRA>) no período de janeiro de 2020 até 31 de outubro de 2021 com viagens do Brasil para o aeroporto Pinto Martins (Fortaleza). É importante ressaltar que devido a cidade de Fortaleza ser periférica,

nem todas as regiões possuem voos diretos. Desse modo, as regiões Centro-Oeste e Sudeste estão acrescidas dos voos provenientes de cidades do Sul.

4.10.2. Índice de permanência domiciliar

Os dados são provenientes do programa *Google Mobility Report* e os cálculos foram realizados segundo descrito em (BARRETO, 2020). Os dados contemplam as macrorregiões do estado do Ceará até o dia 02 de dezembro de 2021.

O Índice de permanência domiciliar, apesar de ser calculado por meio de porcentagens, não corresponde, de fato, a uma. Portanto, um valor positivo indica uma menor circulação de pessoas.

4.10.3. Vacinação

Os dados referentes a data e a quantidade de segundas doses aplicadas foram retirados do sítio <https://opendatasus.saude.gov.br/dataset>, visitado no dia 02/12/2021.

Devido a possíveis diferenças em bancos de dados, pode haver discordâncias entre os dados aqui apresentados e os presentes em outros repositórios.

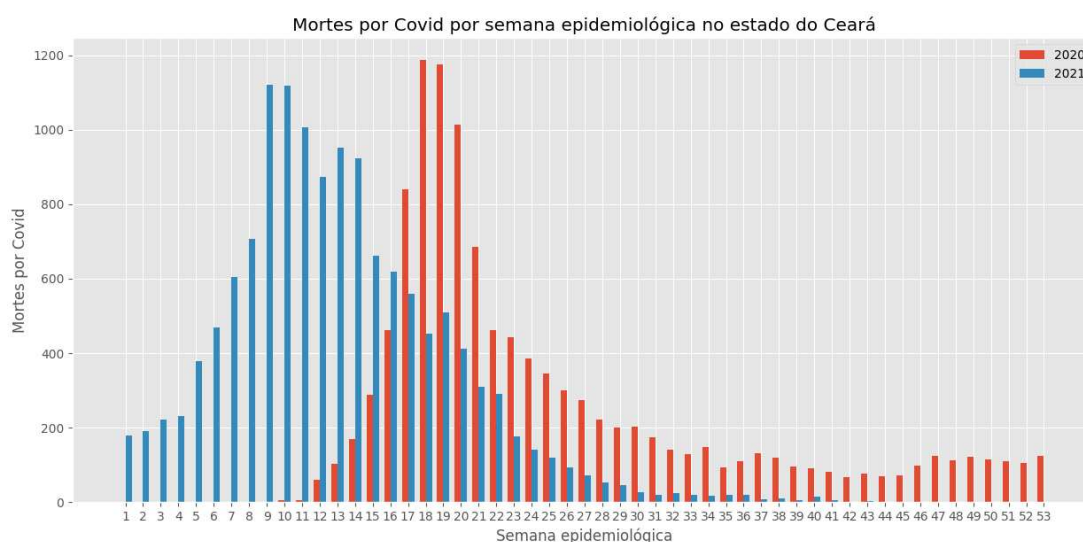
Foram agregadas, por dia, a quantidade de segundas doses aplicadas no estado do Ceará no ano de 2021. Em seguida, o valor referente a cada dia foi dividido pela população cearense estimada pelo IBGE (<https://www.ibge.gov.br/>) para 2021 e multiplicada por 100 para obter o valor de porcentagem de pessoas vacinadas.

5. RESULTADOS E DISCUSSÃO

5.1. Evolução de casos no Ceará

Segundo os dados que compõem o **gráfico 1**, o número total de mortes confirmadas por SARS-CoV-2 no Ceará em 2020 foi de 11344 e média de 214,04 mortes por semana, enquanto no ano de 2021 até 43ª semana epidemiológica foi de 13679 e média de 318,12 por semana, com os respectivos picos de mortes com 1187 na 18ª semana (abril) de 2020 e com 1121 na 9ª semana (fevereiro) de 2021. O aumento no número de casos no Ceará foi similar ao que aconteceu em outros estados do Nordeste, apresentando um crescimento alarmante, principalmente, no final de 2020 e início de 2021 (KERR et al., 2021).

Gráfico 1 - Distribuição do número de óbitos por SARS-CoV-2 no estado do Ceará nos anos de 2020 e 2021, separado por semana epidemiológica



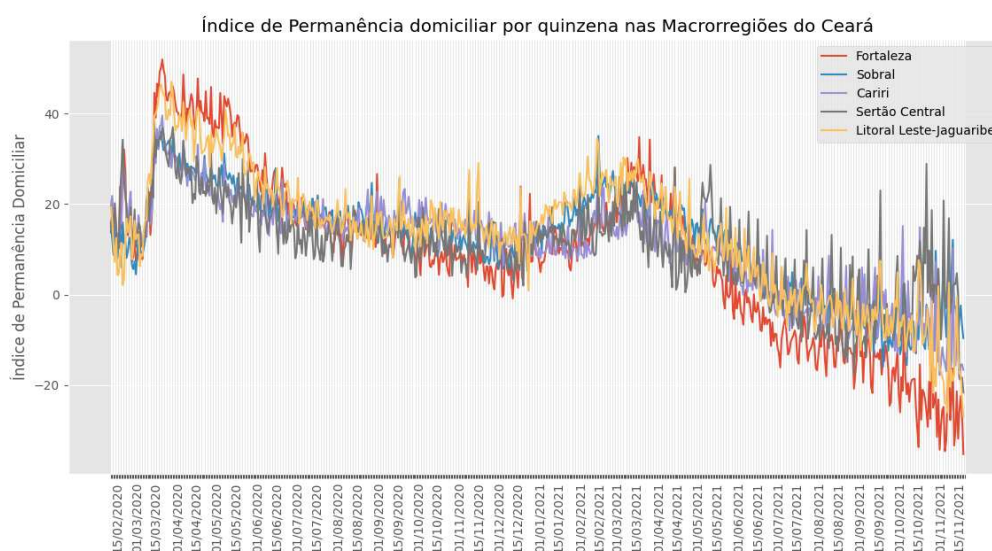
Fonte: Elaborado pelo autor com os dados presentes em <https://opendatasus.saude.gov.br/>.

Acesso em 13 de novembro de 2021

A menor mortalidade observada na primeira onda da pandemia pode ser explicada por um conjunto de fatores, tal como o decreto realizado pelo poder executivo da cidade de Fortaleza e do estado do Ceará no início de março de 2020, o qual instituiu medidas de isolamento rígidas: o controle da circulação de pessoas e veículos; controle de entrada e saída dos municípios; e o dever de permanência domiciliar (CEARÁ, 2020). Ao longo do ano do início da pandemia, outros decretos foram estabelecidos para aumentar o rigor do isolamento, afetando outros campos sociais. (CORONAVIRUS CEARÁ, 2021). Estas medidas foram responsáveis pelo pico de isolamento observado em diversas capitais do nordeste (KERR et al.,

2021), o qual também pode ser visto no pico presente nos meses de março a junho, principalmente nas macrorregiões de Fortaleza e Litoral Leste-Jaguaribe (**gráfico 2**). A manutenção do número de óbitos a partir da semana 29 até a semana 54 de 2020 representa a eficiência desta medida não farmacológica, quando comparada ao período de pico deste mesmo ano.

Gráfico 2 - Índice de Permanência Domiciliar (IPD) por quinzena nas Macrorregiões do estado do Ceará nos anos de 2020 e 2021.



Fonte: Elaborado pelo autor. Adaptado de BARRETO (2020)

Ademais, neste momento da pandemia no Brasil, as principais cepas circulantes eram pertencentes as linhagens B.1.1, B.1.1.28 e B.1.1.33 (PERICO et al., 2021). O perfil mutacional destas linhagens variam por somente uma ou duas mutações sinapomórficas (**tabela 4**), totalizando um máximo de 7 mutações definidoras de linhagem, no caso da B.1.1.33. Aliado a isso, como observado em (PERICO et al., 2021), é perceptível um equilíbrio entre essas três linhagens de março a setembro de 2020, revelando que nenhuma em especial possuía grande vantagem de transmissão em relação a outra. Este cenário passou a mudar no início de setembro, com a reabertura gradual de alguns espaços e atividades para até 100 pessoas (CEARÁ, 2020). Não obstante, o surgimento da linhagem P.1 (Variante de Preocupação Gama) em meados de novembro de 2020 vai de encontro com o que foi notificado até então, tendo em vista que esta nova linhagem apresenta 21 mutações definidoras (NAVECA; SOUZA; COSTA, 2021).

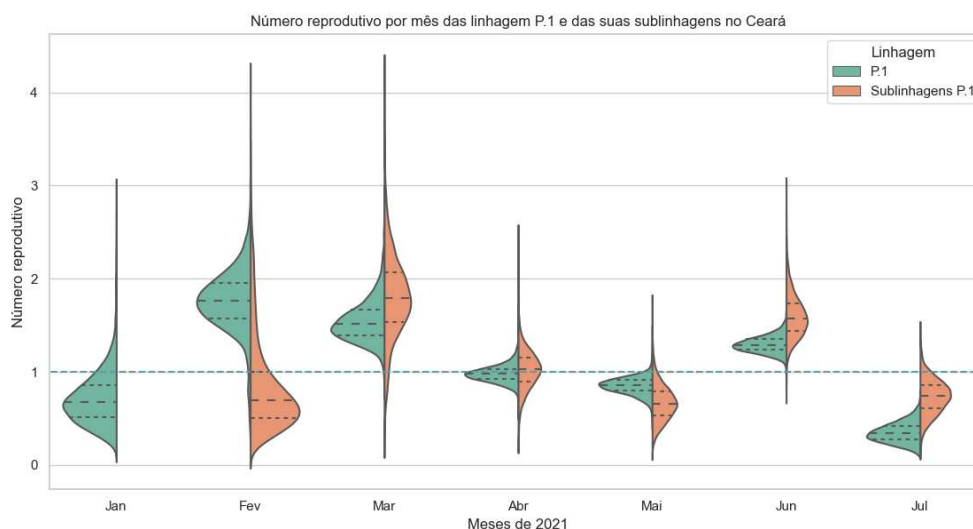
Tabela 4 - Mutações definidoras das linhagens B.1.1, B.1.1.28 e B.1.1.33 separadas por

	Mutações							
	ORF1b P314L	S D614G	S V1176F	ORF6 I33T	ORF8 S84L	N R203K	N G204R	N I292T
B.1.1	Sim	Sim	Não	Não	Sim	Sim	Sim	Não
B.1.1.28	Sim	Sim	Sim	Não	Sim	Sim	Sim	Não
B.1.1.33	Sim	Sim	Não	Sim	Sim	Sim	Sim	Sim

Fonte: Elaborado pelo autor com dados disponíveis em <https://outbreak.info/>

O **gráfico 3**, o qual representa o número reprodutivo da linhagem P.1 e de suas sublinhagens ao longo do tempo, indica os meses em que a mediana da distribuição a posteriori ficou superior a 1 (aumento da dispersão do vírus), sendo estes os meses de fevereiro, março e junho de 2021, ao passo que, para as sublinhagens da Gama, foram os meses de março, abril e junho.

Gráfico 3 - Distribuição do número reprodutivo das linhagens P.1 e de suas sublinhagens no estado do Ceará para os meses de janeiro a julho de 2021.



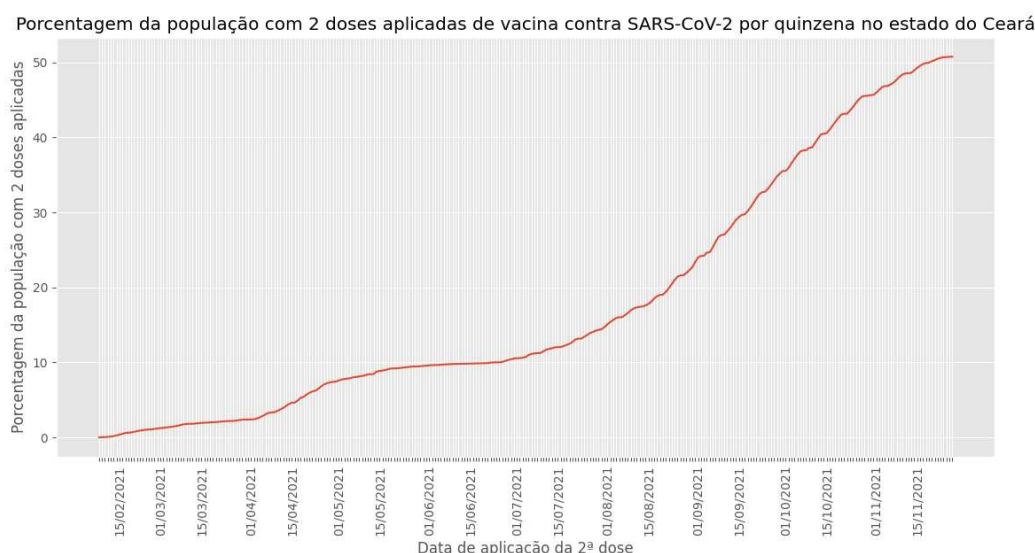
Fonte: Elaborado pelo autor.

O aumento da dispersão do vírus pode ser associado com a maior transmissibilidade, até 2,6 vezes maior, e capacidade de reinfecção presente na linhagem P.1 (COUTINHO et al., 2021), bem como com a abertura progressiva da economia e de outras atividades com números crescentes de pessoas desde 2020 (CEARÁ, 2020), além do descumprimento, por parte da população, dos decretos estabelecidos, principalmente, na capital no período de fim de ano, como pode ser observado no **gráfico 2**. Ainda neste gráfico, é perceptível a redução da mobilidade dos indivíduos, aumento no índice, nos períodos que compreendem as duas ondas de casos e óbitos, seguido de uma queda no valor do índice. Os períodos de maior

distanciamento social foram nos meses de março a maio de 2020 e fevereiro a abril de 2021. Ao mesmo tempo, os períodos de menos adesão a essa medida foram nos meses de junho de 2020 a janeiro de 2021 e maio a novembro de 2021.

Neste intervalo de meses, principalmente na capital, onde se teve o maior número de casos (INTEGRASUS, 2021), é perceptível o aumento posterior no valor do número reprodutivo, tanto da linhagem P.1, quanto de suas sublinhagens. Neste período a vacinação (**gráfico 4**) no estado encontrava-se em um platô, em torno de 10% da população vacinada com as duas doses. É relevante especificar, ainda, que o indivíduo é considerado como totalmente vacinado somente cerca de 2 semanas após a aplicação da segunda dose, como utilizado em outros trabalhos (NANDURI et al., 2021; SELF et al., 2021). Deste modo, encontrava-se um cenário propenso para o possível ressurgimento da Variante Gama. Contudo, como pode ser observado no **gráfico 5**, é no mês de julho onde começam a ser sequenciadas as primeiras amostras, em grande quantidade, da Variante de preocupação Delta, a qual possui uma capacidade de transmissão ainda maior que a Gama (ALIZON et al., 2021; CAMPBELL et al., 2021). Ao ponderar sobre esse fato, a substituição da P.1 e de suas sublinhagens em julho deve-se a soma de dois fatores, primordialmente: ao aumento no número de pessoas vacinadas com as duas doses; e ao surgimento de uma linhagem mais transmissível.

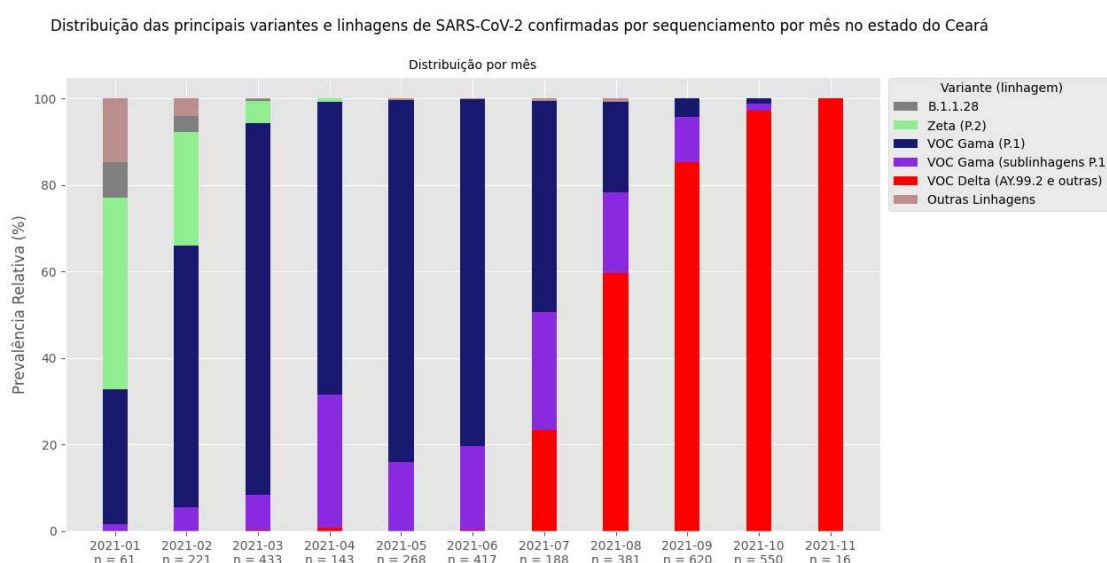
Gráfico 4 - Porcentagem da população com 2 doses aplicadas de vacina contra SARS-CoV-2 por quinzena no estado do Ceará no ano de 2021.



Fonte: Elaborado pelo autor com os dados disponíveis em <https://opendatasus.saude.gov.br/dataset>. Acesso em 02 de dezembro de 2021.

A capacidade de evasão do sistema imune de pessoas já contaminadas previamente por parte da linhagem P.1 podia ser teorizada a partir da incomum quantidade de mutações acumuladas (NAVECA et al., 2021a), mas pôde ser observada na prática, uma vez que esta linhagem afetou duramente o estado do Amazonas, o qual já possuía grande parte da sua população infectada até outubro de 2020 (BUSS et al., 2021).

Gráfico 5 - Distribuição das principais variantes e linhagens de SARS-CoV-2 confirmadas por sequenciamento por mês no estado do Ceará no período de janeiro a novembro de 2021.



Fonte: Elaborado pelo autor

O atraso no plano de vacinação (KERR et al., 2021) conjuntamente com as outras peculiaridades descritas anteriormente formulam a hipótese que explica o aumento vertiginoso de óbitos observados no início de 2021.

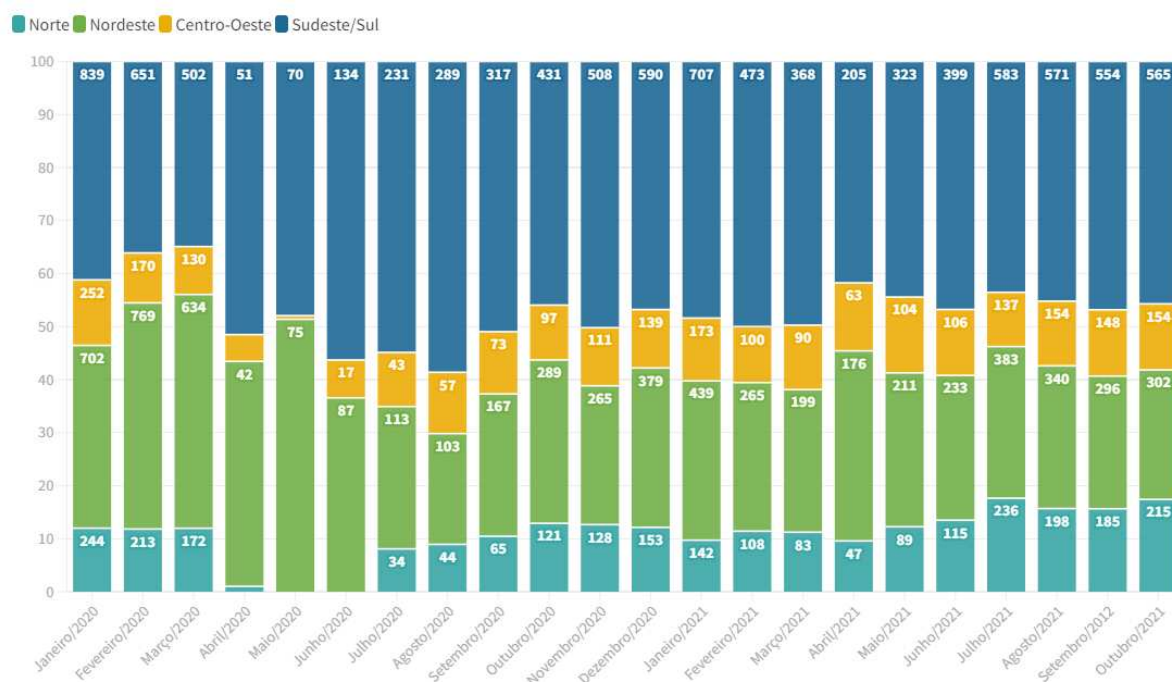
Ao final do ano, nos meses de agosto a outubro de 2021, semanas 27 a 43, um rápido aumento no percentual de pessoas completamente vacinadas é acompanhado de uma diminuição do número de óbitos. Este comportamento é previsível, já que, como demonstrado por (NANDURI et al., 2021; NASREEN et al., 2021; THOMPSON et al., 2021) a resposta imunológica proporcionada pela vacina consegue ser mais efetiva e duradoura, evitando o possível surgimento de uma terceira onda, provocada pelo surgimento de outras Variantes.

5.2. Surgimento da linhagem P.1 no Ceará

O sequenciamento da linhagem P.1 já podia ser observado em janeiro de 2021, onde aumentou sua prevalência até o mês de julho, quando suas sublinhagens e a Variante de Preocupação Delta começou a substituí-la (**gráfico 5**).

A substituição das linhagens P.2, B.1.1.28, dentre outras, pela P.1 no período de 4 meses, no começo do ano, pode ser analisada de um ponto de vista epidemiológico devido: ao grande número de voos oriundos da região Norte do País (**gráfico 6**), local onde a linhagem primeiramente surgiu (NAVECA et al., 2021a), no período de novembro de 2020 a fevereiro de 2021 (Número de voos = 531), por serem considerados como alta temporada; o transporte de pacientes, necessário para aliviar o sistema de saúde de Manaus (ADRIANO, 2021); o índice de distanciamento social abaixo de 45% no estado neste mesmo período (KERR et al., 2021); e a uma queda no índice de permanência domiciliar (**gráfico 2**). Deste modo, um cenário favorável para uma cepa que apresenta uma carga viral até 10 vezes maior do que as não-P.1 (NAVECA et al., 2021a) e que, assim como outras Variantes de Preocupação à época, possuía capacidade de evasão da neutralização por anticorpos (DEJNIRATTISAI et al., 2021; WANG et al., 2021).

Gráfico 6 - Número de voos do Brasil para o aeroporto de Fortaleza Pinto Martins para os anos de 2020 e 2021.



Fonte: Elaborado pelo autor com dados de <https://www.anac.gov.br/>. Acesso em 23 de novembro de 2021.

Contudo, a análise da data de surgimento do Ancestral Comum Mais Recente (**tabela 5**) indica que a primeira introdução feita no Ceará aconteceu ainda no final de dezembro de 2020. Embora ainda houvesse certa limitação para algumas atividades e eventos, em dezembro

de 2020 ainda se continuava o processo de reabertura gradual da economia cearense (CEARÁ, 2020), favorecendo a dispersão dessa Variante. A mesma análise para as sublinhagens indicou sua introdução no início de janeiro de 2021

Tabela 5 - Datas estimadas de introdução da Linhagem P.1 e de suas sublinhagens por reconstrução da data do Ancestral Comum mais Recente.

Variante	Mediana	HPD	Data
P.1	2020,978	2020,8578 - 2021,0195	24/12/2020 (10/11/2020 - 08/01/2021)
Sublinhagens P.1	2021,013	2020,8567 - 2021,0728	05/01/2021 (09/11/2020 - 27/01/2021)

Fonte: Elaborado pelo autor

A observação da árvore de Máxima Verossimilhança Temporal (**figura 2**) com os nós ancestrais reconstruídos, levando-se em consideração os estados do Brasil, permite a inferência de 40 eventos de importação (Norte = 17; Nordeste = 17; Sudeste = 5; Sul = 1) para o Ceará e 6 de exportação (Nordeste = 3; Sudeste = 2; Sul = 1) para o Brasil (**tabela 6**).

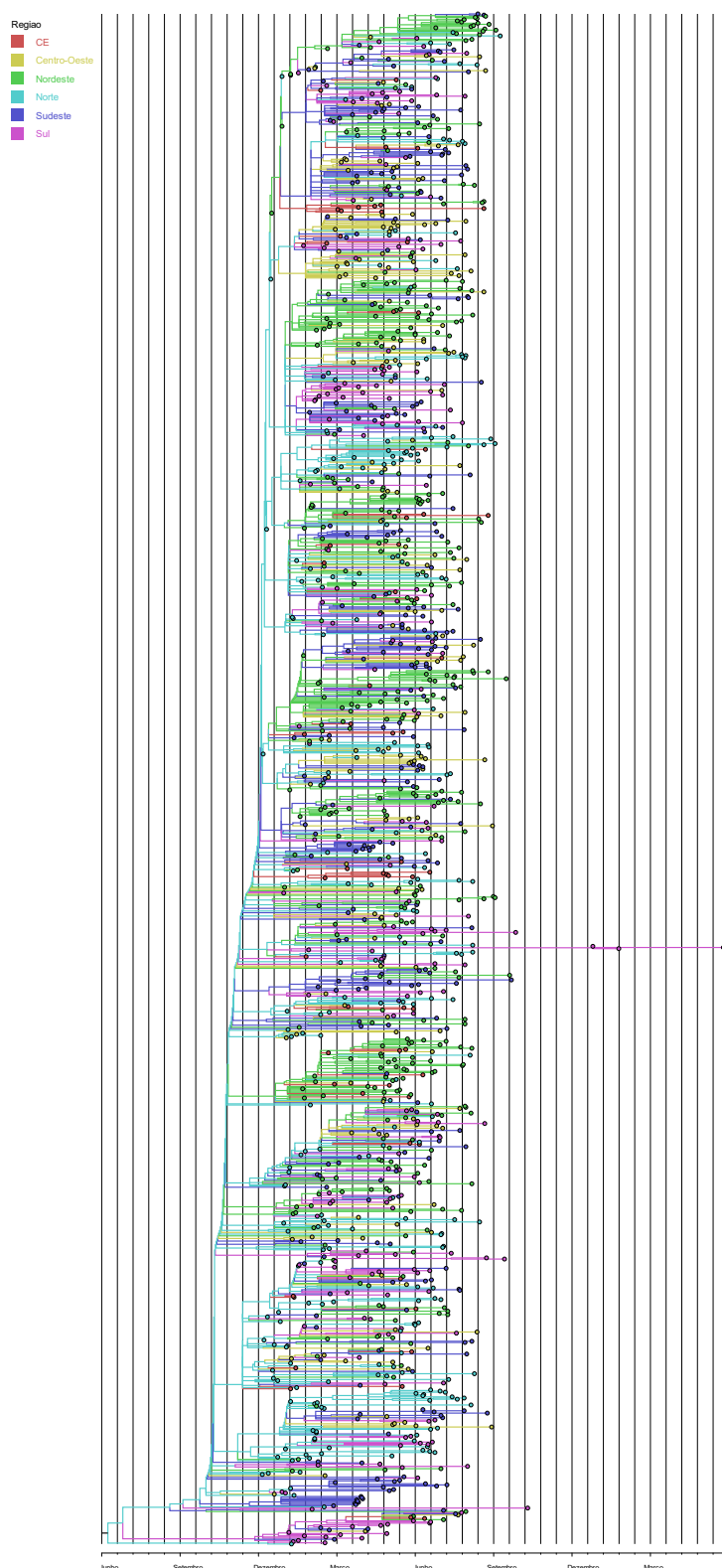
Tabela 6 - Número de importações das regiões brasileiras para o estado do Ceará e de exportações do Ceará para as regiões brasileiras.

Região	Importação	Exportação
Norte	17	0
Nordeste	17	3
Centro-Oeste	0	0
Sudeste	5	2
Sul	1	1
Total		

Fonte: Elaborado pelo autor

Tendo isto em vista, é estimada uma forte contribuição dessas regiões para o início e estabelecimento da Variante Gama em território cearense, ao mesmo tempo em que a baixa quantidade de exportações a partir do Ceará aponta para uma baixa participação na transmissão para outros estados. É pertinente esclarecer que durante o processo de incorporação da variável temporal na árvore de máxima verossimilhança, parâmetros de confiabilidade como *bootstrap* e baseados no procedimento do tipo Shimodaira-Hasegawa não estão mais presentes. Esta constatação indica a necessidade de cautela ao se interpretar estes dados.

Figura 2 - Árvore de Máxima Verossimilhança temporal usando o dataset_4 com amostras P.1 do Brasil, separados por região e Ceará em 2020 e 2021.

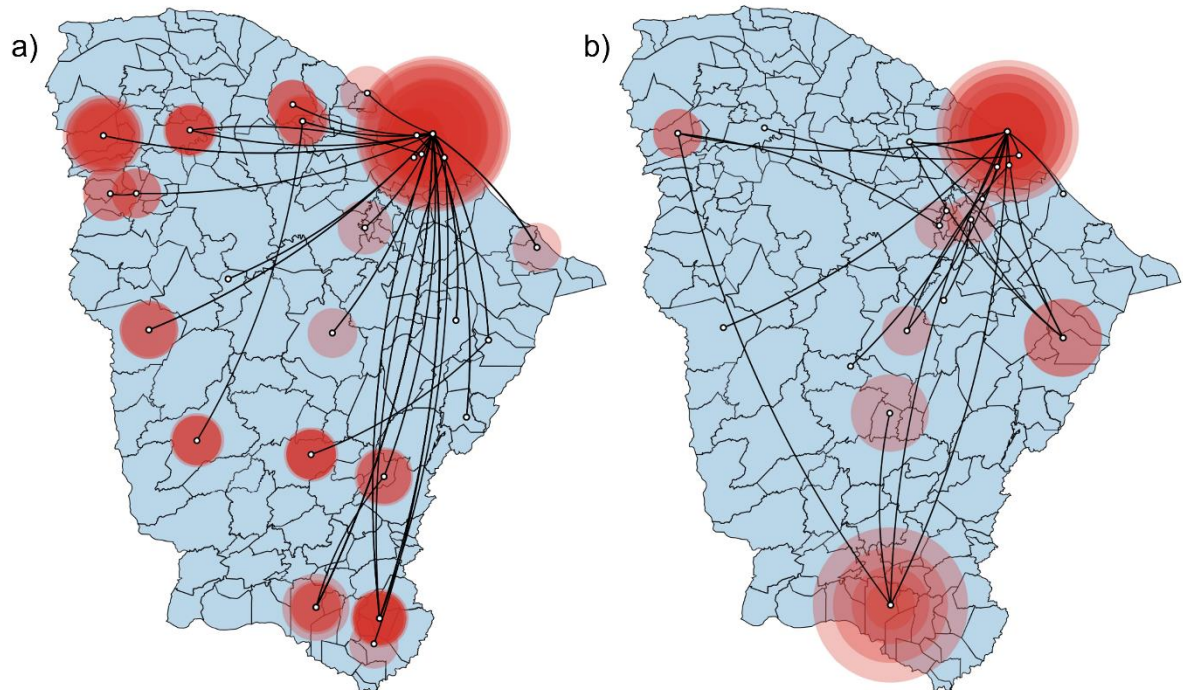


Fonte: Elaborado pelo autor com dados de <https://www.gisaid.org/>.

5.3. Distribuição geográfica

A análise de filogeografia Bayesiana indica a cidade de Fortaleza como a mais provável fonte do início de contágio para os demais municípios do Ceará, tanto quando se considera a linhagem P.1, probabilidade a posteriori = 1, quanto para as suas sublinhagens, probabilidade a posteriori = 1. A **figura 3** indica o fluxo de transmissão, saindo da capital para o restante do Ceará. Contudo, devido à ausência de suporte nos ramos que relacionam os municípios do interior entre si, não se pode identificar se houve, dentre as outras cidades, outro *hub* de exportação. A baixa confiança apresentada por esses ramos pode ser explicada devido à baixa diversidade genética esperada presente dentro do grupo correspondente à linhagem P.1, uma vez que a ideia de mutações definidoras para cada linhagem é uma regra para a sua classificação (O'TOOLE et al., 2021a). Outro fator foi devido ao uso somente da proteína spike, gene com grande número de mutações, o que diminui ainda mais a diversidade genética do grupo de amostras, para permitir que as análises pudessem ser executadas, já que análises bayesianas demandam longos períodos computacionais para serem executadas (SAGULENKO; PULLER; NEHER, 2018). Embora as sublinhagens da P.1 apresentem entre si uma diversidade maior, por pertencerem a grupos levemente diferentes, no conjunto de dados analisados, apenas as sublinhagens P.1.7 (n = 168) e P.1.10 (n = 109) sozinhas compunham 69,60% do total (n = 398) de sublinhagens sequenciadas. Logo, as mesmas limitações da técnica também incidem sobre esse segundo grupo.

Figura 3 - Fluxo de transmissão da Variante de Preocupação Gama entre os municípios do estado do Ceará por inferência Bayesiana: em a) a dispersão geográfica da linhagem P.1; em b) a dispersão geográfica das sublinhagens da P.1

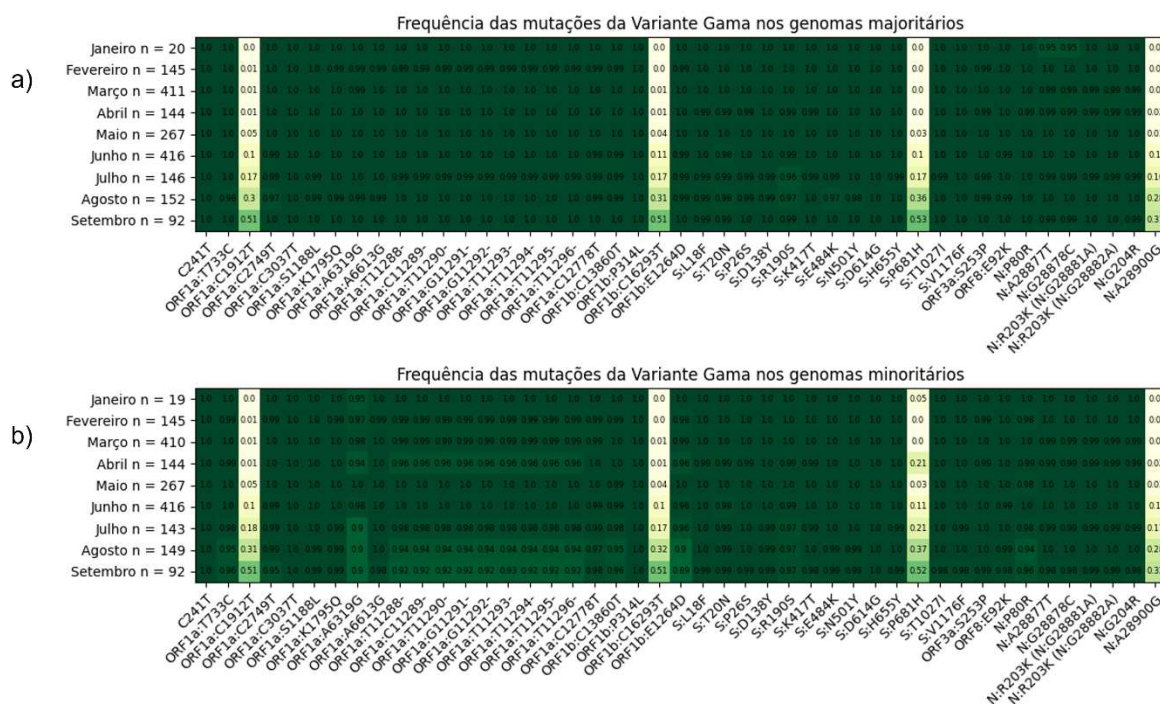


Fonte: Elaborado pelo autor.

5.4. Sublinhagens da P.1

Segundo o **gráfico 7a e 7b**, quatro mutações foram se acumulando na Variante Gama ao longo dos meses de janeiro a setembro de 2021 para os genomas majoritários: ORF1a:C1912T (0 – 51%); ORF1b:C16293T (0 – 51%); S:P681H (0 – 53%); e N:A28900G (0 – 31%). Frequências similares são observadas para os genomas minoritários: ORF1a:C1912T (0 – 51%); ORF1b:C16293T (0 – 51%); S:P681H (0 – 52%); e N:A28900G (0 – 32%).

Gráfico 7 - Frequência das mutações da Variante Gama: em a) para os genomas majoritários; em b) para os genomas minoritários



Fonte: Elaborado pelo autor

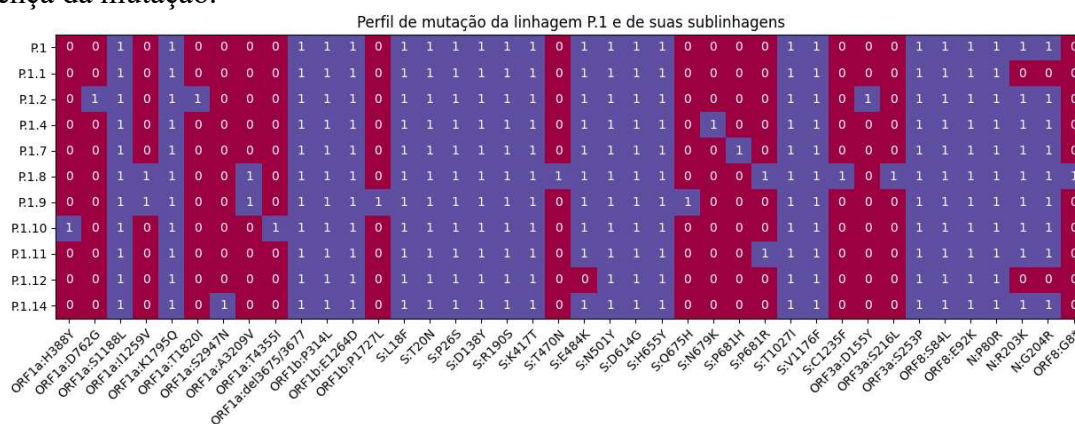
A hipótese a ser testada era a de que, devido a presença de *pools* de populações virais dentro de um mesmo indivíduo (GRUBAUGH et al., 2019), mutações que ocorressem minoritariamente dentro de um indivíduo poderiam se tornar dominantes em outro. Esta possibilidade abriria uma correspondência entre esses dois tipos de genomas. Contudo, as frequências em ambos os tipos de genoma são bastante similares.

Ademais, no **gráfico 7b**, flutuações na frequência em mutações definidoras da Variante Gama, como a deleção de 9 nucleotídeos na região ORF1a, possivelmente indicam a contaminação, por populações correspondentes a outras linhagens. Embora isto não comprometa a qualidade dos genomas majoritários, é plausível de acontecer devido a técnica inerente à detecção dessa diversidade genética, já que a conversão de um genoma em outro é basicamente realizado ao se substituir as bases específicas por outras que estão em baixa frequência (DEZORDI et al., 2021).

Por outro lado, curiosamente, percebe-se a retenção de 3 mutações, dentre as 4 observadas, silenciosas, as quais não alteram o aminoácido correspondente devido ao código genético ser degenerado.

Dentre as mutações com aumento de frequência, somente a mutação S:P681H foi destacada. Esta mutação é uma sinapomorfia que distingue a linhagem P.1.7 de sua ancestral P.1 (**gráfico 8**). Como pode ser visto no **gráfico 9**, a frequência desta sublinhagem aumentou a partir do mês de maio (3,01%) até setembro (50%), explicando a sua participação no **gráfico 7a**. Por outro lado, embora a linhagem P.1.10 também possua um aumento em sua frequência a partir do mês de fevereiro (2,78%) até abril (25,71%), devido a linha de corte estabelecida para que as mutações sejam apresentadas (30%), suas mutações definidoras de linhagem (ORF1a:H388Y; ORF1a:T4355I) não são observadas.

Gráfico 8 - Perfil de mutação da linhagem P.1 e de suas sublinhagens. O eixo x representa as mutações características de cada linhagem, onde a cor vinho (0) representa a ausência e lilás (1) a presença da mutação.

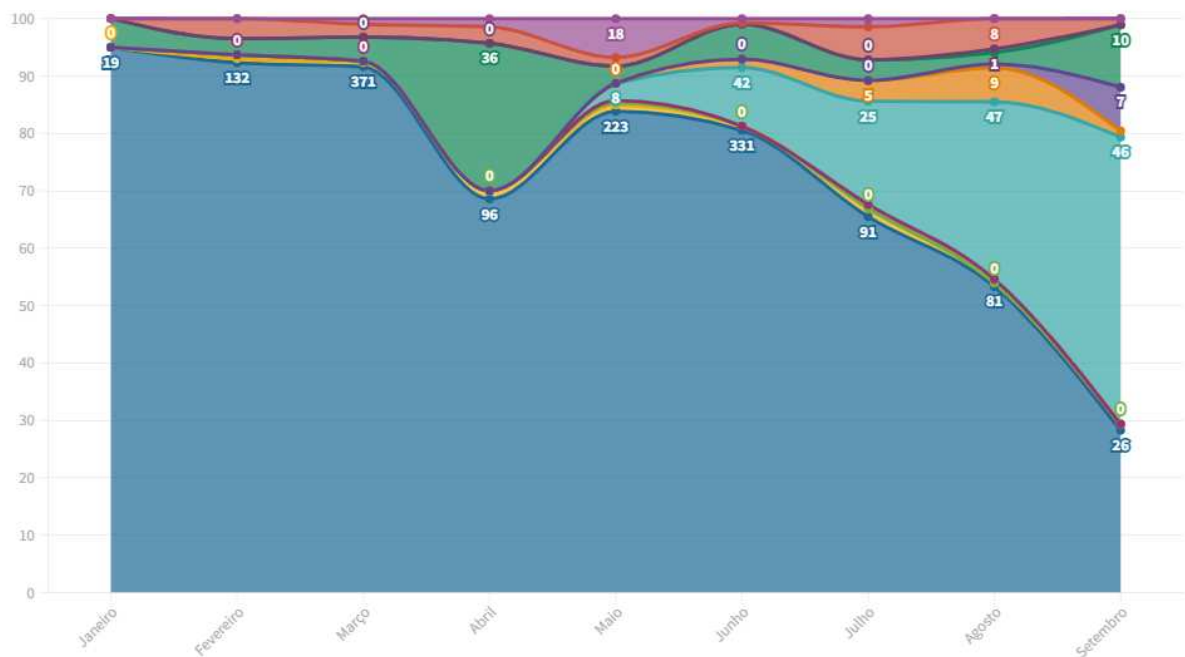


Fonte: Elaborado pelo autor com dados de <https://outbreak.info/>.

A mutação (S P681H) que difere a linhagem P.1.7 de sua ancestral P.1 está localizada na região de clivagem S1/S2, também presente na Variante de Preocupação Alfa (B.1.1.7) a qual foi primeiramente detectada em setembro de 2020 (CHOI; SMITH, 2021). Esta mutação é capaz de facilitar levemente o processo de clivagem das regiões S1 e S2 da proteína *spike*, embora não se tenha identificado um maior nível de infecção celular envolvido (LUBINSKI et al., 2021). Na sublinhagem P.1.10, a ORF1a:H388Y ocorre na proteína não estrutural 2 (nsp2), a qual sua atuação é sugerida como sendo a perturbação do ambiente celular (YOSHIMOTO, 2020), além de cooperar com a proteína não estrutural 4 (nsp4) na replicação viral (BANERJEE et al., 2021). Enquanto isso, ainda nesta sublinhagem, a ORF1a:T4355I está presente na proteína não estrutural 10 (nsp10), responsável por estimular a ação de duas outras proteínas não estruturais: 14 e 16. O complexo nsp10-nsp14 forma uma exorribonuclease importantíssima para manter a fidelidade da replicação viral, ao mesmo tempo em que também

é responsável por excisar inibidores análogos a nucleotídeos incorporados ao RNA em construção, prejudicando antivirais que possuam essa tecnologia como modo de ação (LIU et al., 2021). Em relação ao complexo nsp10-nsp16, este é essencial para a prevenir e reduzir a resposta imunológica do hospedeiro durante a tradução do RNA viral (ROSAS-LEMUS et al., 2020).

Gráfico 9 - Frequência da linhagem P.1 e de suas sublinhagens por meses no ano de 2021.



Fonte: Elaborado pelo autor.

Embora essas mutações possam potencial para aumentar a resistência contra antivirais e para auxiliar no processo de transcrição e tradução, estas sublinhagens foram prontamente substituídas pela Variante Delta ao redor do mundo (TEGALLY et al., 2021).

6. CONCLUSÃO

Portanto, o presente estudo contribui para uma melhor compreensão de como o agente etiológico da COVID-19, mais especificamente a Variante de Preocupação Gama, se comportou no estado do Ceará desde janeiro até novembro de 2021. É destacável a introdução da linhagem P.1 já no final de dezembro de 2020, ocasionada possivelmente pelo aumento no fluxo voos comerciais no período de novembro de 2020 a fevereiro 2021 da região Norte do País e, posteriormente, outras introduções principalmente da região Norte e Nordeste. Ademais, as sublinhagens de P.1 já detectadas em janeiro de 2021 assumiram brevemente a prevalência relativa no estado até serem substituídas pela Variante de Preocupação Delta, a qual é ainda mais transmissível. De um modo geral, o aumento no número de casos e o surgimento de Variantes com maior capacidade de disseminação parece estar associado com as quedas nos índices de permanência domiciliar enquanto a cobertura vacinal ainda se encontrava no seu estágio inicial. Neste cenário, sendo a capital de Fortaleza a fonte de exportação da Variante Gama (P.1 e suas sublinhagens) para os outros municípios do estado. Não obstante, estudos posteriores são de suma importância para entender melhor a dinâmica de substituição da Variante Gama pela Variante Delta, além de continuar com as metodologias de vigilância em busca de mutações que representem potenciais novas linhagens cada vez mais transmissíveis e capazes de contornar o sistema imune. Por fim, este estudo buscou preencher uma lacuna a respeito da história evolutiva natural do vírus SARS-CoV-2, linhagem e sublinhagens P.1, e suas possíveis particularidades desde o final de 2020 até o final de 2021.

7. REFERÊNCIAS

WORLD HEALTH ORGANIZATION (WHO). **Tracking SARS-CoV-2 variants**. 2021. Disponível em: < <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>>. Acesso em: 28 de dez. de 2021.

GOVERNO DO ESTADO DO CEARÁ. **Documentos oficiais**. Disponível em: <<https://coronavirus.ceara.gov.br/profissional/documentos/>>. Acesso em: 28 de dez. 2021.

CEARÁ. **Decreto nº 33.608, de 30 de maio de 2020**. PRORROGA O ISOLAMENTO SOCIAL NO ESTADO DO CEARÁ, NA FORMA DO DECRETO Nº33.519, DE 19 DE MARÇO DE 2020, E INSTITUI A REGIONALIZAÇÃO DAS MEDIDAS DE ISOLAMENTO SOCIAL, E DÁ OUTRAS PROVIDÊNCIAS. Fortaleza: Casa Civil, [2020]. Disponível em: <https://coronavirus.ceara.gov.br/project/decreto-no-33-608-prorroga-isolamento-social-no-ceara-e-institui-a-regionalizacao-das-medidas-de-isolamento-social/>. Acesso em: 28 dez. 2021.

CEARÁ. **Decreto nº 33.846, de 12 de dezembro de 2020**. PRORROGA O ISOLAMENTO SOCIAL NO ESTADO DO CEARÁ, RENOVA A POLÍTICA DE REGIONALIZAÇÃO DAS MEDIDAS DE ISOLAMENTO SOCIAL, E DÁ OUTRAS PROVIDÊNCIAS. Fortaleza: Casa Civil, [2020]. Disponível em: <https://www.legisweb.com.br/legislacao/?id=405887>. Acesso em: 28 dez. 2021.

CEARÁ. **Decreto nº 33.730, de 29 de agosto de 2020**. PRORROGA O ISOLAMENTO SOCIAL NO ESTADO DO CEARÁ, RENOVA A POLÍTICA DE REGIONALIZAÇÃO DAS MEDIDAS DE ISOLAMENTO SOCIAL, E DÁ OUTRAS PROVIDÊNCIAS. Fortaleza: Casa Civil, [2020]. Disponível em: <https://coronavirus.ceara.gov.br/project/decreto-prorroga-medidas-de-isolamento-social-no-ceara-ate-6-de-setembro/>. Acesso em: 28 dez. 2021.

CEARÁ. **Decreto nº 33.510, de 16 de março de 2020**. DECRETA SITUAÇÃO DE EMERGÊNCIA EM SAÚDE E DISPÕE SOBRE MEDIDAS PARA ENFRENTAMENTO E CONTENÇÃO DA INFECÇÃO HUMANA PELO NOVO CORONAVÍRUS. Fortaleza: Casa Civil, [2020]. Disponível em: <https://coronavirus.ceara.gov.br/project/decreto-no-33-510-de-16-de-marco-de-2020/>. Acesso em: 28 dez. 2021.

GOVERNO DO ESTADO DO CEARÁ. **Cenário Epidemiológico COVID-19**. 2021. Disponível em: <<https://integrasus.saude.ce.gov.br/#/indicadores/indicadores-coronavirus/cenario-epidemiologico-covid-19>> . Acesso em: 28 de dez. 2021

LAVOR, Adriano de. Amazônia sem respirar: falta de oxigênio causa mortes e revela colapso em Manaus. **RADIS: Comunicação e Saúde**, n.221, p.20-23, fev. 2021.

ALEXEI J. DRUMMOND; REMCO R. BOUCKAERT. **Bayesian Evolutionary Analysis with BEAST**. [s.l: s.n.].

ALIZON, Samuel; HAIM-BOUKOBZA, Stéphanie; FOULONGNE, Vincent; VERDURME, Laura; TROMBERT-PAOLANTONI, Sabine; LECORCHE, Emmanuel; ROQUEBERT, Bénédicte; SOFONEA, Mircea T. Rapid spread of the SARS-CoV-2 Delta variant in some French regions, June 2021. **Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin**, [S. l.], v. 26, n. 28, p. 1–5, 2021. DOI: 10.2807/1560-7917.ES.2021.26.28.2100573. Disponível em: <http://dx.doi.org/10.2807/1560-7917.ES.2021.26.28.2100573>.

AMATO-LOURENÇO, Luís Fernando; DE SOUZA XAVIER COSTA, Natália; DANTAS, Kátia Cristina; LOMBARDI, Suzette Cleuza Ferreira Spina; JÚNIOR, Alfredo Mendroni; LINDOSO, José Angelo Lauletta; LIMA, Felipe Gallego; CARVALHO-OLIVEIRA, Regiani; MAUAD, Thais. Quantification of airborne SARS-CoV-2 genomic particles in different hospital settings. **Scientific Reports**, [S. l.], v. 11, n. 1, p. 1–6, 2021. DOI: 10.1038/s41598-021-00761-1. Disponível em: <https://doi.org/10.1038/s41598-021-00761-1>.

ANGELETTI, Silvia; BENVENUTO, Domenico; BIANCHI, Martina; GIOVANETTI, Marta; PASCARELLA, Stefano; CICCOCCHI, Massimo. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. **Journal of Medical Virology**, [S. l.], v. 92, n. 6, p. 584–588, 2020. DOI: 10.1002/jmv.25719.

ARONS, Melissa M. et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. **New England Journal of Medicine**, [S. l.], v. 382, n. 22, p. 2081–2090, 2020. DOI: 10.1056/nejmoa2008457.

BANERJEE, Abhik K. et al. SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. **Cell**, [S. l.], v. 183, n. 5, p. 1325–1339.e21, 2020. DOI: 10.1016/j.cell.2020.10.004. Disponível em: <http://dx.doi.org/10.1016/j.cell.2020.10.004>.

BANERJEE, Shuvam; SEAL, Sohan; DEY, Riju; MONDAL, Kousik Kr; BHATTACHARJEE, Pritha. Mutational spectra of SARS-CoV-2 orflab polyprotein and signature mutations in the United States of America. **Journal of Medical Virology**, [S. l.], v.

93, n. 3, p. 1428–1435, 2021. DOI: 10.1002/jmv.26417.

BAR-ON, Yinon M. et al. Protection of BNT162b2 Vaccine Booster against Covid-19 in Israel. **New England Journal of Medicine**, [S. l.], v. 385, n. 15, p. 1393–1400, 2021. DOI: 10.1056/nejmoa2114255.

BARRETO, Ivana Cristina de Holanda Cunha et al. Colapso na Saúde em Manaus : o fardo de não aderir às medidas não farmacológicas de redução da transmissão da COVID-19. **Scielo Preprints**, [S. l.], v. 18, n. versão 1, p. 20, 2020.

BARROS-MARTINS, Joana et al. Humoral and cellular immune response against SARS-CoV-2 variants following heterologous and homologous ChAdOx1 nCoV-19/BNT162b2 vaccination. **medRxiv**, [S. l.], p. 2021.06.01.21258172, 2021. DOI: 10.1101/2021.06.01.21258172. Disponível em: <https://www.medrxiv.org/content/10.1101/2021.06.01.21258172v1>.

BAZZAZPOUR, Shahriyar et al. The detection of SARS-CoV-2 RNA in indoor air of dental clinics during the COVID-19 pandemic. **Environmental Science and Pollution Research**, [S. l.], 2021. DOI: 10.1007/s11356-021-15607-6.

BEHJATI, Sam; TARPEY, Patrick S. What is next generation sequencing? **Archives of Disease in Childhood: Education and Practice Edition**, [S. l.], v. 98, n. 6, p. 236–238, 2013. DOI: 10.1136/archdischild-2013-304340.

BERARDI, Alberto; PERINELLI, Diego R.; MERCHANT, Hamid A.; BISHARAT, Lorina; BASHETI, Iman A.; BONACUCINA, Giulia; CESPI, Marco; PALMIERI, Giovanni F. Hand sanitisers amid CoViD-19: A critical review of alcohol-based products on the market and formulation approaches to respond to increasing demand. **International Journal of Pharmaceutics**, [S. l.], v. 584, n. April, p. 119431, 2020. DOI: 10.1016/j.ijpharm.2020.119431. Disponível em: <https://doi.org/10.1016/j.ijpharm.2020.119431>.

BOLZE, Alexandre et al. SARS-CoV-2 Variant Delta Rapidly Displaced Variant Alpha in the United States and Led to Higher Viral Loads. **SSRN Electronic Journal**, [S. l.], p. 1–11, 2021. DOI: 10.2139/ssrn.3952083.

BOON, Siaw S. et al. Seroprevalence of Unidentified SARS-CoV-2 Infection in Hong Kong during 3 Pandemic Waves. **JAMA Network Open**, [S. l.], v. 4, n. 11, p. 4–10, 2021. DOI: 10.1001/jamanetworkopen.2021.32923.

BOSON, Bertrand; LEGROS, Vincent; ZHOU, Bingjie; SIRET, Eglantine; MATHIEU, Cyrille; COSSET, François Loïc; LAVILLETTE, Dimitri; DENOLLY, Solène. The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein,

allowing assembly of virus-like particles. **Journal of Biological Chemistry**, [S. l.], v. 296, n. 14, p. 100111, 2021. DOI: 10.1074/jbc.RA120.016175. Disponível em: <https://doi.org/10.1074/jbc.RA120.016175>.

BOTELHO-SOUZA, Luan Felipo et al. SARS-CoV-2 genomic surveillance in Rondônia, Brazilian Western Amazon. **Scientific Reports**, [S. l.], v. 11, n. 1, p. 1–12, 2021. DOI: 10.1038/s41598-021-83203-2. Disponível em: <https://doi.org/10.1038/s41598-021-83203-2>.

BOUCKAERT, Remco et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. **PLoS Computational Biology**, [S. l.], v. 15, n. 4, p. 1–28, 2019. DOI: 10.1371/journal.pcbi.1006650.

BRAUNER, Jan M. et al. Inferring the effectiveness of government interventions against COVID-19. **Science**, [S. l.], v. 371, n. 6531, 2021. DOI: 10.1126/science.abd9338.

BRONNER, Iraad F.; QUAIL, Michael A. Best Practices for Illumina Library Preparation. **Current Protocols in Human Genetics**, [S. l.], v. 102, n. 1, p. 1–48, 2019. DOI: 10.1002/cphg.86.

BUSS, Lewis F. et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. **Science**, [S. l.], v. 371, n. 6526, p. 288–292, 2021. DOI: 10.1126/science.abe9728.

CAMPBELL, Finlay et al. Increased transmissibility and global spread of SARSCoV-2 variants of concern as at June 2021. **Eurosurveillance**, [S. l.], v. 26, n. 24, p. 1–6, 2021. DOI: 10.2807/1560-7917.ES.2021.26.24.2100509. Disponível em: <http://dx.doi.org/10.2807/1560-7917.ES.2021.26.24.2100509>.

CAO, Zengguo; XIA, Hongjie; RAJSBAUM, Ricardo; XIA, Xianzhu; WANG, Hualei; SHI, Pei Yong. Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response. **Cellular and Molecular Immunology**, [S. l.], v. 18, n. 3, p. 746–748, 2021. DOI: 10.1038/s41423-020-00603-6. Disponível em: <http://dx.doi.org/10.1038/s41423-020-00603-6>.

CEVIK, Muge; TATE, Matthew; LLOYD, Ollie; MARAOLO, Alberto Enrico; SCHAFERS, Jenna; HO, Antonia. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. **The Lancet Microbe**, [S. l.], v. 2, n. 1, p. e13–e22, 2021. DOI: 10.1016/S2666-5247(20)30172-5.

CHAN, Jasper Fuk-woo; KOK, Kin-hang; ZHU, Zheng; CHU, Hin; KAI-WANG, Kelvin; YUAN, Shuofeng; YUEN, Kwok-yung. Genomic characterization of the 2019 novel

human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *[S. l.]*, 2020. DOI: 10.1080/22221751.2020.1719902.

CHANG, Fengqi; LI, Marilyn M. Clinical application of amplicon-based next-generation sequencing in cancer. **Cancer Genetics**, *[S. l.]*, v. 206, n. 12, p. 413–419, 2013. DOI: 10.1016/j.cancergen.2013.10.003. Disponível em: <http://dx.doi.org/10.1016/j.cancergen.2013.10.003>.

CHEN, Chia Yen et al. Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. **Journal of Infectious Diseases**, *[S. l.]*, v. 196, n. 3, p. 405–415, 2007. DOI: 10.1086/519166.

CHEN, Paul Z.; BOBROVITZ, Niklas; PREMJI, Zahra; KOOPMANS, Marion; FISMAN, David N.; GU, Frank X. Heterogeneity in transmissibility and shedding SARS-CoV-2 via droplets and aerosols. **eLife**, *[S. l.]*, v. 10, p. 1–32, 2021. DOI: 10.7554/ELIFE.65774.

CHEN, Shifu; ZHOU, Yanqing; CHEN, Yaru; GU, Jia. Fastp: An ultra-fast all-in-one FASTQ preprocessor. **Bioinformatics**, *[S. l.]*, v. 34, n. 17, p. i884–i890, 2018. DOI: 10.1093/bioinformatics/bty560.

CHOI, Jun Yong; SMITH, Davey M. SARS-CoV-2 Variants of Concern. **Yonsei Medical Journal**, *[S. l.]*, v. 62, n. 11, p. 961, 2021. DOI: 10.3349/ymj.2021.62.11.961.

CHOWELL, Gerardo; DAHAL, Sushma; BONO, Raquel; MIZUMOTO, Kenji. Harnessing testing strategies and public health measures to avert COVID-19 outbreaks during ocean cruises. **Scientific Reports**, *[S. l.]*, v. 11, n. 1, p. 1–12, 2021. DOI: 10.1038/s41598-021-95032-4. Disponível em: <https://doi.org/10.1038/s41598-021-95032-4>.

COCK, Peter J. A.; FIELDS, Christopher J.; GOTO, Naohisa; HEUER, Michael L.; RICE, Peter M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. **Nucleic Acids Research**, *[S. l.]*, v. 38, n. 6, p. 1767–1771, 2009. DOI: 10.1093/nar/gkp1137.

CORNILLEZ-TY, Cromwell T.; LIAO, Lujian; YATES, John R.; KUHN, Peter; BUCHMEIER, Michael J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. **Journal of Virology**, *[S. l.]*, v. 83, n. 19, p. 10314–10318, 2009. DOI: 10.1128/jvi.00842-09.

COTTAM, Eleanor M.; WHELBAND, Matthew C.; WILEMAN, Thomas. Coronavirus NSP6 restricts autophagosome expansion. **Autophagy**, *[S. l.]*, v. 10, n. 8, p. 1426–1441, 2014. DOI: 10.4161/auto.29309.

COUTINHO, Renato Mendes; MARIA, Flavia; MARQUITTI, Darcie; SOUTO, Leonardo; EDUARDO, Marcelo. Model-based estimation of transmissibility and. *[S. l.]*, p. 1–3, 2021.

DAI, Wenhao et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. **Science**, *[S. l.]*, v. 368, n. 6497, p. 1331–1335, 2020. DOI: 10.1126/science.abb4489.

DANIELL, Henry et al. Debulking SARS-CoV-2 in saliva using angiotensin converting enzyme 2 in chewing gum to decrease oral virus transmission and infection. **Molecular Therapy**, *[S. l.]*, v. 30, n. 4, p. 1–13, 2021. DOI: 10.1016/j.ymthe.2021.11.008.

DE MAIO, Flavio et al. Improved binding of SARS-CoV-2 Envelope protein to tight junction-associated PALS1 could play a key role in COVID-19 pathogenesis. **Microbes and Infection**, *[S. l.]*, v. 22, n. 10, p. 592–597, 2020. DOI: 10.1016/j.micinf.2020.08.006. Disponível em: <https://doi.org/10.1016/j.micinf.2020.08.006>.

DEJNIRATTISAI, Wanwisa et al. Antibody evasion by the P.1 strain of SARS-CoV-2. **Cell**, *[S. l.]*, v. 184, n. 11, p. 2939–2954.e9, 2021. DOI: 10.1016/j.cell.2021.03.055.

DEMING, Damon J.; GRAHAM, Rachel L.; DENISON, Mark R.; BARIC, Ralph S. Processing of Open Reading Frame 1a Replicase Proteins nsp7 to nsp10 in Murine Hepatitis Virus Strain A59 Replication. **Journal of Virology**, *[S. l.]*, v. 81, n. 19, p. 10280–10291, 2007. DOI: 10.1128/jvi.00017-07.

DEZORDI, Filipe Zimmer; CAMPOS, Túlio de Lima; JERONIMO, Pedro Miguel Carneiro; AKSENEN, Cleber Furtado; ALMEIDA, Suzana Porto; WALLAU, Gabriel Luz. ViralFlow: an automated workflow for SARS-CoV-2 genome assembly, lineage assignment, mutations and intrahost variants detection. **medRxiv**, *[S. l.]*, 2021.

DÍAZ, José. SARS-CoV-2 Molecular Network Structure. **Frontiers in Physiology**, *[S. l.]*, v. 11, n. July, p. 1–8, 2020. DOI: 10.3389/fphys.2020.00870.

DOREMALEN, Neeltje Van et al. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. **The New England Journal of Medicine**, *[S. l.]*, p. 1564–1567, 2020.

DOUADY, Christophe J.; DELSUC, Frédéric; BOUCHER, Yan; DOOLITTLE, W. Ford; DOUZERY, Emmanuel J. P. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. **Molecular Biology and Evolution**, *[S. l.]*, v. 20, n. 2, p. 248–254, 2003. DOI: 10.1093/molbev/msg042.

FANG, Shou Guo; SHEN, Hongyuan; WANG, Jibin; TAY, Felicia P. L.; LIU, Ding

Xiang. Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells. **Virology**, [S. l.], v. 379, n. 2, p. 175–180, 2008. DOI: 10.1016/j.virol.2008.06.038.

FELTES, Bruno César; MAGALHÃES, Camila S. De; STAATS, Charley Christian; JUNQUEIRA, Dennis Maletich; BONATTO, Diego; YATES, Edwin A.; CUSTÓDIO, Fabio Lima. **Bioinformática da Biologia à Flexibilidade Molecular**. [s.l: s.n.].

FLAXMAN, Seth et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. **Nature**, [S. l.], v. 584, n. 7820, p. 257–261, 2020. DOI: 10.1038/s41586-020-2405-7.

FU, Limin; NIU, Beifang; ZHU, Zhengwei; WU, Sitao; LI, Weizhong. CD-HIT: Accelerated for clustering the next-generation sequencing data. **Bioinformatics**, [S. l.], v. 28, n. 23, p. 3150–3152, 2012. DOI: 10.1093/bioinformatics/bts565.

GADHAVE, Kundlik; KUMAR, Prateek; KUMAR, Ankur; BHARDWAJ, Taniya; GARG, Neha; GIRI, Rajanish. NSP 11 of SARS-CoV-2 is an Intrinsically Disordered Protein. **bioRxiv**, [S. l.], n. October, p. 2020.10.07.330068, 2020. DOI: 10.1101/2020.10.07.330068. Disponível em: <https://doi.org/10.1101/2020.10.07.330068>.

GAO, Wenjing; LV, Jun; PANG, Yuanjie; LI, Li-Ming. Role of asymptomatic and pre-symptomatic infections in covid-19 pandemic. **Bmj**, [S. l.], p. n2342, 2021. DOI: 10.1136/bmj.n2342.

GAO, Yan et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. **Science**, [S. l.], v. 368, n. 6492, p. 779–782, 2020. DOI: 10.1126/science.abb7498.

GARRISON, Erik; MARTH, Gabor. Haplotype-based variant detection from short-read sequencing. [S. l.], p. 1–9, 2012. Disponível em: <http://arxiv.org/abs/1207.3907>.

GREGORY, T. Ryan. Understanding Evolutionary Trees. **Evolution: Education and Outreach**, [S. l.], v. 1, n. 2, p. 121–137, 2008. DOI: 10.1007/s12052-008-0035-x.

GRUBAUGH, Nathan D. et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. **Genome Biology**, [S. l.], v. 20, n. 1, p. 1–19, 2019. DOI: 10.1186/s13059-018-1618-7.

GRUNEWALD, Matthew E.; FEHR, Anthony R.; ATHMER, Jeremiah; PERLMAN, Stanley. The coronavirus nucleocapsid protein is ADP-ribosylated. **Virology**, [S. l.], v. 517, n. November 2017, p. 62–68, 2018. DOI: 10.1016/j.virol.2017.11.020. Disponível em: <https://doi.org/10.1016/j.virol.2017.11.020>.

GUINDON, Stéphane; DUFAYARD, Jean Franois; LEFORT, Vincent; ANISIMOVA, Maria; HORDIJK, Wim; GASCUEL, Olivier. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. **Systematic Biology**, [S. l.], v. 59, n. 3, p. 307–321, 2010. DOI: 10.1093/sysbio/syq010.

HACKBART, Matthew; DENG, Xufang; BAKER, Susan C. Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. **Proceedings of the National Academy of Sciences of the United States of America**, [S. l.], v. 117, n. 14, p. 8094–8103, 2020. DOI: 10.1073/pnas.1921485117.

HARCOURT, Brian H.; JUKNELIENE, Dalia; KANJANAHALUETHAI, Amornrat; BECHILL, John; SEVERSON, Kari M.; SMITH, Catherine M.; ROTA, Paul A.; BAKER, Susan C. Identification of Severe Acute Respiratory Syndrome Coronavirus Replicase Products and Characterization of Papain-Like Protease Activity. **Journal of Virology**, [S. l.], v. 78, n. 24, p. 13600–13612, 2004. DOI: 10.1128/jvi.78.24.13600-13612.2004.

HASEGAWA, Masami; KISHINO, Hirohisa; YANO, Taka aki. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. **Journal of Molecular Evolution**, [S. l.], v. 22, n. 2, p. 160–174, 1985. DOI: 10.1007/BF02101694.

HE, Xi et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. **Nature Medicine**, [S. l.], v. 26, n. 5, p. 672–675, 2020. DOI: 10.1038/s41591-020-0869-5.

HEALEY, Adam; FURTADO, Agnelo; COOPER, Tal; HENRY, Robert J. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. **Plant Methods**, [S. l.], v. 10, n. 1, p. 1–8, 2014. DOI: 10.1186/1746-4811-10-21.

HILLEN, Hauke S.; KOKIC, Goran; FARNUNG, Lucas; DIENEMANN, Christian; TEGUNOV, Dmitry; CRAMER, Patrick. Structure of replicating SARS-CoV-2 polymerase. **Nature**, [S. l.], v. 584, n. 7819, p. 154–156, 2020. DOI: 10.1038/s41586-020-2368-8. Disponível em: <http://dx.doi.org/10.1038/s41586-020-2368-8>.

HOFFMANN, Markus et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. **Cell**, [S. l.], v. 181, n. 2, p. 271–280.e8, 2020. DOI: 10.1016/j.cell.2020.02.052.

HUANG, Gene K. L.; STEWARDSON, Andrew J.; GRAYSON, Michael L. Back to basics: Hand hygiene and isolation. **Current Opinion in Infectious Diseases**, [S. l.], v. 27, n. 4, p. 379–389, 2014. DOI: 10.1097/QCO.0000000000000080.

HUMPHRIES, Elizabeth M.; WINKER, Kevin. Working through polytomies: Auklets

revisited. **Molecular Phylogenetics and Evolution**, [S. l.], v. 54, n. 1, p. 88–96, 2010. DOI: 10.1016/j.ympev.2009.07.023. Disponível em: <http://dx.doi.org/10.1016/j.ympev.2009.07.023>.

HUTCHISON, Clyde A. DNA sequencing: Bench to bedside and beyond. **Nucleic Acids Research**, [S. l.], v. 35, n. 18, p. 6227–6237, 2007. DOI: 10.1093/nar/gkm688.

ISHIKAWA, Sohta A.; ZHUKOVA, Anna; IWASAKI, Wataru; GASCUEL, Olivier; PUPKO, Tal. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. **Molecular Biology and Evolution**, [S. l.], v. 36, n. 9, p. 2069–2085, 2019. DOI: 10.1093/molbev/msz131.

IVANOV, Konstantin A.; THIEL, Volker; DOBBE, Jessika C.; VAN DER MEER, Yvonne; SNIJDER, Eric J.; ZIEBUHR, John. Multiple Enzymatic Activities Associated with Severe Acute Respiratory Syndrome Coronavirus Helicase. **Journal of Virology**, [S. l.], v. 78, n. 11, p. 5619–5632, 2004. DOI: 10.1128/jvi.78.11.5619-5632.2004.

JANG, Kyoung Jin; JEONG, Seonghwan; KANG, Dong Young; SP, Nipin; YANG, Young Mok; KIM, Dong Eun. A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA. **Scientific Reports**, [S. l.], v. 10, n. 1, p. 1–13, 2020. DOI: 10.1038/s41598-020-61432-1. Disponível em: <http://dx.doi.org/10.1038/s41598-020-61432-1>.

KANTOR, Jonathan. Behavioral considerations and impact on personal protective equipment use: Early lessons from the coronavirus (COVID-19) pandemic. **Journal of the American Academy of Dermatology**, [S. l.], v. 82, n. 5, p. 1087–1088, 2020. DOI: 10.1016/j.jaad.2020.03.013. Disponível em: <https://doi.org/10.1016/j.jaad.2020.03.013>.

KARABA, Authors Andrew H. et al. A Third Dose of SARS-CoV-2 Vaccine Increases Neutralizing Antibodies Against Variants of Concern in Solid Organ Transplant Recipients. **medRxiv**, [S. l.], 2021.

KATOH, Kazutaka; STANDLEY, Daron M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. **Molecular Biology and Evolution**, [S. l.], v. 30, n. 4, p. 772–780, 2013. DOI: 10.1093/molbev/mst010.

KEARSE, Matthew et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, [S. l.], v. 28, n. 12, p. 1647–1649, 2012. DOI: 10.1093/bioinformatics/bts199.

KERR, Ligia Regina Franco Sansigolo et al. Covid-19 no Nordeste do Brasil: primeiro ano de pandemia e incertezas que estão por vir. **Revista de Saúde Pública**, [S. l.], v. 55, p. 35, 2021. DOI: 10.11606/s1518-8787.2021055003728.

KIMURA, Motoo. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**, [S. l.], v. 16, n. 2, p. 111–120, 1980. DOI: 10.1007/BF01731581.

KOBOLDT, Daniel C. Best practices for variant calling in clinical sequencing. **Genome Medicine**, [S. l.], v. 12, n. 1, p. 1–13, 2020. DOI: 10.1186/s13073-020-00791-w.

LAMARCA, Alessandra P. et al. Genomic Surveillance Tracks the First Community Outbreak of the SARS-CoV-2 Delta (B.1.617.2) Variant in Brazil. **Journal of Virology**, [S. l.], n. November, 2021. DOI: 10.1128/jvi.01228-21.

LEVIN, Einav G. et al. Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months. **New England Journal of Medicine**, [S. l.], v. 385, n. 24, p. e84, 2021. DOI: 10.1056/nejmoa2114583.

LI, Baisheng et al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. **medRxiv**, [S. l.], p. 2021.07.07.21260122, 2021. Disponível em: <https://www.medrxiv.org/content/10.1101/2021.07.07.21260122v2%0Ahttps://www.medrxiv.org/content/10.1101/2021.07.07.21260122v2.abstract>.

LI, Heng. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, [S. l.], v. 27, n. 21, p. 2987–2993, 2011. DOI: 10.1093/bioinformatics/btr509.

LI, Heng. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [S. l.], v. 00, n. 00, p. 1–3, 2013. Disponível em: <http://arxiv.org/abs/1303.3997>.

LI, Qun et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. **New England Journal of Medicine**, [S. l.], v. 382, n. 13, p. 1199–1207, 2020. DOI: 10.1056/nejmoa2001316.

LIU, Chang; SHI, Wei; BECKER, Scott T.; SCHATZ, David G.; LIU, Bin; YANG, Yang. Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. **Science**, [S. l.], v. 373, n. 6559, p. 1142–1146, 2021. DOI: 10.1126/science.abi9310.

LIU, Ding Xiang; FUNG, To Sing; CHONG, Kelvin Kian Long; SHUKLA, Aditi; HILGENFELD, Rolf. Accessory proteins of SARS-CoV and other coronaviruses. **Antiviral Research**, [S. l.], v. 109, n. 1, p. 97–109, 2014. DOI: 10.1016/j.antiviral.2014.06.013. Disponível em: <http://dx.doi.org/10.1016/j.antiviral.2014.06.013>.

LUBINSKI, Bailey; FERNANDES, Maureen H. V.; FRAZIER, Laura; TANG, Tiffany; DANIEL, Susan; DIEL, Diego; JAIMES, Javier; WHITTAKER, Gary R. Functional

Evaluation of Proteolytic Activation for the SARS-CoV-2 Variant B.1.1.7: Role of the P681H Mutation. **SSRN Electronic Journal**, [S. l.], p. 1–28, 2021. DOI: 10.2139/ssrn.3889709.

MACHADO, Leandro C.; GUTIERREZ, Sara Duarte; BARCELOS, Vitória Amaral Araujo; AGUIAR, Pedro Igor Portalea; MARQUES, Aureliano Moreira; PINHEIRO, Ana Luiza. Discovery: The use of the N95 / PFF2/KF94 filter mask by 100% of the South Korean herd has saved thousands of lives on COVID-19. **SSRN Electronic Journal**, [S. l.], 2020.

MCCRONE, John T.; LAURING, Adam S. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. **Journal of Virology**, [S. l.], v. 90, n. 15, p. 6884–6895, 2016. DOI: 10.1128/jvi.00667-16.

METZKER, Michael L. Emerging technologies in DNA sequencing. **Genome Research**, [S. l.], v. 15, n. 12, p. 1767–1776, 2005. DOI: 10.1101/gr.3770505.

METZKER, Michael L. Sequencing technologies the next generation. **Nature Reviews Genetics**, [S. l.], v. 11, n. 1, p. 31–46, 2010. DOI: 10.1038/nrg2626.

MIORIN, Lisa et al. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. **Proceedings of the National Academy of Sciences of the United States of America**, [S. l.], v. 117, n. 45, p. 28344–28354, 2020. DOI: 10.1073/pnas.2016650117.

MIZRAHI, Barak; LOTAN, Roni; KALKSTEIN, Nir; PERETZ, Asaf; PEREZ, Galit; BEN-TOV, Amir; CHODICK, Gabriel; GAZIT, Sivan; PATALON, Tal. Correlation of SARS-CoV-2-breakthrough infections to time-from-vaccine. **Nature Communications**, [S. l.], v. 12, n. 1, p. 1–5, 2021. DOI: 10.1038/s41467-021-26672-3.

MU, Jingfang et al. SARS-CoV-2-encoded nucleocapsid protein acts as a viral suppressor of RNA interference in cells. **Science China Life Sciences**, [S. l.], v. 63, n. 9, 2020. DOI: 10.1007/s11427-020-1692-1.

NANDURI, Srinivas et al. Effectiveness of Pfizer-BioNTech and Moderna Vaccines in Preventing SARS-CoV-2 Infection Among Nursing Home Residents Before and During Widespread Circulation of the SARS-CoV-2 B.1.617.2 (Delta) Variant — National Healthcare Safety Network, March 1–August. **MMWR. Morbidity and Mortality Weekly Report**, [S. l.], v. 70, n. 34, p. 1163–1166, 2021. DOI: 10.15585/mmwr.mm7034e3.

NASREEN, Sharifa et al. Effectiveness of mRNA and ChAdOx1 COVID-19 vaccines against symptomatic SARS-CoV-2 infection and severe outcomes with variants of concern in Ontario. **medRxiv**, [S. l.], p. 2021.06.28.21259420, 2021. Disponível em: <https://www.medrxiv.org/content/10.1101/2021.06.28.21259420v3%0Ahttps://www.medrxiv>.

org/content/10.1101/2021.06.28.21259420v3.abstract.

NAVECA, Felipe Gomes et al. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. **Nature Medicine**, [S. l.], v. 27, n. 7, p. 1230–1238, 2021. a. DOI: 10.1038/s41591-021-01378-7. Disponível em: <http://dx.doi.org/10.1038/s41591-021-01378-7>.

NAVECA, Felipe Gomes 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**, [S. l.], p. 2021.09.12.21263453, 2021. b. Disponível em: <https://www.medrxiv.org/content/10.1101/2021.09.12.21263453v1%0Ahttps://www.medrxiv.org/content/10.1101/2021.09.12.21263453v1.abstract>.

NAVECA, Felipe; SOUZA, Victor; COSTA, Cristiano. 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. **Research Square Square**, [S. l.], n. February, p. 1–21, 2021. Disponível em: <https://www.researchsquare.com/article/rs-275494/v1>.

NGUYEN, Lam Tung; SCHMIDT, Heiko A.; VON HAESELER, Arndt; MINH, Bui Quang. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Molecular Biology and Evolution**, [S. l.], v. 32, n. 1, p. 268–274, 2015. DOI: 10.1093/molbev/msu300.

O'TOOLE, Áine et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. **Virus Evolution**, [S. l.], p. 1–9, 2021. a. DOI: 10.1093/ve/veab064.

O'TOOLE, Áine et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. **Virus Evolution**, [S. l.], v. 7, n. 2, p. 1–9, 2021. b. DOI: 10.1093/ve/veab064.

OLIVERA, Adriana; COAGLIO, Thabata; ALDO, Robert. O Que a Pandemia Da Covid-19 Tem Nos Ensinado Sobre Adoção De Medidas De Prevenção? **Texto & Contexto Enfermagem**, [S. l.], v. 29, p. 13, 2020. Disponível em: https://www.scielo.br/pdf/tce/v29/pt_1980-265X-tce-29-e20200106.pdf.

PANCER, Katarzyna; MILEWSKA, Aleksandra; OWCZAREK, Katarzyna; DABROWSKA, Agnieszka; KOWALSKI, Michał; LABAJ, Paweł; BRANICKI, Wojciech; SANAK, Marek; PYRC, Krzysztof. The SARS-CoV-2 ORF10 is not essential in vitro or in

vivo in humans. **PLoS Pathogens**, [S. l.], v. 16, n. 12, p. 1–8, 2020. DOI: 10.1371/journal.ppat.1008959. Disponível em: <http://dx.doi.org/10.1371/journal.ppat.1008959>.

PARRY, H.; BRUTON, R.; STEPHENS, C.; BROWN, K.; AMIRTHALINGAM, G.; HALLIS, B.; OTTER, A.; ZUO, J.; MOSS, P. Extended interval BNT162b2 vaccination enhances peak antibody generation in older people. **medRxiv**, [S. l.], n. December 2020, 2021.

PEDRO, Nicole et al. Dynamics of a dual sars-cov-2 lineage co-infection on a prolonged viral shedding COVID-19 case: Insights into clinical severity and disease duration. **Microorganisms**, [S. l.], v. 9, n. 2, p. 1–10, 2021. DOI: 10.3390/microorganisms9020300.

PERICO, Camila P.; PIERRI, Camilla R. De; NETO, Giuseppe P.; FERNANDES, Danrley R.; PEDROSA, Fabio O.; SOUZA, Emanuel M. De; RAITTZ, Roberto T. Genomic landscape of SARS-CoV-2 pandemic in Brazil suggests an external P . 1 variant origin. [S. l.], p. 1–21, 2021.

PILLON, Monica C. et al. Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. **Nature Communications**, [S. l.], v. 12, n. 1, p. 1–12, 2021. DOI: 10.1038/s41467-020-20608-z. Disponível em: <http://dx.doi.org/10.1038/s41467-020-20608-z>.

PIROOZNIA, Mehdi; KRAMER, Melissa; PARLA, Jennifer; GOES, Fernando S.; POTASH, James B.; MCCOMBIE, W. Richard; ZANDI, Peter P. Validation and assessment of variant calling pipelines for next-generation sequencing. **Human genomics**, [S. l.], v. 8, p. 14, 2014. DOI: 10.1186/1479-7364-8-14.

POLACK, Fernando P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. **New England Journal of Medicine**, [S. l.], v. 383, n. 27, p. 2603–2615, 2020. DOI: 10.1056/nejmoa2034577.

RAMBAUT, Andrew; DRUMMOND, Alexei J.; XIE, Dong; BAELE, Guy; SUCHARD, Marc A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. **Systematic Biology**, [S. l.], v. 67, n. 5, p. 901–904, 2018. DOI: 10.1093/sysbio/syy032.

RAMBAUT, Andrew; HOLMES, Edward C.; TOOLE, Áine O.; HILL, Verity; MCCRONE, John T.; RUIS, Christopher; PLESSIS, Louis; PYBUS, Oliver G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. **Nat Microbiol**, [S. l.], v. 5, n. 11, p. 1403–1407, 2020. DOI: 10.1038/s41564-020-0770-5.A.

RAMBAUT, Andrew; LAM, Tommy T.; CARVALHO, Luiz Max; PYBUS, Oliver G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-

O-Gen). **Virus Evolution**, [S. l.], v. 2, n. 1, p. 1–7, 2016. DOI: 10.1093/ve/vew007.

REDONDO, Natalia; ZALDÍVAR-LÓPEZ, Sara; GARRIDO, Juan J.; MONTOYA, Maria. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. **Frontiers in Immunology**, [S. l.], v. 12, n. December 2019, p. 1–8, 2021. DOI: 10.3389/fimmu.2021.708264.

REN, Yujie et al. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. **Cellular and Molecular Immunology**, [S. l.], v. 17, n. 8, p. 881–883, 2020. DOI: 10.1038/s41423-020-0485-9.

RODRIGUEZ, F.; OLIVER, J. L.; MARIN, A. The General Stochastic Model of Nucleotide Substitution The G4H Model. **J. theor. Biol.**, [S. l.], v. 142, p. 485–501, 1990.

ROSAS-LEMUS, Monica et al. The crystal structure of nsp10-nsp16 heterodimer from SARS CoV-2 in complex with S-adenosylmethionine. **bioRxiv**, [S. l.], p. 1–22, 2020. DOI: 10.1101/2020.04.17.047498.

SAGULENKO, Pavel; PULLER, Vadim; NEHER, Richard A. TreeTime: Maximum-likelihood phylodynamic analysis. **Virus Evolution**, [S. l.], v. 4, n. 1, p. 1–9, 2018. DOI: 10.1093/ve/vex042.

SAKAI, Yusuke; KAWACHI, Kengo; TERADA, Yutaka; OMORI, Hiroko; MATSUURA, Yoshiharu; KAMITANI, Wataru. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. **Virology**, [S. l.], v. 510, n. June, p. 165–174, 2017. DOI: 10.1016/j.virol.2017.07.019. Disponível em: <http://dx.doi.org/10.1016/j.virol.2017.07.019>.

SANTARPIA, Joshua L. et al. Aerosol and surface contamination of SARS-CoV-2 observed in quarantine and isolation care. **Scientific Reports**, [S. l.], v. 10, n. 1, p. 1–8, 2020. DOI: 10.1038/s41598-020-69286-3. Disponível em: <https://doi.org/10.1038/s41598-020-69286-3>.

SANTOS, Ana Clara; CARVALHO, André; AMORIM, Andressa; RIBEIRO, Cleide Felicia; SANTANA, Heiko. Nota Técnica GVIMS/GGTES/ANVISA Nº 04/2020 Orientações para serviços de saúde: medidas de prevenção e controle que devem ser adotadas durante a assistência aos casos suspeitos ou confirmados de infecção pelo novo coronavírus (SARS-CoV-2) - Atualizada em 2. NOTA TÉCNICA GVIMS/GGTES/ANVISA Nº 04/2020 ORIENTAÇÕES PARA SERVIÇOS DE SAÚDE: MEDIDAS DE PREVENÇÃO E CONTROLE QUE DEVEM SER ADOTADAS DURANTE A ASSISTÊNCIA AOS CASOS SUSPEITOS OU CONFIRMADOS DE INFECÇÃO PELO NOVO CORONAVÍRUS (SARS-

CoV-2). (atualizada em 31/03/2020) Gerência de Vigilância e Monitoramento em Serviços de Saúde Gerência Geral de Tecnologia em Serviços de Saúde Agência Nacional de Vigilância Sanitária. **Agência Nacional de Vigilância Sanitária**, 2021. p. 118.

SANTOS PAIVA, Marcelo Henrique et al. Multiple introductions followed by ongoing community spread of SARS-CoV-2 at one of the largest metropolitan areas in the Northeast of Brazil. **medRxiv**, [S. l.], 2020. DOI: 10.1101/2020.08.25.20171595.

SCHIRMER, Melanie; IJAZ, Umer Z.; D'AMORE, Rosalinda; HALL, Neil; SLOAN, William T.; QUINCE, Christopher. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. **Nucleic Acids Research**, [S. l.], v. 43, n. 6, 2015. DOI: 10.1093/nar/gku1341.

SCHUBERT, Katharina et al. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. **Nature Structural and Molecular Biology**, [S. l.], v. 27, n. 10, p. 959–966, 2020. DOI: 10.1038/s41594-020-0511-8.

SELF, Wesley H. et al. Comparative Effectiveness of Moderna, Pfizer-BioNTech, and Janssen (Johnson & Johnson) Vaccines in Preventing COVID-19 Hospitalizations among Adults without Immunocompromising Conditions - United States, March-August 2021. **Morbidity and Mortality Weekly Report**, [S. l.], v. 70, n. 38, p. 1337–1343, 2021. DOI: 10.15585/MMWR.MM7038E1.

SHIN, Donghyuk et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. **Nature**, [S. l.], v. 587, n. 7835, p. 657–662, 2020. DOI: 10.1038/s41586-020-2601-5. Disponível em: <http://dx.doi.org/10.1038/s41586-020-2601-5>.

SHROTRI, Madhumita et al. Spike-antibody waning after second dose of BNT162b2 or ChAdOx1. **The Lancet**, [S. l.], v. 398, n. 10298, p. 385–387, 2021. DOI: 10.1016/S0140-6736(21)01642-1. Disponível em: [http://dx.doi.org/10.1016/S0140-6736\(21\)01642-1](http://dx.doi.org/10.1016/S0140-6736(21)01642-1).

SIMS, David; SUDBERY, Ian; ILOTT, Nicholas E.; HEGER, Andreas; PONTING, Chris P. Sequencing depth and coverage: Key considerations in genomic analyses. **Nature Reviews Genetics**, [S. l.], v. 15, n. 2, p. 121–132, 2014. DOI: 10.1038/nrg3642.

SIU, Kam Leung et al. Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. **FASEB Journal**, [S. l.], v. 33, n. 8, p. 8865–8877, 2019. DOI: 10.1096/fj.201802418R.

SNIJDER, Eric J.; VAN DER MEER, Yvonne; ZEVENHOVEN-DOBBE, Jessika; ONDERWATER, Jos J. M.; VAN DER MEULEN, Jannes; KOERTEN, Henk K.; MOMMAAS, A. Mieke. Ultrastructure and Origin of Membrane Vesicles Associated with the

by the Nsp1 protein of SARS-CoV-2. **Science**, [S. l.], v. 369, n. 6508, p. 1249–1256, 2020. DOI: 10.1126/SCIENCE.ABC8665.

TSOI, Ho; LI, Li; CHEN, Zhefan S.; LAU, Kwok Fai; TSUI, Stephen K. W.; CHAN, Ho Yin Edwin. The SARS-coronavirus membrane protein induces apoptosis via interfering with PDK1PKB/Akt signalling. **Biochemical Journal**, [S. l.], v. 464, n. 3, p. 439–447, 2014. DOI: 10.1042/BJ20131461.

VOYSEY, Merryn et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. **The Lancet**, [S. l.], v. 397, n. 10269, p. 99–111, 2021. DOI: 10.1016/S0140-6736(20)32661-1.

WALLS, Alexandra C.; PARK, Young Jun; TORTORICI, M. Alejandra; WALL, Abigail; MCGUIRE, Andrew T.; VEESLER, David. Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. **bioRxiv**, [S. l.], n. January, 2020. DOI: 10.1101/2020.02.19.956581.

WALSH, Edward E. et al. Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. **New England Journal of Medicine**, [S. l.], v. 383, n. 25, p. 2439–2450, 2020. DOI: 10.1056/nejmoa2027906.

WANG, Chen; HORBY, Peter W.; HAYDEN, Frederick G.; GAO, George F. A novel coronavirus outbreak of global health concern. **The Lancet**, [S. l.], v. 395, n. 10223, p. 470–473, 2020. DOI: 10.1016/S0140-6736(20)30185-9.

WANG, Pengfei et al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. **Cell Host and Microbe**, [S. l.], v. 29, n. 5, p. 747–751.e4, 2021. DOI: 10.1016/j.chom.2021.04.007. Disponível em: <https://doi.org/10.1016/j.chom.2021.04.007>.

WHEATLEY, Adam K. et al. Evolution of immune responses to SARS-CoV-2 in mild-moderate COVID-19. **Nature Communications**, [S. l.], v. 12, n. 1, p. 1–11, 2021. DOI: 10.1038/s41467-021-21444-5. Disponível em: <http://dx.doi.org/10.1038/s41467-021-21444-5>.

WONG, Hui Hui; FUNG, To Sing; FANG, Shouguo; HUANG, Mei; LE, My Tra; LIU, Ding Xiang. Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3. **Virology**, [S. l.], v. 515, n. November 2017, p. 165–175, 2018. DOI: 10.1016/j.virol.2017.12.028. Disponível em: <https://doi.org/10.1016/j.virol.2017.12.028>.

WOROBAY, By Michael. Dissecting the early COVID-19 cases in Wuhan. **Science**, [S. l.], v. 374, n. 6572, p. 1202–1204, 2021. DOI: 10.1126/science.abm4454.

WRAPP, Daniel; WANG, Nianshuang; CORBETT, Kizzmekia S.; GOLDSMITH, Jory A.; HSIEH, Ching Lin; ABIONA, Olubukola; GRAHAM, Barney S.; MCLELLAN, Jason S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. **Science**, [S. l.], v. 367, n. 6483, p. 1260–1263, 2020. DOI: 10.1126/science.aax0902.

XIA, Hongjie; CAO, Zengguo; XIE, Xuping; ZHANG, Xianwen; CHEN, John Yun Chung; WANG, Hualei; MENACHERY, Vineet D.; RAJSBAUM, Ricardo; SHI, Pei Yong. Evasion of Type I Interferon by SARS-CoV-2. **Cell Reports**, [S. l.], v. 33, n. 1, p. 108234, 2020. DOI: 10.1016/j.celrep.2020.108234. Disponível em: <https://doi.org/10.1016/j.celrep.2020.108234>.

YANG, Z. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. **Molecular Biology and Evolution**, [S. l.], v. 10, n. 6, p. 1396–1401, 1993. DOI: 10.1093/oxfordjournals.molbev.a040082.

YOSHIMOTO, Francis K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. **Protein Journal**, [S. l.], v. 39, n. 3, p. 198–216, 2020. DOI: 10.1007/s10930-020-09901-4. Disponível em: <https://doi.org/10.1007/s10930-020-09901-4>.

ZHANG, Linlin; LIN, Daizong; SUN, Xinyuanyuan; CURTH, Ute; DROSTEN, Christian; SAUERHERING, Lucie; BECKER, Stephan; ROX, Katharina; HILGENFELD, Rolf. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. **Science**, [S. l.], v. 368, n. 6489, p. 409–412, 2020. DOI: 10.1126/science.abb3405.

ZHANG, Yan et al. In vivo structure and dynamics of the SARS-CoV-2 RNA genome. **Nature Communications**, [S. l.], v. 12, n. 1, p. 1–12, 2021. DOI: 10.1038/s41467-021-25999-1. Disponível em: <http://dx.doi.org/10.1038/s41467-021-25999-1>.

ZHOU, Ziliang et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. **iScience**, [S. l.], v. 24, n. 3, p. 102187, 2021. DOI: 10.1016/j.isci.2021.102187. Disponível em: <https://doi.org/10.1016/j.isci.2021.102187>.