



**UNIVERSIDADE FEDERAL DO CEARÁ  
CENTRO DE CIÊNCIAS  
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR  
PROGRAMA DE GRADUAÇÃO EM BIOTECNOLOGIA**

**CLEBER FURTADO AKSENEN**

**ASPECTOS TÉCNICOS E OPERACIONAIS DE UM MODELO DE SERVIÇO DE  
VIGILÂNCIA GENÔMICA PARA COVID-19 NO ESTADO DO CEARÁ**

**FORTALEZA**

**2022**

CLEBER FURTADO AKSENEN

ASPECTOS TÉCNICOS E OPERACIONAIS DE UM MODELO DE SERVIÇO DE  
VIGILÂNCIA GENÔMICA PARA COVID-19 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  
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

A1a Aksenen, Cleber Furtado.

Aspectos técnicos e operacionais de um modelo de serviço de vigilância genômica para COVID-19 no estado do Ceará / Cleber Furtado Aksenen. – 2022.

79 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. Variante de preocupação (VOC). 3. Vigilância genômica. 4. Serviço de referência.. 5. Saúde pública. I. Título.

CDD 661

---

**ASPECTOS TÉCNICOS E OPERACIONAIS DE UM MODELO DE SERVIÇO DE  
VIGILÂNCIA GENÔMICA PARA COVID-19 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

Aprovada em: 04/02/2022.

**BANCA EXAMINADORA**

---

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

---

Prof. Dra. Márjory Lima Holanda Araújo  
Universidade Federal do Ceará (UFC)

---

Prof. Dra. Denise Cavalcante Hissa  
Universidade Federal do Ceará (UFC)

Dedico este trabalho à minha família, que me ensinou, acima de tudo, o tipo de homem que quero ser.

## AGRADECIMENTOS

A minha mãe, Kátia Furtado Aksenen, pelo apoio completo em todas as fases de minha vida. Por todo o suporte e carinho com meus estudos e minha vocação. Por tentar me ajudar mesmo sem entender o problema. E acima de tudo, por acreditar em mim.

Ao meu pai, Dirceu José Aksenen, por me ensinar a aprender com meus erros e tentar não os cometer novamente, seguindo sempre pelo caminho correto.

A minha irmã, Thalia Furtado Aksenen, por me ajudar a ser um educador e me fazer valorizar o estudo como ferramenta de mudar vidas. Por sempre tentar me corrigir quando estou errado e sempre agir com a mais profunda virtude.

Ao Professor Dr. Fábio Miyajima, que me acolheu e confiou em minhas capacidades. Pelo acompanhamento, orientação e determinação em transformar falhas em conquistas, me ensinando a buscar conhecimento e melhores maneiras de resolver os problemas.

A toda a equipe do Laboratório de Biologia Molecular, Inovação e Diagnóstico (Biomind): Dr. Fábio Miyajima, Dr. Fernando Braga, Dra. Veridiana Miyajima, Thais Costa, Thais Ferreira, Joaquim Cesar, Carlos Leonardo, Ticiane Cavalcante, Jamille Maria, Pedro Miguel, Suzana Almeida, Igor Oliveira, Arlene Almeida, Guilherme Lobo e Bárbara Guimarães, pelo apoio nesse período tão conturbado que estamos vivendo. Tenho a certeza de que estamos fazendo um excelente trabalho e seguiremos na luta contra a Covid. Um agradecimento especial ao Dr. Francisco Éder, que apesar de não fazer mais parte da equipe, me ajudou bastante no desenvolvimento desse trabalho e no meu próprio desenvolvimento pessoal. Um grande abraço, é um prazer enorme fazer parte desse time.

A todos os demais companheiros e colaboradores da Rede Genômica da Fiocruz e do Hemocentro do Ceará, por todo o esforço, respeito e confiança. Sem a ajuda de todos, nada disso seria possível.

A Unadig, ao Lacen e a todos os centros de testagem do Ceará, pelo fornecimento de dados e amostras que compõem parte desse estudo e por todo o trabalho prestado durante a pandemia para a população cearense.

Ao Filipe Zimmer, do Instituto Aggeu Magalhães da Fiocruz PE, por ensinar não só a mim, como também aos meus outros colegas da equipe de Bioinformática. Por acreditar no nosso potencial em trabalharmos juntos no desenvolvimento de uma grande parceria.

Aos meus amigos e companheiros de curso, Pedro, Lucas, João Matheus, Daniel, Melissa, Natália, Lara, Isabele, Liandra e Nicole, que permaneceram comigo nesses longos 5

anos, me ajudando, ensinando e mostrando as maravilhas dessa área do conhecimento que escolhi. Tenho a certeza de que mais anos de amizade virão.

Aos meus amigos de escola, Guilherme, Ricardo, Vinicius, Kevyn, Luan, Matheus, Aleksandr, Lucas Saraiva, Lucas Gabriel, Maria, André, Silas, Fábio, Cleiton, Isac, Arthur, Breno, Eliezer e Caio, que em muito me ajudaram ao longo desses mais de 10 anos.

A todos os professores do Bacharelado em Biotecnologia e antigos orientadores e coorientadores por todos os ensinamentos adquiridos. Assim como todas as unidades de apoio à pesquisa que me financiaram durante todos esses anos.

Muito obrigado!

“Em algum lugar, há algo incrível esperando para ser descoberto.”

(Carl Sagan)

## RESUMO

As variantes de preocupação (VOCs) de SARS-CoV-2 são o resultado do processo evolutivo do vírus desde sua emergência e constituem uma contínua ameaça para a saúde pública mundial. A identificação precisa dessas variantes se tornou essencial para o direcionamento de medidas de contenção da pandemia da COVID-19, inclusive no estado do Ceará. Essa identificação é feita principalmente através de sequenciamento do genoma completo do vírus, um método de alto custo e laborioso que exige diversas etapas de preparo e processamento de amostras. Métodos mais rápidos de melhor custo-efetividade têm sido desenvolvidos para a determinação objetiva de VOCs predominantes no estado e de maior relevância epidemiológica, devido sua maior capacidade de transmissão, disseminação e escape vacinal. Cronologicamente, a primeira VOC identificada no estado do Ceará foi a variante Gama, que originalmente emergiu no final de 2020 no estado do Amazonas e rapidamente se disseminou para os demais estados brasileiros e exterior, tornando-se a principal responsável pela segunda onda pandêmica no Ceará. A segunda VOC de maior expressão que foi introduzida no Estado foi a variante Delta, que apesar de ter surgido no final de 2020 na Índia, os primeiros registros foram registrados no Brasil a partir de abril de 2021, tornando-se a variante predominante durante o segundo semestre. Nesse contexto, a instalação de um Serviço de Vigilância Genômica no Ceará atuando em conjunto com um laboratório referenciado para diagnóstico da COVID-19 da rede de saúde estadual (Centro de Hematologia e Hemoterapia do Ceará/Hemoce) constituiu-se na verdade em uma necessidade de saúde pública, tanto para triagem prioritária de casos sob investigação, como para o monitoramento da população viral e seu processo evolutivo, e finalmente para a notificação em tempo hábil aos órgãos de vigilância em saúde. O presente trabalho propõe um modelo de serviço para a Vigilância Genômica no estado do Ceará, concebido por uma rede multicêntrica de laboratórios participantes da Fiocruz, através do desenho de fluxos de trabalho, padronização e otimização do modelo proposto, por meio da avaliação da abrangência do serviço em contemplar a maioria dos municípios do estado e análise comparativa dos dados com outros serviços de vigilância no Brasil. Em um segundo momento, foram reunidos os principais problemas e intercorrências associados ao modelo proposto e apresentação de potenciais soluções analíticas, através da testagem e re-testagem com contraprova, concordância de resultados com ensaios moleculares de rt-qPCR, por método indireto de inferência e a otimização de procedimentos de sequenciamento genômico e de controle de qualidade de dados. Por fim, é relatado o impacto e a relevância da constituição de um serviço de referência em vigilância genômica para o estado.

**Palavras-chave:** SARS-CoV-2. Variante de Preocupação (VOC). Vigilância Genômica. Serviço de Referência. Saúde Pública.

## ABSTRACT

The variants of concern (VOCs) of SARS-CoV-2 are the result of the evolutionary process of the virus since its emergence and constitute a threat to public health worldwide. The precise identification of these variants has become essential for targeting measures to contain the COVID-19 pandemic, including in the state of Ceará. This identification is mainly done through sequencing the complete genome of the virus, a costly and laborious method that requires several steps of sample preparation and processing. Faster, more cost-effective methods have been developed for the objective determination of VOCs predominant in the state and of greater epidemiological relevance due to their greater capacity for transmission, dissemination, and vaccine escape. Chronologically, the first VOC identified in the state of Ceará was the Gamma variant, which originally emerged in late 2020 in the state of Amazonas and quickly spread to other Brazilian states and abroad, becoming the main responsible for the second pandemic wave in Ceará. The second most expressive VOC that was introduced in the State was the Delta variant, which despite having appeared in late 2020 in India, the first records were registered in Brazil from April 2021, becoming the predominant variant, including in the State, during the second half of 2021. In this context, the installation of a Genomic Surveillance Service in Ceará working together with a referenced laboratory for the diagnosis of COVID-19 of the state health network (Centro de Hematologia e Hemoterapia do Ceará/Hemoce) it constituted a public health need, both for priority screening of cases under investigation, for monitoring the viral population and its evolutionary process, and finally for timely reporting to health surveillance agencies. The present work proposes a service model for Genomic Surveillance in the state of Ceará, conceived by a multicentric network of laboratories participating in Fiocruz, through the design of workflows, standardization, and optimization of the proposed model, through the evaluation of the scope of the service to cover most municipalities in the state and comparative analysis of data with other surveillance services in Brazil. In a second moment, the main problems and complications associated with the proposed model were gathered and potential analytical solutions were presented, through testing and re-testing with control, agreement of results with molecular assays of rt-PCR by indirect method of inference and the optimization of genomic sequencing procedures and data quality control. Finally, the impact and relevance of constituting a reference service in genomic surveillance for the state is reported.

**Keywords:** SARS-CoV-2. Variant of Concern (VOC). Genomic Surveillance. Reference Service. Public health.

## LISTA DE FIGURAS

Figura 1 - Representação da estrutura do vírus SARS-CoV-2 e da enzima conversora de angiotensina 2 (ACE2).....	19
Figura 2 - Arquitetura do genoma de SARS-CoV-2. O painel superior indica os nomes dos genes, com o círculo vermelho na extremidade esquerda representando a sequência reguladora de transcrição (TRS-L). O painel inferior representa as proteínas não estruturais (NSPs), derivados do processamento das pp1a e pp1b.....	20
Figura 3 - Árvore filogenética de máxima verossimilhança de sequências amostrais globais do SARS-CoV-2 obtidas no banco de dados Gisaid.....	22
Figura 4 - Arquitetura do genoma da VOC Gama (P.1) destacando suas mutações definidoras de linhagem.....	24
Figura 5 - Gráfico da taxa móvel de casos confirmados de indivíduos com COVID-19 durante o período entre 01 de janeiro de 2021 e 22 de dezembro de 2021, com o pico no número de casos entre março e maio de 2021.....	25
Figura 6 - Arquitetura do genoma da VOC Delta (B.1.617.2) destacando suas mutações definidoras de linhagem.....	26
Figura 7 - Arquitetura do genoma da VOC Ômicron (B.1.1.529) destacando suas mutações definidoras de linhagem.....	27
Figura 8 - Gráfico de quantificação de produtos de PCR em termociclador em tempo real. O gráfico mostra as quantidades absolutas de material genético em função do número de ciclos. Em verde: curvas de amplificação; Em amarelo: valor de threshold.....	28
Figura 9 – Desenho esquemático de genotipagem por meio de PCR em tempo real na qual mostra a intensidade da fluorescência para determinação de alelos. Em vermelho: homozigoto para o alelo 1; Em azul: homozigoto para o alelo 2; Em verde: heterozigoto para ambos os alelos; Em preto e roxo: resultado negativo/ controle negativo.....	29
Figura 10 - Esquema de fluxo de trabalho do ViralFlow. a. Divisão do fluxo em cinco etapas; b. Configuração do fluxo de trabalho para funcionar em diferentes ambientes computacionais; c. Várias saídas geradas pelo fluxo de trabalho por amostra.....	31
Figura 11 - Principais variantes de preocupação sequenciadas pela Rede Genômica da Fiocruz na região do nordeste no período de novembro de 2020 a dezembro de 2021.....	33
Figura 12 - Frequência de variantes de preocupação sequenciadas pela Rede Genômica da Fiocruz no Brasil durante o período de fevereiro de 2020 a dezembro de 2021.....	34
Figura 13 - Fluxo dos dados elaborado para o Serviço de Vigilância Genômica da Fiocruz Ceará.	

.....	42
Figura 14 - Mapa coroplético dos municípios contemplados por Mesorregiões do Ceará. Em branco: municípios que não tiveram amostras sequenciadas. ....	46
Figura 15 -Gráfico de fluxo do quantitativo de variantes de preocupação no Ceará em 2021. ....	48
Figura 16 - Curvas de amplificação de amostras analisadas por inferência baseados na amplificação de uma região de inserção de quatro nucleotídeos característico da VOC Gama (P.1).....	51
Figura 17 - Curvas de amplificação de amostras analisadas por inferência baseados na amplificação de uma região de deleção de nove nucleotídeos característico da VOC Gama (P.1). ....	52
Figura 18 - Ensaio de genotipagem medindo a intensidade da fluorescência da mutação K417T, baseado na detecção de alelos característicos da VOC Gama (P.1). Em vermelho: Homozigotos K/K (não mutantes); Em Azul: Homozigotos T/T (mutantes); Em preto: indeterminados. ....	52
Figura 19 - Diagrama circular do percentual do resultado da validação do ensaio de inferência comparado com os dados de sequenciamento. ....	54
Figura 20 - Mapas de profundidade de amostras classificadas como variante Gama (P.1) sequenciadas utilizando os iniciadores Midnight™. ....	55
Figura 21 - Mapas de profundidade de amostras classificadas como variante Gama (P.1) sequenciadas utilizando os iniciadores ArticV3™. ....	56
Figura 22 - Mapas de profundidade de amostras classificadas como variante Delta (B.1.617.2) sequenciadas utilizando os iniciadores Midnight™. ....	57
Figura 23 - Mapas de profundidade de amostras classificadas como variante Delta (B.1.617.2) sequenciadas utilizando os iniciadores ArticV3™. ....	58
Figura 24 - Mapa das mutações na região da proteína Spike (gene S), utilizando os iniciadores Midnight™ e ArticV3™, alinhadas com uma amostra referência de número de acesso NC_045512.2; a) Amostras classificadas como Gama (P.1); b) Amostras classificadas como Delta (B.1.617.2); em branco) ausência da mutação; em verde) presença da mutação. ....	59

## **LISTA DE TABELAS**

Tabela 1 - Divisão do número de amostras sequenciadas por municípios contemplados das Mesorregiões Cearenses.....	45
Tabela 2 - Quantitativo das principais variantes sequenciadas no Ceará ao longo dos meses de 2021.....	47
Tabela 3 - Lista dos principais problemas técnicos e logísticos observados e possíveis soluções encontradas .....	50
Tabela 4 - Comparativo do resultado dos ensaios de inferência por lote de sequenciamento..	53

## LISTA DE ABREVIATURAS E SIGLAS

ACE2	enzima conversora de angiotensina 2
CEP	código de endereçamento postal
cDNA	DNA complementar
CoV	coronavírus
COVID-19	<i>coronavirus infection disease 2019</i>
CT	limite de ciclo
CTV	centro de testagem ao viajante
EBOV	vírus Ebola
Fiocruz	Fundação Oswaldo Cruz
GAL	gerenciador de ambiente laboratorial
hCoV	coronavírus humano
Hemoce	hemocentro do Ceará
HIV	vírus da imunodeficiência humana
IBGE	Instituto Brasileiro de Geografia e Estatística
Lacen	Laboratório Central de Saúde Pública
MAFFT	<i>multiple sequence alignment program</i>
MERS-CoV	<i>Middle East respiratory syndrome Coronavirus</i>
NCBI	<i>National Center for Biotechnology Information</i>
NGS	sequenciamento de nova geração
NSPs	proteínas não estruturais
OMS	Organização Mundial da Saúde
ORF	<i>open reading frame</i>
PCR	reação em cadeia da DNA polimerase

PP1a	poliproteína 1a
PP1b	poliproteína 1b
RT-qPCR	reação em cadeia da polimerase em tempo real
SARS-CoV-2	<i>severe acute respiratory syndrome coronavírus 2</i>
SESA	Secretaria de Saúde
SUS	Sistema Único de Saúde
Tm	temperatura de <i>melting</i>
Unadig	Unidade de Apoio ao Diagnóstico da COVID-19
VOC	variante de preocupação
VOI	variante de interesse

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>14</b>
<b>2 OBJETIVOS .....</b>	<b>17</b>
<b>2.1 Objetivo geral.....</b>	<b>17</b>
<b>2.2 Objetivos específicos .....</b>	<b>17</b>
<b>3 REVISÃO BIBLIOGRÁFICA .....</b>	<b>18</b>
<b>3.1 Agente etiológico e fisiopatologia do SARS-CoV-2.....</b>	<b>18</b>
<b>3.2 Nomenclatura dinâmica para linhagens de SARS-CoV-2 .....</b>	<b>20</b>
<b>3.3 Variantes de preocupação (VOCs) .....</b>	<b>23</b>
<b>3.3.1 Variante Gama (P.1) .....</b>	<b>23</b>
<b>3.3.2 Variante Delta (B.1.617.2) .....</b>	<b>25</b>
<b>3.3.3 Variante Ômicron (B.1.1.529) .....</b>	<b>26</b>
<b>3.4 Detecção de novas variantes.....</b>	<b>27</b>
<b>3.5 Ferramentas de bioinformática para predição de linhagens .....</b>	<b>30</b>
<b>3.5.1 Pangolin: atribuição filogenética de linhagens .....</b>	<b>32</b>
<b>3.5.2 Nextclade: atribuição de clado, chamada de mutação e verificação de qualidade de sequências.....</b>	<b>32</b>
<b>3.6 Prevalência das principais variantes e linhagens de SARS-CoV-2 no Brasil em 2021.....</b>	<b>33</b>
<b>4 METODOLOGIA.....</b>	<b>35</b>
<b>4.1 Elaboração e otimização do fluxo operacional.....</b>	<b>35</b>
<b>4.2 Avaliação do quantitativo e da abrangência de amostras sequenciadas pelo Serviço de Vigilância Genômica da Fiocruz Ceará .....</b>	<b>36</b>
<b>4.2.1 Obtenção dos dados.....</b>	<b>36</b>
<b>4.2.2 Critérios de exclusão .....</b>	<b>37</b>
<b>4.2.3 Distribuição geográfica e randomização das amostras.....</b>	<b>37</b>
<b>4.3 Avaliação da prevalência de variantes de SARS-CoV-2 no estado do Ceará no ano de 2021, em comparação com outros estados brasileiros .....</b>	<b>37</b>
<b>4.3.1 Predição de linhagens.....</b>	<b>37</b>
<b>4.3.2 Critérios de exclusão .....</b>	<b>38</b>
<b>4.3.3 Organização amostral.....</b>	<b>38</b>
<b>4.3.4 Comparação dos resultados.....</b>	<b>38</b>
<b>4.4 Listagem dos principais desafios e limitações para o serviço, assim como possíveis soluções para os problemas enfrentados.....</b>	<b>38</b>

<b>4.5 Validação dos protocolos de inferência como testes preliminares para a predição de variantes de SARS-CoV-2 .....</b>	<b>39</b>
<b>4.5.1 Categorização do percentual de ensaios de inferência confirmados através de sequenciamento.....</b>	<b>40</b>
<b>4.6 Comparação e otimização de diferentes protocolos de preparo de amostras para sequenciamento .....</b>	<b>40</b>
<b>5 RESULTADOS E DISCUSSÃO .....</b>	<b>41</b>
<b>5.1 Elaboração e otimização do fluxo operacional .....</b>	<b>41</b>
<b>5.1.1 Fluxo da triagem de amostras.....</b>	<b>43</b>
<b>5.1.2 Fluxo do sequenciamento.....</b>	<b>43</b>
<b>5.1.3 Fluxo da bioinformática.....</b>	<b>44</b>
<b>5.2 Avaliação do quantitativo e da abrangência de amostras sequenciadas pelo Serviço de Vigilância Genômica da Fiocruz Ceará .....</b>	<b>44</b>
<b>5.3 Avaliação da prevalência de variantes de SARS-CoV-2 no estado do Ceará no ano de 2021, em comparação com outros estados brasileiros .....</b>	<b>46</b>
<b>5.4 Listagem dos principais desafios e limitações para o serviço, assim como possíveis soluções para os problemas enfrentados.....</b>	<b>49</b>
<b>5.5 Levantamento do percentual de ensaios de inferência confirmados através de sequenciamento .....</b>	<b>51</b>
<b>5.6 Comparação e otimização de diferentes protocolos de preparo de amostras para sequenciamento .....</b>	<b>55</b>
<b>6 PERSPECTIVAS .....</b>	<b>61</b>
<b>7 CONCLUSÃO.....</b>	<b>62</b>
<b>8 REFERÊNCIAS .....</b>	<b>63</b>
<b>ANEXO A – INICIADORES DO PROTOCOLO MIDNIGHT™.....</b>	<b>68</b>
<b>ANEXO B – INICIADORES DO PROTOCOLO ARTICV3™.....</b>	<b>70</b>

## 1 INTRODUÇÃO

Desde quando foi designado pela Organização Mundial da Saúde (OMS) como um grande surto epidemiológico sem precedentes em janeiro de 2020 e posteriormente como uma pandemia de abrangência mundial em março do mesmo ano, o vírus da Síndrome Respiratória Aguda Grave do Coronavírus 2 (SARS-CoV-2), responsável por causar a doença do Coronavírus 2019 (COVID-19), rapidamente se espalhou pelo mundo, impactando severamente os sistemas de saúde de inúmeros países (CHAVEZ et al., 2021). A ausência de tratamento específico imediato, somado à fácil disseminação do vírus pelo ar, gerou e ainda tem gerado complicações substanciais no controle da infecção (BHAGAT et al., 2021; CASCELLA et al., 2020), influenciado pela grande aglomeração de pessoas em meios de transporte e locais públicos, em uma verdadeira globalização viral (TSANG et al., 2021). O agente causador desse surto é um vírus, que está sujeito a alterações esporádicas em seu material genético, resultado do acúmulo natural de mutações aleatórias e de seu processo evolutivo.

No final de 2020, o termo Variante de Preocupação (do inglês, VOC – *variant of concern*) emergiu em meio ao grande número de variantes já catalogadas, mas que estavam associadas com uma ampla expansão e ao aumento do número de casos e da patogenicidade. Essas variantes se caracterizam principalmente por meio de mutações em seu gene da proteína *Spike* (gene S), afetando a afinidade do domínio de ligação de receptores e a eficiência de neutralização de anticorpos (YANIV et al., 2021). Nesses casos, mutações que ocorrem durante o processo de replicação do vírus na célula hospedeira se fixam no genoma, gerando formas mais competentes devido a uma vantagem ganha como parte do processo evolutivo (YANIV et al., 2021). À época, uma nova linhagem designada como P.1, conhecida posteriormente como variante Gama, foi detectada inicialmente em dezembro de 2020 no estado do Amazonas e logo foi reconhecida como uma linhagem emergente (NAVECA et al., 2021). Essa nova variante se disseminou rapidamente em território nacional e foi apontada como uma das causas prováveis da segunda onda em diferentes estados do Brasil, devido às altas taxas de transmissão e maior escape imunológico, ou seja, estavam associados com altos índices de reinfecção em pessoas que já haviam sido infectadas anteriormente. No final de 2020, a linhagem de origem india designada como B.1.617.2 e, posteriormente, como variante Delta, foi detectada inicialmente no estado de Maharashtra na Índia e se espalhou para o resto do mundo, incluindo o Brasil,

devido ao maior escape vacinal aos anticorpos humanos, além de sua maior eficiência de replicação (MLCOCHOVA et al., 2021).

Nesse cenário, a Rede Genômica da Fiocruz tem como foco a identificação e o monitoramento da circulação das diferentes variantes de SARS-CoV-2 no Brasil, tendo vital importância para iniciar uma resposta política de contenção da pandemia. O monitoramento do perfil epidemiológico e laboratorial dos casos de COVID-19 é feito através das unidades sentinelas da Rede Genômica em diversos estados do Brasil, por meio da coleta de amostras clínicas de pacientes pelos hospitais e unidades de saúde, seguido pelo encaminhamento aos laboratórios de referência para pesquisa e notificação dos casos. O surgimento e expansão de novas linhagens do vírus evidenciam o desafio de reforçar a capacidade de detecção, através da ampliação e fortalecimento da vigilância genômica, do acompanhamento do caminho dessas transmissões e do mapeamento do material genético no decorrer do tempo, uma forma de monitorar quais linhagens do vírus merecem atenção e preocupação nos estados (SALLAS et al., 2021). Durante o surto de SARS-CoV-2, a vigilância sanitária se encontra ainda mais presente em diversos países, fornecendo informações fundamentais sobre o rastreamento e evolução do vírus, otimização de testes moleculares, tratamentos e vacinas, orientando, assim, as respostas para a saúde pública (BRITO et al., 2021).

Os procedimentos de rastreamento do SARS-CoV-2 que utilizam ferramentas genômicas mostraram reduzir o tamanho e duração de um surto. Essas ferramentas também fornecem informações detalhadas sobre as linhagens de transmissão (HARPER et al., 2021). Até o momento, a melhor e mais confiável forma de determinar linhagens do SARS-CoV-2 é através de sequenciamento. Apesar de completa, essa abordagem se mostrou cara e limitante para o grande número de amostras recebidas durante uma pandemia, sendo inacessível para conclusões primárias. Além disso, amostras com Limite de Ciclo (C<sub>t</sub>) elevado tem altas chances de falha no sequenciamento, devido à baixa quantidade de sequência alvo presente na amostra (SILVA et al., 2020).

Nesse momento, o diagnóstico laboratorial fornece um papel fundamental tanto para o prognóstico e acompanhamento, bem como para estudos de epidemiologia molecular, através da utilização da transcrição reversa, seguida da reação em cadeia da polimerase em tempo real (RT-qPCR) para o diagnóstico do SARS-CoV-2 (SILVA et al., 2020). Esse procedimento pesquisa o patógeno diretamente no seu ácido nucleico, permitindo a detecção precoce e a sua diferenciação. Os ensaios utilizados buscam definir mudanças no genoma viral, como a inserção, deleção ou substituição de nucleotídeos, que fornecem informações

preliminares para a seleção de amostras a serem confirmadas através do sequenciamento. Os ensaios de genotipagem rastreiam apenas algumas partes ou alvos específicos do genoma viral, diferente do sequenciamento, que busca identificar o genoma completo do vírus. Os protocolos de genotipagem já são bem estabelecidos para outros vírus, como o vírus da imunodeficiência humana (HIV), no entanto, um painel de genotipagem de SARS-CoV-2 vem sendo desenvolvido e utilizado para o fornecimento de dados genômicos para a identificação dos surtos (HARPER et al., 2021). Para o sequenciamento de novas variantes que estão sendo detectadas no Ceará, são necessários métodos cada vez mais sofisticados para englobar a detecção de novas mutações definidoras de linhagem em amostras sequenciadas, e a rápida notificação para o Ministério da Saúde. Para isso, diferentes protocolos foram testados tanto para a conversão do RNA viral em cDNA, como para a amplificação desse cDNA através da PCR, utilizando diferentes conjuntos de iniciadores (*primers*).

## 2 OBJETIVOS

### 2.1 Objetivo geral

Elaboração e otimização de um modelo de processo operacional, abordando todas as etapas logísticas e técnicas para o estabelecimento local de um Serviço de Vigilância Genômica de SARS-CoV-2 da Rede Fiocruz, em parceria com o serviço de assistência diagnóstica da COVID-19 do Centro de Hematologia e Hemoterapia do Ceará (Hemoce).

### 2.2 Objetivos específicos

1. Avaliação e otimização do fluxo operacional para o Serviço de Vigilância Genômica de SARS-CoV-2 no Ceará, padronizado pela Rede Genômica da Fiocruz;
2. Determinação da quantidade e abrangência de amostras sequenciadas pelo Serviço de Vigilância Genômica da Fiocruz Ceará;
3. Investigar a prevalência de variantes de SARS-CoV-2 sequenciadas no estado do Ceará no ano de 2021, em comparação com outros estados brasileiros;
4. Elencar e descrever os principais desafios e limitações para o serviço, assim como possíveis soluções para os problemas enfrentados;
5. Avaliar comparativamente os protocolos de inferência de variantes de SARS-CoV-2 por RT-qPCR, através do levantamento estatístico do percentual de ensaios confirmados através de sequenciamento;
6. Comparar e otimizar diferentes protocolos de preparo de amostra e de painéis de amplificação para sequenciamento através da utilização de ferramentas de bioinformática e análises de qualidade.

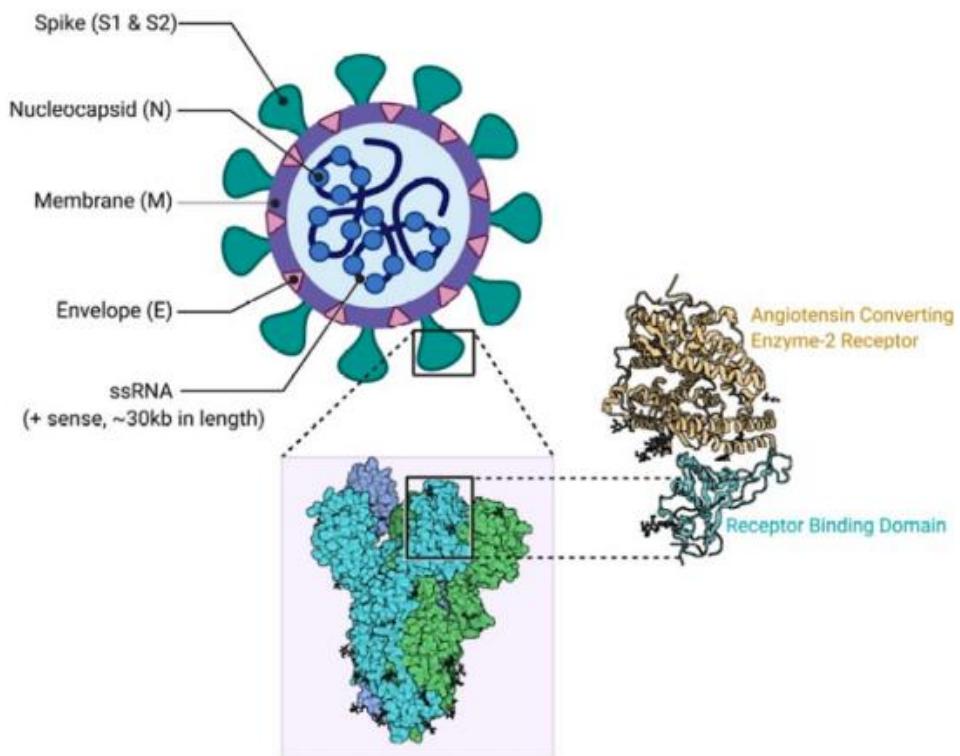
### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Agente etiológico e fisiopatologia do SARS-CoV-2

Os coronavírus (CoV) são um grupo de vírus respiratórios de RNA com grande potencial de infecção e disseminação. Antes de 2019, apenas seis cepas de coronavírus eram conhecidas por infectar humanos e causar doenças respiratórias, a saber: (HCoV-229E, HCoV-OC43, HCoV-NL63, HKU1, SARS-CoV e MERS-CoV). Dentre eles, os de maior letalidade são os SARS-CoV e MERS-CoV, pois podem levar a uma condição severa de saúde em humanos, chamada de SRAG, ou síndrome respiratória aguda grave (Hasöksüz, Kılıç, & Saraç, 2020). No entanto, mais recentemente em dezembro de 2019, surgiu na China um novo tipo de Coronavírus humano denominado de SARS-CoV-2, causador da doença COVID-19, que se disseminou muito rapidamente pelo mundo, superando os quadros de epidemia observados anteriormente, tanto em número de pessoas infectadas, como em óbitos, comprovando sua grande transmissão e adaptabilidade (Wu et al., 2020; Hui et al., 2020).

A família Coronaviridae é pertencente à ordem Nidovirales, que correspondem a vírus compostos por um genoma de RNA de fita simples, medindo em média 30 kilobases. A subfamília Orthocoronavirinae contém quatro gêneros (*Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* e *Deltacoronavirus*), sendo o SARS-CoV-2 pertencente ao gênero *Betacoronavirus*. Esse vírus possui um genoma de RNA não segmentado, contendo quatro proteínas estruturais principais, responsáveis pela montagem do capsídeo viral: a proteína do nucleocapsídeo (gene N), a proteína transmembrana (gene M), a proteína do envelope (gene E) e a proteína Spike (gene S) (Figura 1). Para infectar uma célula humana, o SARS-CoV-2 emprega a enzima conversora de angiotensina 2 (ACE2), uma proteína transmembrana ligada às células humanas, como seu receptor. Depois de se ligar ao receptor, o vírus obtém acesso ao citosol da célula hospedeira, que ocorre geralmente através da proteína *Spike*, seguida pela fusão das membranas celular e viral (HASOKSUZ et al., 2020).

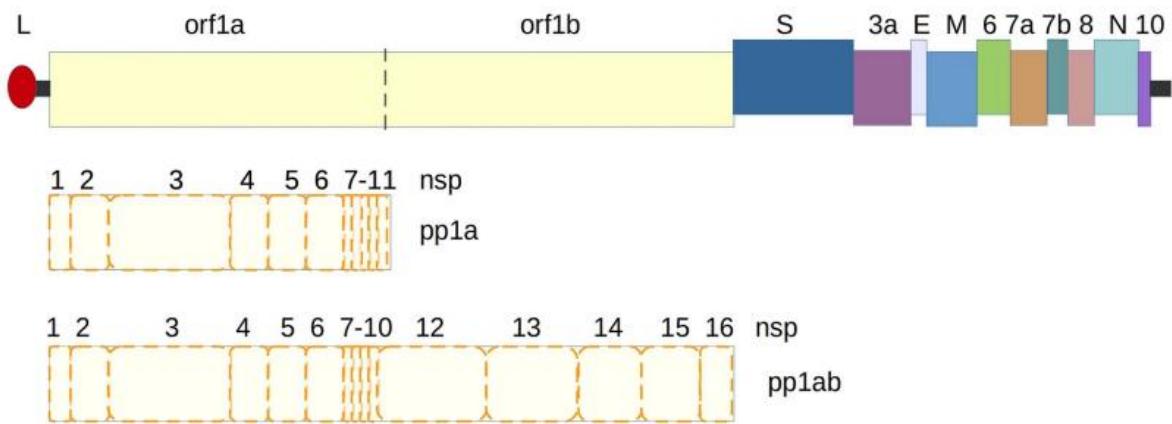
Figura 1 - Representação da estrutura do vírus SARS-CoV-2 e da enzima conversora de angiotensina 2 (ACE2).



Fonte: Hasöksüz, 2020.

O arranjo do genoma de SARS-CoV-2 não é atípico. O gene replicase, que consiste em duas longas e sobrepostas cadeias de leitura (ORF1a e ORF1b), ocupa cerca de dois terços dos aproximados 30.000 pb do genoma (Figura 2). A ORF1a e ORF1b são responsáveis pela tradução da poliproteína 1a (pp1a) e poliproteína 1b (pp1b), respectivamente. Essas poliproteínas são posteriormente processadas em 16 proteínas não estruturais (NSPs) necessárias para a transcrição e replicação do genoma viral (CHIARA et al., 2021).

Figura 2 - Arquitetura do genoma de SARS-CoV-2. O painel superior indica os nomes dos genes, com o círculo vermelho na extremidade esquerda representando a sequência reguladora de transcrição (TRS-L). O painel inferior representa as proteínas não estruturais (NSPs), derivados do processamento das pp1a e pp1b.



Fonte: Chiara, 2021.

### 3.2 Nomenclatura dinâmica para linhagens de SARS-CoV-2

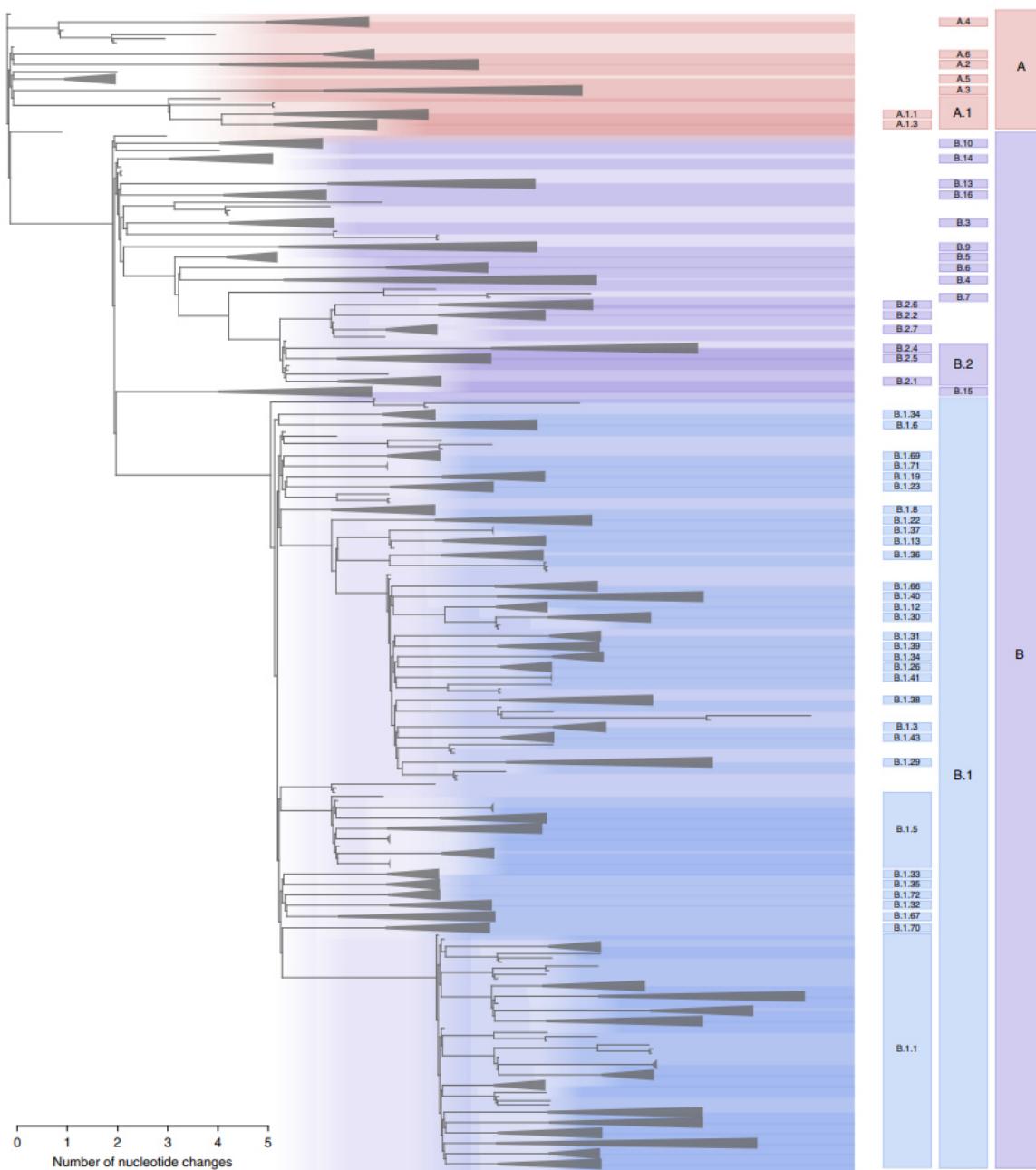
Como outros vírus de RNA, o SARS-CoV-2 está sujeito constantemente a mudanças em seu material genético durante sua replicação em células humanas. No caso do coronavírus, essas mutações vêm sendo acompanhadas em tempo real, através de sequenciamento e estudos mais detalhados.

Não existe uma abordagem universal para classificar a diversidade genética viral. Normalmente essa diversidade é categorizada em “clados” distintos, cada um correspondendo a um grupo monofilético em uma árvore filogenética. Os clados podem ser referidos por uma variedade de termos, dependendo do seu nível taxonômico ou a literatura científica estabelecida para esse vírus. Mudanças e atualizações na taxonomia de vírus ocorrem frequentemente e são o resultado de um processo de várias etapas (WALKER et al., 2019). No caso do SARS-CoV-2, é utilizado o termo “linhagem” ao invés de “clado”, “genótipo” ou outra designação, devido ao seu grande dinamismo, não dependendo de uma estrutura hierárquica estática e exclusiva. Em uma epidemia contínua e em rápida mudança como a do SARS-CoV-2, um sistema de nomenclatura auxilia na epidemiologia e vigilância, fornecendo rótulos comumente aceitos para se referir a determinada versão do vírus circulando em diferentes partes do mundo (RAMBAUT et al., 2020).

Segundo RAMBAUT et al., 2020, os principais desafios para um sistema de nomenclatura utilitários são: i) ser válido e amplamente aceito; ii) capturar os padrões globais e locais da diversidade genética do vírus em tempo hábil; iii) rastrear as novas versões do vírus à medida em que se movem entre os países e dentro de cada país; iv) ser suficientemente robusto e flexível para acomodar a diversidade de novas linhagens à medida em que são gerados e; v) ter dinamismo para incorporar tanto o nascimento quanto a morte de linhagens virais ao longo do tempo. Especificamente em relação à COVID-19, há um desafio ainda maior, pois os dados de sequenciamento estão sendo gerados rapidamente e em grande quantidade.

O sistema de classificação que vem sendo utilizado é chamado de nomenclatura PANGO, tendo foco nas mudanças genéticas associadas à importantes eventos epidemiológicos e biológicos, baseada na presença e ausência de determinadas mutações (RAMBAUT et al., 2020). Os rótulos das linhagens principais começam com uma letra, sendo as duas linhagens presentes na raiz da árvore filogenética chamadas de linhagens A e B. Para as linhagens descendentes são atribuídos valores numéricos, como por exemplo linhagem A.1 e B.2. Dessa forma, essas novas linhagens do vírus servem como ancestrais para os novos descendentes que estão surgindo em outras áreas geográficas, resultando em novas designações de linhagens, como por exemplo A.1.1. Esse procedimento interativo prossegue por no máximo três subníveis, por exemplo A.1.1.1. Novos descendentes dessa linhagem recebem uma letra na sequência alfabética do inglês, de modo que a linhagem A.1.1.1 se tornaria C.1, do mesmo jeito a linhagem A.1.1.2 se tornaria C.2 e assim respectivamente, como mostrado na árvore filogenética de máxima verossimilhança (Figura 3). O sistema PANGO aborda e complementa outros dois sistemas de nomenclatura de SARS-CoV-2 (NextStrain e Gisaid), que foca em clados filogenéticos mais amplos, incorporando critérios para prevalência e persistência mínimos (O'TOOLE et al., 2021).

Figura 3 - Árvore filogenética de máxima verossimilhança de sequências amostrais globais do SARS-CoV-2 obtidas no banco de dados Gisaid.



Fonte: Rambaut et al., 2020.

### 3.3 Variantes de preocupação (VOCs)

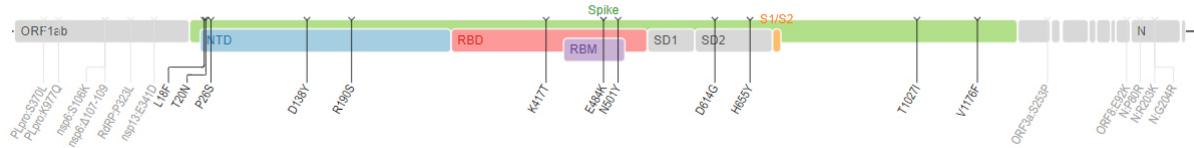
Quando um conjunto de linhagens reúnem mutações distintas em comum, passam a ser chamadas de variante pela OMS e ganham uma nomenclatura específica, dependendo do seu parentesco evolutivo com a cepa original. Geralmente, quando novas variantes se tornam mais comuns, elas o fazem por causa de alguma vantagem seletiva para o vírus (WALENSKY et al., 2021).

Entre a numerosa quantidade de variantes de SARS-CoV-2 que foram detectadas, apenas um pequeno grupo é denominado inicialmente como variante de interesse (VOI) e posteriormente como variante de preocupação (VOC), por serem mais transmissíveis, causarem quadros de doenças respiratórias mais graves e iludir a resposta imune, causando uma infecção e possível escape vacinal (WALENSKY et al., 2021). As VOCs são classificadas pela OMS por uma letra do alfabeto grego, para permitir uma comunicação clara entre grupos de pesquisa. A primeira VOC foi detectada no Reino Unido em setembro de 2020 e classificada como linhagem B.1.1.7 (conhecida como 20I/501Y.V1) ou variante Alfa. Na época, dados de modelagem ilustraram como uma variante muito mais contagiosa que a versão original do vírus, tendo potencial de estender a pandemia nos Estados Unidos e reverter a tendência da época de queda do número de infecções (WALENSKY et al., 2021).

#### 3.3.1 Variante Gama (P.1)

Em novembro de 2020, surgiu no Amazonas a linhagem P.1 (conhecida como GR/501Y.V3), que evoluiu de um clado da linhagem B.1.1.28, sendo posteriormente denominada pela OMS como VOC Gama (NAVECA et al., 2021). A VOC Gama se tornou a causa putativa da segunda onda tanto no Amazonas como no restante do país, sendo exportada também para outros lugares do mundo, se tornando alvo de serviços de vigilância genômica que surgiram em diferentes polos da Fiocruz nos estados brasileiros no final de 2020 e primeiro semestre de 2021. O motivo principal de sua grande disseminação se encontra no grande número de mutações, sendo a maior parte delas concentrada no gene S (Figura 4).

Figura 4 - Arquitetura do genoma da VOC Gama (P.1) destacando suas mutações definidoras de linhagem.

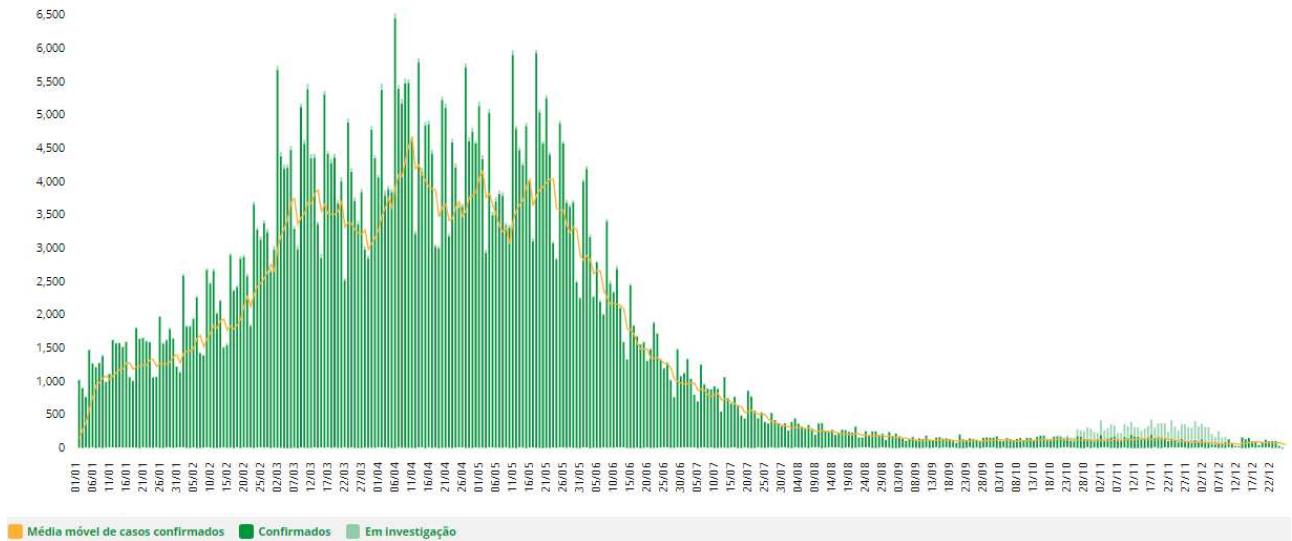


Fonte: Adaptado do banco de dados de SARS-CoV-2 da Universidade de Stanford. Disponível em: <https://covdb.stanford.edu/page/mutation-viewer/>.

O surgimento dessa variante coincidiu com o aumento de internações na região Norte e Sul do Brasil. No Rio Grande do Sul, o número de casos graves de COVID-19 aumentou em 3,8 vezes em fevereiro de 2021, constatando-se através de dados de sequenciamento a ligação entre esse quadro e a predominância da linhagem P.1 na região (YANIV et al., 2021). A segunda onda de crescimento do número de casos no Amazonas também coincidiu com a emergência da VOC Gama no final de novembro de 2020. Dados estatísticos apoiam a conclusão de que sucessivas substituições de linhagem no Amazonas pela P.1 foram impulsionadas pela combinação complexa de níveis variáveis de medidas de distanciamento social, que coroboram com a maior infectividade e transmissão da P.1 (NAVECA et al., 2021).

No Ceará, de acordo com os dados do Boletim Epidemiológico da COVID-19, disponibilizado pela Secretaria de Saúde do Estado, a segunda onda da doença começou por volta do final de fevereiro e início de março de 2021, com um aumento exponencial do número de casos confirmados, se extendendo até julho do mesmo ano (Figura 5).

Figura 5 - Gráfico da taxa móvel de casos confirmados de indivíduos com COVID-19 durante o período entre 01 de janeiro de 2021 e 22 de dezembro de 2021, com o pico no número de casos entre março e maio de 2021.

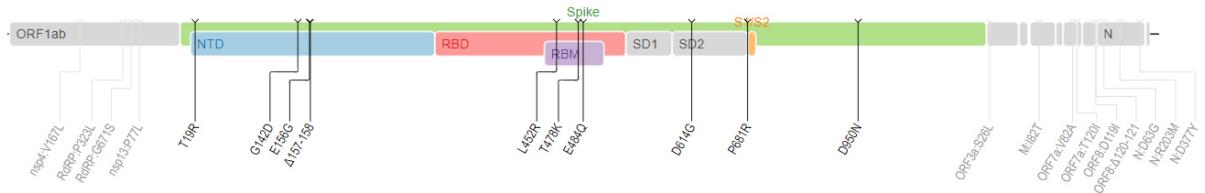


Fonte: Boletim Epidemiológico da COVID-19, disponível em: <https://integrasus.saude.ce.gov.br/> - acesso 27/12/2021.

### 3.3.2 Variante Delta (B.1.617.2)

Ainda no final de 2020, foi identificada no estado de Maharashtra na Índia a linhagem B.1.617.2 (conhecida como GK/478K.V1), sendo posteriormente classificada como VOC Delta, superando linhagens pré-existentes, incluindo a B.1.617.1 (Kappa) e B.1.1.7 (Alfa) (MLCOCHOVA et al., 2021). A VOC Delta e sua nova sublinhagem, a Delta Plus (B.1.617.2.1 ou AY.1), se espalharam pelo mundo, tornando-se a variante mais predominante do vírus, incluindo o Brasil no segundo semestre de 2021, que gerou uma necessidade de vigilância constante para a surgimento de novas sublinhagens com maior infectividade e transmissão. A Delta possui muitas mutações definidoras de linhagem em seu genoma, incluindo as principais regiões do gene S (Figura 6). Já a Delta Plus apresenta um número ainda mais significativo de mutações de alta prevalência. Mutações de assinatura na *Spike* (G142D, A222V e T95I) existem em uma porcentagem mais significativa na variante Delta Plus do que na variante Delta (KANNAN et al., 2021).

Figura 6 - Arquitetura do genoma da VOC Delta (B.1.617.2) destacando suas mutações definidoras de linhagem.



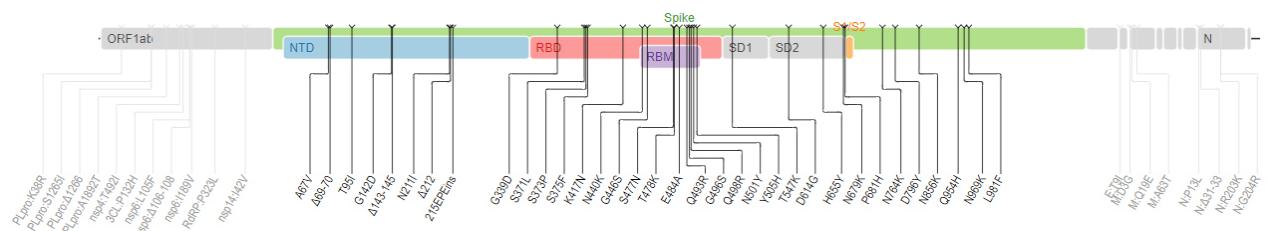
Fonte: Adaptado do banco de dados de SARS-CoV-2 da Universidade de Stanford. Disponível em: <https://covdb.stanford.edu/page/mutation-viewer/>.

Inúmeras sublinhagens derivadas da variante Delta surgiram em diversos países no ano de 2021. A sublinhagem B.1.617.2.99.2 ou simplesmente AY.99.2, surgiu no Brasil e se tornou a mais predominante considerando o número de casos relatados no segundo semestre de 2021. Essa nova sublinhagem possui duas mutações adicionais no gene ORF1a em relação à versão ancestral da Delta, segundo dados do sistema de classificação PANGO, disponível em <https://cov-lineages.org/>.

### 3.3.3 Variante Ômicron (B.1.1.529)

Muito recentemente, em novembro de 2021, foi descoberta a variante Ômicron (B.1.1.529) em Botswana na África. Esta nova variante mostrou grau acentuado de mutações, em comparação com as versões anteriores do SARS-CoV-2, incluindo ao menos 32 na região da proteína *Spike* (Figura 7). Estudos indicam que essas mutações aumentaram a capacidade da Spike em se conectar ao receptor ACE2 de células hospedeiras (KANDEEL et al., 2021).

Figura 7 - Arquitetura do genoma da VOC Ômicron (B.1.1.529) destacando suas mutações definidoras de linhagem.



Fonte: Adaptado do banco de dados de SARS-CoV-2 da Universidade de Stanford. Disponível em: <https://covdb.stanford.edu/page/mutation-viewer/>

Em dezembro de 2021, a Agência de Segurança em Saúde do Reino Unido confirmou a primeira morte causada pela variante Ômicron, considerando estado de emergência no país, devido ao escape vacinal relacionado à variante e ao número alarmante de novos casos (mais de 70.000 em 15 de dezembro de 2021) (BURKI et al., 2021). Estudos estão sendo feitos por todo o mundo e novos dados ainda precisam ser disponibilizados para comprovar a real capacidade da variante Ômicron em aumentar o número de casos e agravar a condições clínica dos pacientes positivos.

### 3.4 Detecção de novas variantes

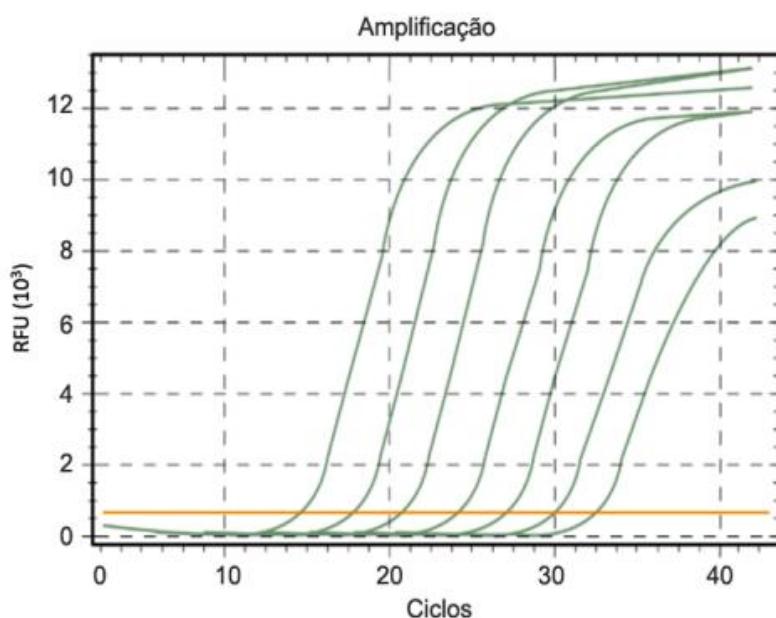
O potencial de novas variantes para escape da resposta imune, reduzindo a eficácia das vacinas, permanece uma ameaça. Os laboratórios de todo o mundo se encontram na busca de identificar e catalogar essas novas linhagens do vírus que circulam na população o mais rápido possível. Para acelerar a vigilância genômica, novos ensaios de diagnóstico e métodos abordando ensaios de inferência vêm sendo desenvolvidos para uma resposta mais rápida e precisa (SILVA et al., 2020).

As primeiras sequências genômicas completas do novo *betacoronavírus* foram obtidas no final de dezembro de 2019, através de abordagens metagenômicas e sequenciamento Sanger. A disponibilidade de um conjunto de genomas de referência no NCBI (1798172431 - NC\_045512.2) facilitou o desenvolvimento de testes diagnósticos baseados em RT-qPCR e sequenciamento de nova geração (NGS). Várias estratégias baseadas em sequenciamento foram usadas com sucesso no passado para estudar a origem e compreender a evolução dos agentes infecciosos, investigando as cadeias de propagação e transmissão de surtos (CHIARA et al.,

2021). O sequenciamento se mostrou o método mais utilizado para a determinação precisa da linhagem da amostra e monitoramento de eventuais alterações no código genético, por identificar, na ordem correta, a sequência de nucleotídeos da molécula de DNA ou RNA, visando conhecer a informação genética contida nessa estrutura (GRADA & WEINBRECHT, 2013).

Os dados de sequenciamento de SARS-CoV-2 derivam principalmente de amostras com alta carga viral, provindas do diagnóstico clínico prévio por RT-qPCR. Diferente da PCR convencional, onde os resultados são visualizados através de padrões de bandas de eletroforese, a RT-qPCR utiliza iniciadores específicos e sondas marcadas com compostos fluorescentes. A emissão e detecção da fluorescência acontece durante a reação de PCR, em que a fluorescência mínima da fase exponencial de amplificação corresponde ao *threshold*, traçado horizontalmente no gráfico (Figura 8). Esse valor é utilizado para calcular o limite de ciclo (Ct) de cada amostra. O Ct corresponde ao número de ciclos de PCR necessários para o início da amplificação, sendo assim, inversamente proporcional à quantidade de sequência alvo presente na amostra (SILVA et al., 2020).

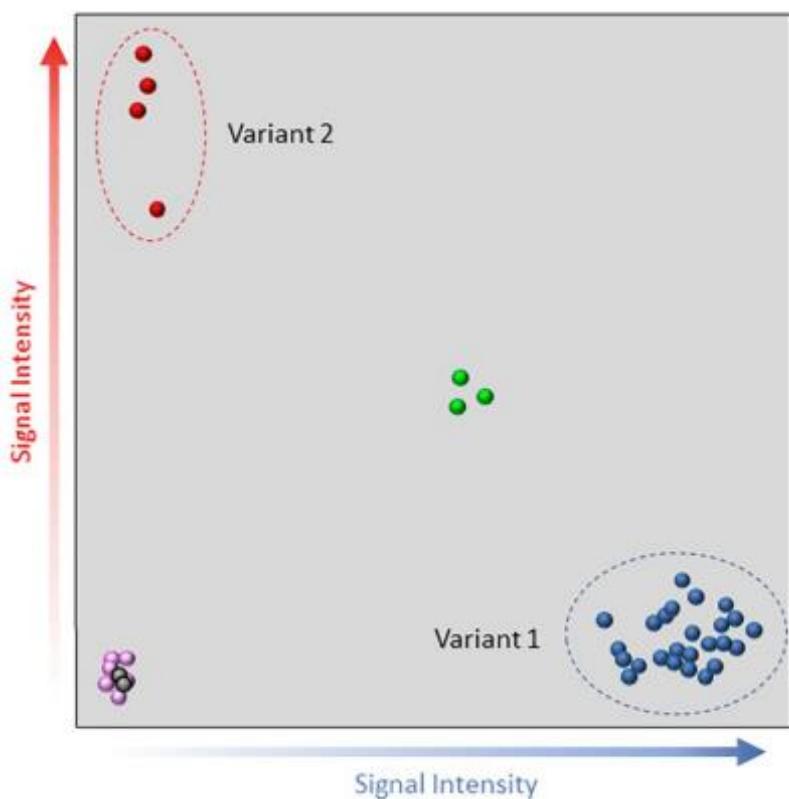
Figura 8 - Gráfico de quantificação de produtos de PCR em termociclador em tempo real. O gráfico mostra as quantidades absolutas de material genético em função do número de ciclos. Em verde: curvas de amplificação; Em amarelo: valor de *threshold*.



Fonte: Silva, 2020.

Devido principalmente ao alto custo envolvido e o tempo de espera pelo resultado do sequenciamento, novas alternativas vêm sendo desenvolvidas para um resultado preliminar, necessitando ser confirmado posteriormente através de sequenciamento. Os testes de inferência e genotipagem são uma alternativa rápida, de alto rendimento e menor custo para a triagem de amostras positivas (HARPER et al., 2021). Esses testes identificam um painel de marcador mínimo que pode definir genótipos distintos, como uma substituição, deleção ou inserção de nucleotídeos. A abordagem também pode utilizar a intensidade da fluorescência para a determinação de diferentes alelos (Figura 9), onde quanto maior o sinal mais forte a informação da amostra tendendo para um determinado valor.

Figura 9 – Desenho esquemático de genotipagem por meio de PCR em tempo real na qual mostra a intensidade da fluorescência para determinação de alelos. Em vermelho: homozigoto para o alelo 1; Em azul: homozigoto para o alelo 2; Em verde: heterozigoto para ambos os alelos; Em preto e roxo: resultado negativo/ controle negativo.



Fonte: Harper, 2021.

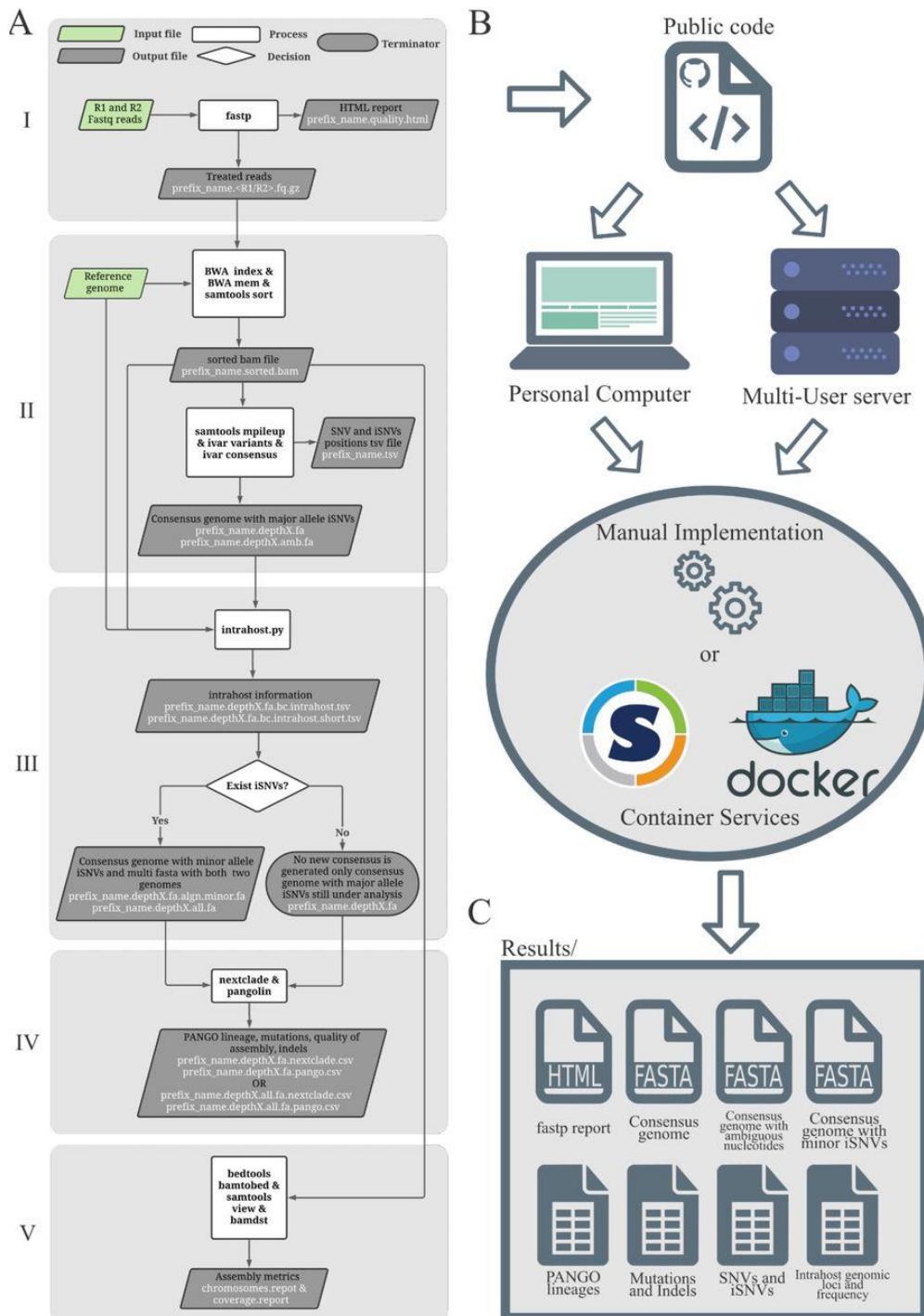
Dentre as variantes a serem identificadas, a Gama e a Delta possuem alguns padrões de mutações específicos. Um deles ocorre no final da região da ORF8 e início do gene N (28227-28286pb) para a VOC Gama, que inclui uma inserção de quatro nucleotídeos específicos. A VOC Delta, por sua vez, apresenta a deleção 157/158 localizada na região do gene S (21989-22083 pb). Ambas são alvo de ensaios com iniciadores e sondas para o seu monitoramento, através de tecnologias integradas com o DNA e a utilização de fluoróforos (YANIV et al., 2021).

### 3.5 Ferramentas de bioinformática para predição de linhagens

O esforço global para sequenciar e identificar as linhagens circulantes de SARS-CoV-2 resultou na rápida disponibilidade de milhões de genomas em bancos de dados internacionais, como o banco EpiCoV™ hospedado no GISAID, representando mais de 1.700 linhagens descritas no sistema PANGO (SHU & MCCUALEY, 2017). Apesar dos métodos de biologia molecular desenvolvidos através de PCR, apenas o sequenciamento do genoma completo pode fornecer informações genéticas suficientes para a discriminação de linhagens de forma confiável e caracterização de padrões de mutações pelo genoma, necessários para a caracterização de VOCs (LAURING & HODCROFT, 2021).

Dessa forma, a utilização de ferramentas de bioinformática, utilizando linguagem de programação e análise de dados se mostra necessária para a montagem correta do genoma após o sequenciamento e verificação correta de mutações. O fluxo de trabalho atualmente em uso pela Rede de Vigilância Genômica de COVID-19 da Fiocruz, denominado ViralFlow, foi desenvolvido pelo Instituto Aggeu Magalhães, unidade da Fiocruz em Pernambuco, funcionando com leituras de sequenciamento de amplicons Illumina emparelhados, através da “conteinerização” com o Docker (<https://www.docker.com/>) e Singularity (<https://sylabs.io/singularity/>) (DEZORDI et al, 2021). O fluxo utiliza um conjunto de *scripts* e ferramentas para o pré-processamento dos dados, corte de adaptadores, alinhamento, predição de linhagens e busca por possíveis casos de contaminação e coinfeção nas amostras (Figura 10). Há também a possibilidade de utilização de outros fluxos para a montagem de genomas, como é o caso do DRAGEN™ Bio-IT, desenvolvido pela Illumina, ou uma montagem manual através do download das ferramentas utilizadas isoladamente.

Figura 10 - Esquema de fluxo de trabalho do ViralFlow. a. Divisão do fluxo em cinco etapas; b. Configuração do fluxo de trabalho para funcionar em diferentes ambientes computacionais; c. Várias saídas geradas pelo fluxo de trabalho por amostra.



Fonte: Dezordi et al, 2021.

### **3.5.1 Pangolin: atribuição filogenética de linhagens**

Uma das principais ferramentas computacionais para atribuição mais provável das linhagens de SARS-CoV-2 a uma determinada sequência do genoma, é chamada de Pangolin (O'TOOLE et al., 2021). Essa ferramenta se baseia no sistema de nomenclatura de linhagem dinâmico PANGO (RAMBAUT et al., 2020). A identificação e atribuição de linhagens filogenéticas de SARS-CoV-2 não é um problema trivial, devido ao grande tamanho e rápido crescimento do conjunto de dados do genoma global do vírus. Os ramos de filogenia acumulam aproximadamente uma substituição de nucleotídeo a cada duas semanas em média (DUCHÈNE et al. 2020). Consequentemente, não é possível definir linhagens de alta resolução usando apenas a presença ou ausência de alterações de nucleotídeos em um subconjunto de locais variáveis. Dessa forma, apenas usando alinhamentos completos do genoma podemos estimar o posicionamento mais provável de uma nova sequência dentro da filogenia global do SARS-CoV-2 e, assim, atribuir um nome para a linhagem ( O'TOOLE et al., 2021). O Pangolin é disponibilizado gratuitamente para download pelo site: <https://cov-lineages.org/resources/pangolin.html>.

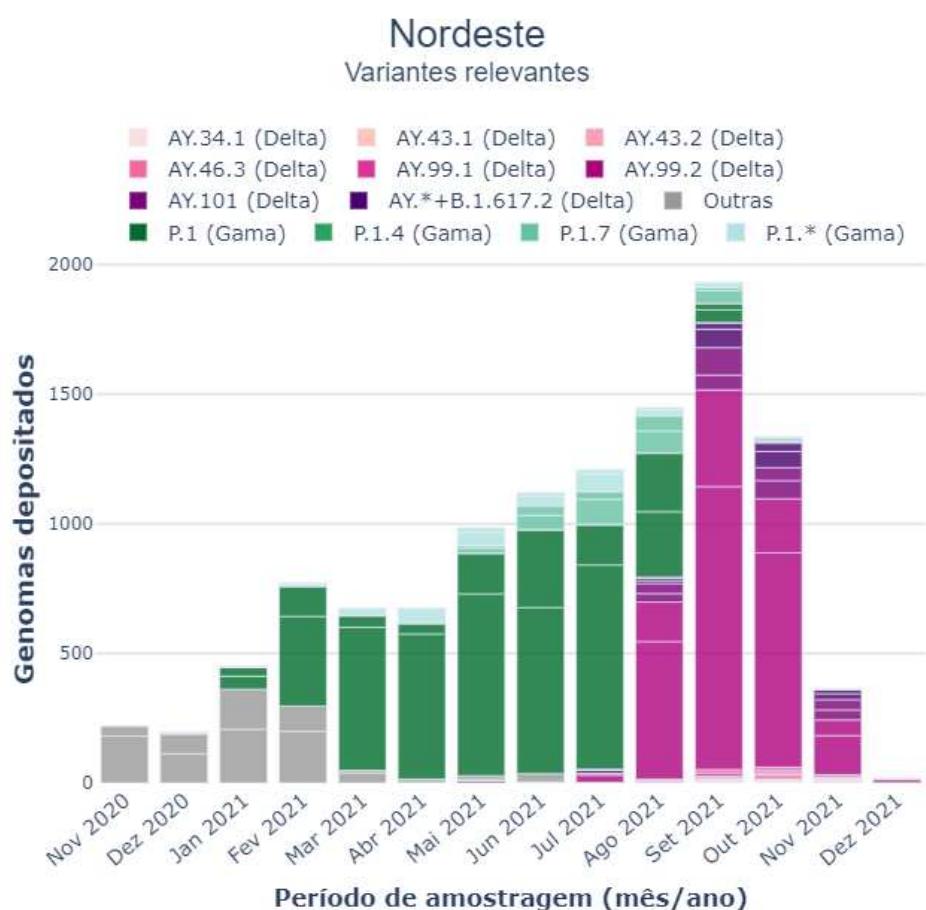
### **3.5.2 Nextclade: atribuição de clado, chamada de mutação e verificação de qualidade de sequências**

Uma outra ferramenta utilizada para predição rápida de linhagens durante a epidemia de SARS-CoV-2 é o Nextclade e sua versão *web*. Essa ferramenta aceita dados de sequência no formato FASTA, realizando um alinhamento com uma sequência de referência, onde a utiliza para detectar mudanças evolutivas (AKSAMENTOV et al., 2021). A ferramenta é construída para um *feedback* rápido, podendo ser acessado remotamente através do site <https://clades.nextstrain.org/> ou via linha de comando. O Nextclade é construído via alinhador de sequências em pares ciente de códons para genomas virais semelhantes, o que evita inequívocos na chamada de alteração de aminoácidos associadas a alterações na sequência de nucleotídeos, permitindo determinar as mutações de cada sequência de consulta, realizando uma pesquisa exaustiva para a correspondência mais próxima em uma árvore filogenética, representando a diversidade populacional.

### 3.6 Prevalência das principais variantes e linhagens de SARS-CoV-2 no Brasil em 2021

Segundo o painel de variantes de SARS-CoV-2, disponibilizado pela Rede Genômica da Fiocruz no banco de dados EpiCoV™ hospedado no GISAID, o Brasil apresenta mais de 91000 genomas sequenciados até dezembro de 2021. Na região do Nordeste, apesar da linhagem P.1 ter surgido no final de 2020, ainda não se encontrava dominante no período até final de janeiro de 2021, com a prevalência de outras linhagens. De fevereiro a julho de 2021, a VOC Gama se tornou predominante em praticamente todas as amostras sequenciadas, dando lugar a VOC Delta a partir de agosto (Figura 11).

Figura 11 - Principais variantes de preocupação sequenciadas pela Rede Genômica da Fiocruz na região do nordeste no período de novembro de 2020 a dezembro de 2021.

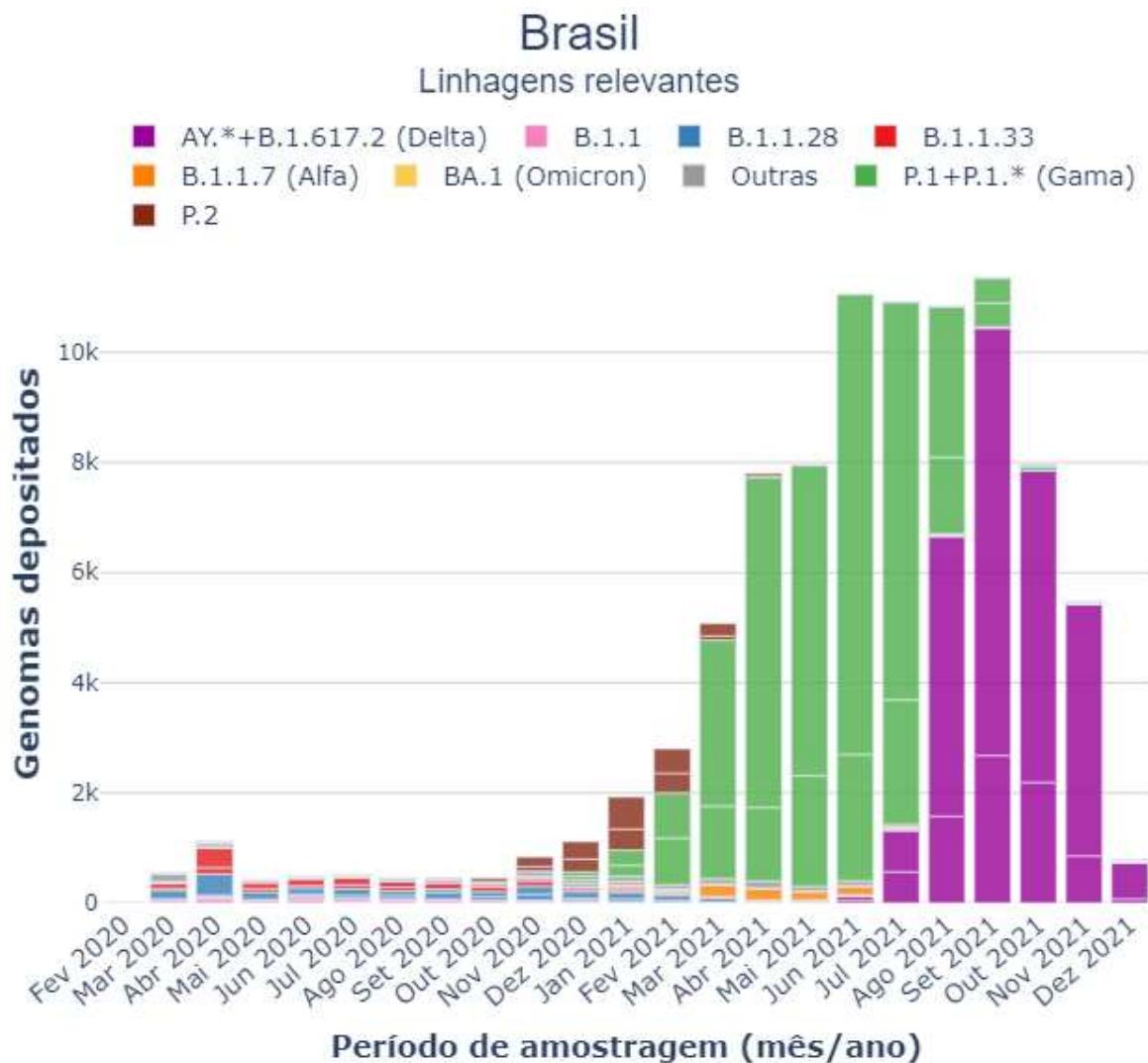


Fonte: Rede Genômica da Fiocruz, disponível em: <http://www.genomahcov.fiocruz.br/dashboard/>.

Acesso em: 31/12/2021.

De acordo com os dados de Naveca et al., 2021, a linhagem P.1 surge no Brasil no período de outubro de 2020. Esse dado é apresentado no gráfico da Figura 12, que mostra o padrão de linhagens sequenciadas para o Brasil como um todo. Nos anos anteriores, a prevalência era de linhagens ancestrais, como a B.1.1.28 e B.1.1.33. A variante Delta surge no Brasil em agosto de 2021, se estendendo até o fim do ano.

Figura 12 - Frequência de variantes de preocupação sequenciadas pela Rede Genômica da Fiocruz no Brasil durante o período de fevereiro de 2020 a dezembro de 2021.



Fonte: Rede Genômica da Fiocruz, disponível em: <http://www.genomahcov.fiocruz.br/dashboard/>.

Acesso em: 30/12/2021.

## 4 METODOLOGIA

Este estudo tem caráter exploratório e experimental e consiste na construção e validação de um modelo de serviço a partir dos trabalhos desenvolvidos e testes realizados pela Rede Genômica da Fiocruz Ceará, no cenário da pandemia mundial causada pela COVID-19.

O planejamento experimental do projeto consistiu na elaboração de um fluxograma analisador, contemplando todas as etapas e processos de vigilância genômica, seguido com a validação do desenho proposto pelos analistas e desenvolvedores do serviço prestado e comparação dos resultados obtidos com dados da Secretaria de Saúde do Ceará e de outros serviços de vigilância genômica distribuídos pelo país. Em um segundo momento, o trabalho buscou aprimorar o modelo apresentado através da compilação dos principais problemas envolvidos, seguido por sugestões de possíveis soluções e testes para aperfeiçoamento de protocolos. Essa atividade foi realizada através da validação dos testes de inferência e a otimização de diferentes protocolos de preparo de amostras para sequenciamento, por meio da utilização de ferramentas de bioinformática e análise de dados.

### 4.1 Elaboração e otimização do fluxo operacional

Esta etapa do estudo buscou elaborar e estabelecer um fluxo de trabalho para o Serviço de Vigilância Genômica de COVID-19 no Ceará, com o intuito de organizar e aprimorar o serviço prestado. O processo metodológico consistiu em três momentos e levou em consideração a trajetória ideal da amostra desde o seu recebimento e triagem até a sua notificação e depósito das sequências genômicas.

O primeiro momento foi realizado através do acompanhamento teórico e prático de todas as atividades envolvidas na vigilância genômica, mediante entrevistas com os analistas responsáveis por cada etapa do processo e coleta de dados em documentos normatizados de cada um dos setores. Dessa forma, foi possível haver uma compreensão mais ampla dos estágios do fluxo partindo de cada atividade isoladamente.

Em um segundo momento, foi feito o desenho do fluxograma analisador, que é uma representação gráfica do processo de trabalho e das ações tomadas a partir da demanda vigente. Os fluxos foram elaborados utilizando o programa Lucid chart, disponível em [www.lucidchart.com.br](http://www.lucidchart.com.br), e divididos em três partes principais, de acordo com a trajetória da amostra: triagem, sequenciamento e bioinformática. Ele é representado por três símbolos,

convencionados universalmente: a elipse representa sempre a entrada ou saída do processo de produção de serviços; o losango indica os momentos em que deve haver uma decisão para continuidade do trabalho ou divisão do fluxo e o retângulo, que diz respeito ao momento de intervenção, ação, sobre o processo (MERHY, 2007, p. 150). Além deles, foram utilizados mais dois símbolos: o hexágono que representa uma preparação para a próxima etapa, através de um controle de qualidade e o cilindro que representa o armazenamento final de todos os dados.

Em um terceiro e último momento, foi realizada uma análise reflexiva do material produzido, a fim de que se identifique possíveis falhas no fluxo, mediante padronização da Fiocruz, através da apresentação, alteração e validação de todas as etapas do processo pelo coordenador da Rede de Vigilância Genômica da Fiocruz Ceará e demais membros da equipe. A análise foi fundamentada na aplicação do fluxograma como orientação das atividades desempenhadas e em eventuais tomadas de decisão.

#### **4.2 Avaliação do quantitativo e da abrangência de amostras sequenciadas pelo Serviço de Vigilância Genômica da Fiocruz Ceará**

Esta etapa do estudo buscou avaliar o quantitativo de amostras sequenciadas pelo serviço e a sua abrangência entre os municípios do Ceará, com a intenção de avaliar a capacidade do fluxo em contemplar amostras de todas as regiões do estado.

##### ***4.2.1 Obtenção dos dados***

Os dados foram obtidos através do Gerenciador de Ambiente Laboratorial (GAL) da Fiocruz e do Hemoce, disponibilizados pelos profissionais em saúde responsáveis por fazer a coleta das amostras. As informações foram organizadas em planilhas pelos analistas da Fiocruz, servindo de base para as análises realizadas.

#### **4.2.2 Critérios de exclusão**

Foram excluídas do estudo repetições e amostras de pacientes que não informaram o CEP ou município de residência. Pacientes que residem em outros estados do Brasil também foram excluídos das análises.

#### **4.2.3 Distribuição geográfica e randomização das amostras**

Foi realizada uma organização das amostras para validação da representatividade do estudo em contemplar a maioria dos municípios do Ceará, de acordo com a localização geográfica entre as mesorregiões do estado: Noroeste Cearense, Norte Cearense, Metropolitana de Fortaleza, Sertões Cearenses, Jaguaribe, Centro-sul Cearense e Sul Cearense. As mesorregiões foram definidas pelo IBGE almejando a organização do espaço geográfico, por meio de três dimensões: o processo social, como determinante; o quadro natural, como condicionante e a rede de comunicação e de lugares, como elemento de articulação espacial. Através da análise integrada destas três dimensões, possibilita-se que o espaço delimitado tenha identidade regional (IBGE, 1990). Para validar a representatividade do estudo em abranger a maioria dos municípios do estado, foi realizado um georreferenciamento utilizando o CEP ou município de residência dos pacientes através da ferramenta Power Bi da Microsoft.

### **4.3 Avaliação da prevalência de variantes de SARS-CoV-2 no estado do Ceará no ano de 2021, em comparação com outros estados brasileiros**

#### **4.3.1 Predição de linhagens**

As linhagens de SARS-CoV-2 foram classificadas utilizando a versão 3.1.17 do software de atribuição filogenética Pangolin. Para a confirmação da variante registrada, foi utilizada a versão *web* do software de atribuição de clado, chamada de mutação de verificação de qualidade Nextclade Covid. O acesso foi feito no dia 03/01/2022.

#### **4.3.2 Critérios de exclusão**

Foram excluídas amostras com Ct elevado ( $>27$ ) e baixos padrões de qualidade, incapazes de predizer com exatidão a variante da amostra mesmo após a repetição. Casos em que o paciente realizou a coleta mais de uma vez e o resultado do sequenciamento resultou na mesma linhagem filogenética também foram descartados.

#### **4.3.3 Organização amostral**

As amostras sequenciadas foram obtidas de todas as quatro fontes mostradas no fluxograma, o que inclui amostras da vigilância genômica, do Hemoce, da Unadig e do Lacen. Foram organizadas de acordo com a data de coleta entre os meses de janeiro a dezembro de 2021, com N amostral de 3513 amostras, fornecendo um estudo temporal da evolução das principais variantes do SARS-CoV-2 ao longo do ano.

#### **4.3.4 Comparação dos resultados**

A partir da análise temporal feita com as amostras sequenciadas do Ceará, foi realizada uma comparação com os dados do Boletim Epidemiológico da COVID-19 da Secretaria de Saúde do Estado do Ceará, e com o painel de variantes da Rede Genômica da Fiocruz, depositadas no banco de dados EpiCoV™, hospedado na plataforma GISAID. O estudo se concentrou na avaliação temporal do crescimento das principais VOCs detectadas pelo serviço em comparação com o crescimento observado no país.

### **4.4 Listagem dos principais desafios e limitações para o serviço, assim como possíveis soluções para os problemas enfrentados**

O serviço, por mais eficiente que seja, nem sempre consegue evitar que erros aconteçam. Falhas são inerentes ao processo, que depende de um grande volume de variáveis que nem sempre estão sobre o controle dos administradores para o fornecimento de um resultado esperado. Processos adequados e sua correta execução são essenciais para se evitar falhas durante as etapas do serviço. Falhas podem ter origem em razão como um projeto mal

desenvolvido, manuseio errado de máquinas e equipamentos ou falhas de pessoal. Além disso, pode ter origem por problemas técnicos, de gestão ou humanos (SLACK et al., 2009).

Levando estas questões em consideração, nessa etapa do estudo foi feito um levantamento de eventuais desafios no serviço prestado, assim como possíveis soluções para os problemas observados, a fim de fornecer um resultado mais rápido e preciso. O processo metodológico contou primeiramente com anotações de eventuais focos de problemas durante o acompanhamento teórico e prático das etapas logísticas e técnicas, assim como uma pesquisa sobre os principais limitantes na execução das atividades. Posteriormente, foram realizadas entrevistas com os analistas responsáveis para uma melhor compreensão da visão da equipe sobre falhas na metodologia.

#### **4.5 Validação dos protocolos de inferência como testes preliminares para a predição de variantes de SARS-CoV-2**

Em 2021, com a prevalência de novas variantes de interesse no Ceará, ensaios de inferência para a detecção rápida de mutações dentre as amostras positivas no diagnóstico para SARS-CoV-2 foram desenvolvidos e aplicados, para futura confirmação através de sequenciamento genômico. Os ensaios apresentados neste trabalho buscam a detecção da VOC Gama, através do acompanhamento de suas principais inserções e deleções. O primeiro ensaio testado buscou monitorar a presença da inserção de quatro nucleotídeos no início da região do Gene N (28263 bp), específico para a variante Gama (P.1), utilizando sonda Taqman™ (5'-/5HEX/CAGACATTTAGTTGTTGTCGT/3MGBEc/-3'). O segundo ensaio buscou monitorar uma deleção de nove nucleotídeos na região ORF1ab (11288 pb), compartilhado entre as variantes Gama (P.1), Alpha (B.1.1.7) e Beta (B.1.351), utilizando também sonda Taqman™ (5'-/56-FAM/GGTTGATACTAGTTGAAGCT/3MGBEc/-3'). Foi feito um ensaio duplex, que utiliza simultaneamente os dois ensaios anteriores. Para todas as estratégias, foi utilizado o kit de reação de PCR MASTERMIX GoTaq da PROMEGA, com iniciadores específicos (senso e anti-senso) tanto para o ensaio de inserção quanto (134ins4fw: 5'-TGCCTGTTCTATGAAGAC-3' e 134ins4rv: 5'-GGTGCATTGCTGATTG-3'), como também para o ensaio de deleção nove (134del9fw: 5'-CACAGTTCTGCCATAAGGA-3' e 134del9rv: 5'-CTGCTAGTTGGGTGATGCGTA-3'), em uma concentração de 300 nM e 0,3 µl, em um volume total de 8,15 µl de Mix e 5 µl de amostra. Para ambos os ensaios realizados foi feito um monitoramento do produto de amplificação através da avaliação da curva

de amplificação após a ciclagem, utilizando como agente fluorescente as sondas TaqMan™ para o monitoramento da reação em tempo real em um termociclador QuantStudio® (Thermo Fisher).

Em seguida, foram realizados ensaios de inferência de genotipagem para a mutação K417T, característico da variante Gama (P.1). Para os ensaios de genotipagem foi utilizado o kit de reação *TaqMan Genotyping Master Mix (Thermo Fisher)*, utilizando um ensaio com sondas Taqman™ customizadas para a região alvo (22761-22861pb), com 0,25 µl da sonda em um volume total de 5 µl de Mix e 5 µl de amostra.

#### **4.5.1 Categorização do percentual de ensaios de inferência confirmados através de sequenciamento**

Foi realizado um levantamento estatístico do percentual de ensaios de inferência confirmados em 20 lotes de sequenciamento. Foram analisadas apenas amostras com alto padrão de qualidade, levando em consideração a cobertura total do genoma, para evitar possíveis erros de predição de linhagem. As análises foram feitas levando em consideração três possíveis resultados. O primeiro deles seria um resultado confirmado, quando o sequenciamento confirma o dado preliminar dos testes de inferência como: “possível P.1” ou “não P.1”. O segundo deles seria não confirmado, quando há uma discordância entre os dados de inferência e sequenciamento. E o terceiro deles seria inconclusivo, quando não foi possível determinar com exatidão o resultado da inferência, não podendo comparar com o dado de sequenciamento.

#### **4.6 Comparação e otimização de diferentes protocolos de preparo de amostras para sequenciamento**

Neste trabalho, foi feita uma comparação entre diferentes lotes de amostras sequenciadas utilizando o protocolo Illumina® COVIDSeq™ tanto para a conversão como para a amplificação do cDNA. Foram testados os iniciadores do protocolo Midnight™ e ArticV3™ para a construção das bibliotecas de cDNA (Anexos 1 e 2).

Em um primeiro teste, três amostras classificadas como VOC Gama, sequenciadas em 2 diferentes lotes utilizando o mesmo cartucho V2 micro, foram utilizadas como parâmetro de comparação entre os iniciadores Midnight™ e ArticV3™. Ambas apresentaram alta cobertura e média de profundidade, o que permitiu a visualização de eventuais problemas através de análises de bioinformática utilizando mapas de profundidade. Através da utilização da

ferramenta de montagem de sequências Geneious prime, foi possível fazer um alinhamento das sequências com uma sequência de referência (1798172431 - NC\_045512.2) utilizando o programa de alinhamento de sequências múltiplo MAFFT, e analisar visualmente as regiões de interesse para a classificação da amostra. Para a melhor comparação entre os iniciadores, foram gerados mapas das mutações encontradas por ambos os ensaios para a proteína *Spike* (gene S), possuidora das principais mutações definidoras da VOC Gama.

Em um segundo testes, outras três amostras classificadas como VOC Delta sequenciadas em um mesmo lote utilizando um cartucho V3 600, foram utilizadas como parâmetro de qualidade dos iniciadores *Midnight™* e *ArticV3™*. Novamente, as amostras foram selecionadas por apresentarem alta cobertura e média de profundidade, o que permitiu a melhor análise dos mapas de profundidade. Para a melhor comparação entre os iniciadores, foram gerados mapas das mutações encontradas por ambos os ensaios para a proteína *Spike* (gene S), possuidora das principais mutações definidoras da VOC Delta.

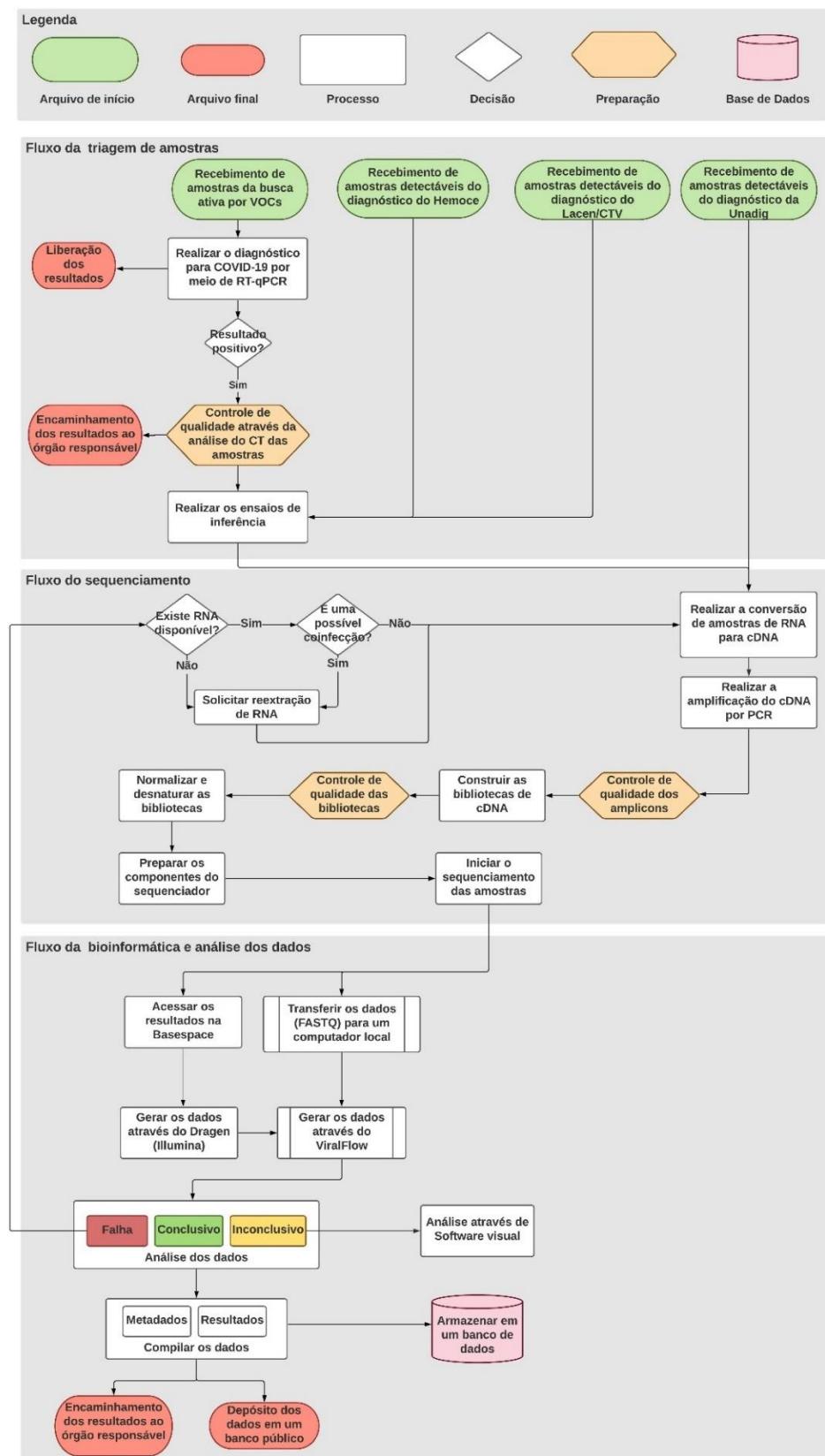
## 5 RESULTADOS E DISCUSSÃO

Nesta sessão são apresentados e discutidos os principais resultados obtidos pela coleta de dados nas etapas quantitativa e qualitativa da pesquisa, buscando uma melhor validação do fluxo e aprimoramentos propostos.

### 5.1 Elaboração e otimização do fluxo operacional

A Figura 13 mostra o fluxograma analisador do processo operacional, abordando todas as etapas logísticas e técnicas do serviço, desde a coleta e triagem das amostras até o sequenciamento e análise dos dados.

Figura 13 - Fluxo dos dados elaborado para o Serviço de Vigilância Genômica da Fiocruz Ceará.



Fonte: Elaborado pelo autor.

### **5.1.1 Fluxo da triagem de amostras**

O primeiro fluxo se inicia com o recebimento das amostras biológicas obtidas do aspirado nasofaríngeo e *swab* nasofaríngeo de quatro fontes distintas:

1. As amostras oriundas de hospitais e postos de saúde de vários municípios do estado do Ceará na busca ativa por VOCs;
2. As amostras detectáveis do diagnóstico do Hemoce, vindas de profissionais da saúde e pacientes infectados;
3. As amostras detectáveis do diagnóstico do Lacen e do Centro de Testagem ao Viajante, através de testes realizados no Aeroporto Internacional Pinto Martins em Fortaleza;
4. As amostras detectáveis do diagnóstico realizado pela Unadig na Fiocruz Ceará.

As amostras da Vigilância Genômica seguem para um diagnóstico preliminar através de RT-qPCR, realizada em um termociclador QuantStudio® (Thermo Fisher), sendo possível predizer o valor de Ct obtido pelo diagnóstico para cada amostra, através de um controle de qualidade. Os dados são liberados de forma online no GAL e o encaminhamento dos resultados é feito à Secretaria de Saúde do Ceará.

São realizados ensaios de inferência para as amostras triadas, resultando em um dado preliminar sobre a possibilidade de o paciente estar contaminado ou não com uma VOC. Os ensaios de inferência realizados e testados estão descritos posteriormente neste trabalho. As amostras da Unadig são encaminhadas através de ensaios de inferência já realizados pelo setor.

### **5.1.2 Fluxo do sequenciamento**

O segundo fluxo se inicia após a triagem completa das amostras e encaminhamento para a Fiocruz. A primeira etapa descrita é a conversão do RNA viral em cDNA, seguida pela amplificação através de PCR, realizada em um termociclador QuantStudio® (Thermo Fisher). Após essa etapa, é realizado o primeiro controle de qualidade do sequenciamento, através de eletroforese para a validação dos amplicons e a possibilidade de seguir com o fluxo. A próxima etapa consiste na construção das bibliotecas de cDNA para cada amostra e a validação das bibliotecas através de um segundo controle de qualidade. Ao final, as amostras são então encaminhadas para um sequenciador Miseq da Illumina e o sequenciamento é iniciado.

### **5.1.3 Fluxo da bioinformática**

O terceiro e último fluxo se inicia após o término do sequenciamento e envio dos dados para um computador local. Os arquivos gerados são analisados através da *Pipeline* ViralFlow e confirmados através da plataforma ultrarrápida DRAGEN™ Bio-IT da Illumina. Os dados para cada amostra gerados pelo controle de qualidade estão descritos como 3 possíveis resultados: conclusivo, inconclusivo e falha. Os resultados conclusivos seguem no fluxo, enquanto os resultados inconclusivos são analisados manualmente através de algum software visual, como o Geneious prime. A análise manual busca detectar possíveis erros de montagem ou de alinhamento dos iniciadores. Caso o resultado seja uma falha, a amostra é enviada novamente para o fluxo do sequenciamento, no qual é solicitada uma reextração da amostra caso não haja RNA suficiente, ou seja, um possível caso de coinfecção. Posteriormente, os resultados são compilados em planilhas e armazenados em um banco de dados local da Fiocruz. Por fim, os dados são liberados novamente no GAL, encaminhados à Secretaria de Saúde do Ceará e disponibilizados no banco de dados EpiCoV™ hospedado no GISAID.

## **5.2 Avaliação do quantitativo e da abrangência de amostras sequenciadas pelo Serviço de Vigilância Genômica da Fiocruz Ceará**

No período de janeiro a dezembro de 2021, o serviço contemplou 3930 amostras de 167 municípios do Ceará, divididos entre as sete mesorregiões do estado. O quantitativo de amostras sequenciadas foi organizado de acordo com os dados de cada região (Tabela 1). Os municípios não contemplados pelo serviço foram: Deputado Irapuan Pinheiro, Potengi, Altaneira, Aurora, Nova Olinda, Pena Forte, Acaraú, Barroquinha, Chaval, Croatá, Ubajara, Flecheirinha, Groaíras, Miraíma, Baixio, Ipaumirim e Umari.

Tabela 1 - Divisão do número de amostras sequenciadas por municípios contemplados das Mesorregiões Cearenses.

<b>Mesorregiões Cearenses</b>	<b>Número de amostras</b>	<b>Municípios contemplados</b>	<b>Total de municípios</b>
Metropolitana de Fortaleza	1.545	11	11
Sertões Cearenses	960	29	30
Norte Cearense	460	35	36
Noroeste Cearense	347	40	47
Sul Cearense	300	20	25
Jaguaribe	174	21	21
Centro-sul Cearense	144	11	14
<b>Total</b>	<b>3.930</b>	<b>167</b>	<b>184</b>

Fonte: Elaborado pelo autor.

O maior número de amostras sequenciadas se concentrou na Região Metropolitana de Fortaleza ( $n = 1545$ ), com 39,3% do total de amostras, seguido pela Região dos Sertões Cearenses ( $n = 960$ ), com 24,4%. A Região do Centro-sul Cearense teve o menor número de amostras ( $n = 144$ ), com apenas 3,7% do total. O serviço sequenciou amostras de 90,7% dos municípios do Ceará e todas as suas mesorregiões, o que permitiu uma boa representatividade para comparação com dados de outros estados do país.

O mapa coroplético da Figura 14 mostra a geolocalização de cada um dos municípios contemplados, divididos por cores entre as sete mesorregiões do Ceará. Os municípios que não tiveram amostras sequenciadas permaneceram em branco na figura. O mapa coroplético foi escolhido devido seu componente visual evidente, permitindo mostrar os dados objetivos dos municípios do território cearense e sua localização através de tonalidades de cores. Nesse tipo de mapa, considera-se que o valor da variável representativa do fenômeno, no caso as amostras sequenciadas, está uniformemente distribuída nos limites de cada área, assumindo assim que o fenômeno coincide com os limites geográficos, como explica SLOCUM et al., 2009.

Figura 14 - Mapa coroplético dos municípios contemplados por Mesorregiões do Ceará. Em branco: municípios que não tiveram amostras sequenciadas.



Fonte: Elaborado pelo autor.

### 5.3 Avaliação da prevalência de variantes de SARS-CoV-2 no estado do Ceará no ano de 2021, em comparação com outros estados brasileiros

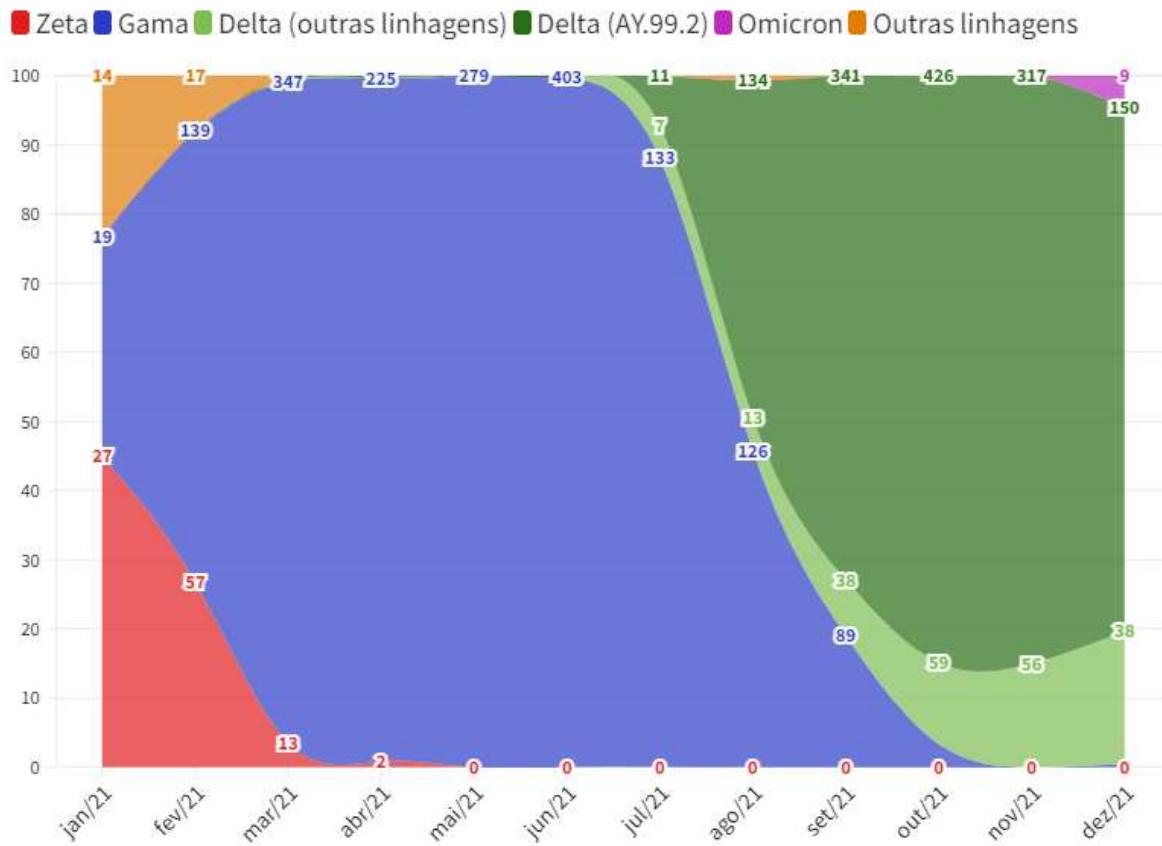
Os dados do quantitativo das VOCs sequenciadas ao longo dos meses de janeiro a dezembro de 2021 estão descritos na Tabela 2 abaixo. As principais variantes sequenciadas nesse período foram a Zeta, Gama, Delta e Ômicron. Outras linhagens presentes, porém, não em números representativos, também foram reunidas e catalogadas. O representativo visual para esses dados foi organizado através da construção de um gráfico de fluxo com o percentual de variantes a cada mês de 2021, como ilustrado na Figura 15.

Tabela 2 - Quantitativo das principais variantes sequenciadas no Ceará ao longo dos meses de 2021.

<b>Período (2021)</b>	<b>Zeta</b>	<b>Gama</b>	<b>Delta</b>	<b>Ômicron</b>	<b>Outras linhagens</b>	<b>Total</b>
<b>Janeiro</b>	27	19	0	0	14	60
<b>Fevereiro</b>	57	139	0	0	17	213
<b>Março</b>	13	347	1	0	2	363
<b>Abril</b>	2	225	1	0	0	228
<b>Maio</b>	0	279	0	0	0	279
<b>Junho</b>	0	403	1	0	0	404
<b>Julho</b>	0	133	18	0	0	151
<b>Agosto</b>	0	126	147	0	2	275
<b>Setembro</b>	0	89	379	0	0	468
<b>Outubro</b>	0	17	485	0	0	502
<b>Novembro</b>	0	0	373	0	0	373
<b>Dezembro</b>	0	0	188	9	0	197
<b>Total</b>	99	1.777	1.593	9	35	3.513

Fonte: Elaborado pelo autor.

Figura 15 -Gráfico de fluxo do quantitativo de variantes de preocupação no Ceará em 2021.



Fonte: Elaborado pelo autor.

A partir da análise temporal realizada, observou-se no mês de janeiro a predominância da variante Zeta, com cerca de 45% das amostras analisadas ( $n = 27$ ), seguida da VOC Gama (31,67%). Além disso, outras linhagens somaram 23,3% do quantitativo total no mês de janeiro de 2021. Dentre as outras linhagens observadas, se destaca a B.1.28 ( $n = 5$ ), que evoluiu dando origem à VOC Gama, segundo NAVECA et al., 2021. Outra linhagem presente é a B.1.33 ( $n = 4$ ), uma linhagem africana detectada no Quênia em 2020, de acordo com GITHINJI et al., 2021. Nos meses seguintes já é nítida a abrangência da VOC Gama no total de amostras, o que coincide com o aumento expressivo do número de casos positivos para COVID-19 no início do mês de março, segundo o Boletim Epidemiológico do COVID-19 do portal Integra SUS da SESA.

No Ceará, a linhagem P.1 demorou mais tempo para ser confirmada e se disseminar pela população, quando comparamos com outras regiões do país, o que resultou em um atraso no aumento do número de casos, o que confirma os dados compilados de SOUZA et al., 2021, que mostram a direta relação entre a linhagem P.1 e o crescimento do número de casos confirmados na região do Amazonas em novembro de 2020. Outros estudos também relatam a associação da

linhagem P.1 e o rápido crescimento dos casos e internações, como exemplificado no trabalho realizado por MARTINS et al., 2021 no Rio Grande do Sul. Os dados apresentados nesse trabalho ressaltam essa relação e confirmam os dados de outros autores. Isso mostra a importância da vigilância genômica em relatar o aumento dos números de determinada linhagem e sua ligação com eventuais aumentos do número de casos.

Entre março e julho de 2021, a variante Delta começou a circular no Ceará e, em agosto, se tornou predominantes dentre as amostras sequenciadas. Comparando com os dados da Rede Genômica da Fiocruz para a região do Nordeste, as mesmas variantes começaram a se tornar predominantes no mesmo período. A linhagem AY.99.2 se mostrou preponderante no Nordeste já no mês de agosto ( $n = 682$ ), com aproximadamente 88% das variantes Delta sequenciadas. Esse padrão se mantém para as amostras sequenciadas no Ceará, com a linhagem AY.99.2 compondo 85,3% das amostras da variante Delta sequenciadas no ano de 2021.

#### **5.4 Listagem dos principais desafios e limitações para o serviço, assim como possíveis soluções para os problemas enfrentados**

Os substanciais desafios relatados foram divididos entre desafios técnicos e logísticos. As possíveis soluções pensadas são apresentadas juntamente na Tabela 3.

Tabela 3 - Lista dos principais problemas técnicos e logísticos observados e possíveis soluções encontradas.

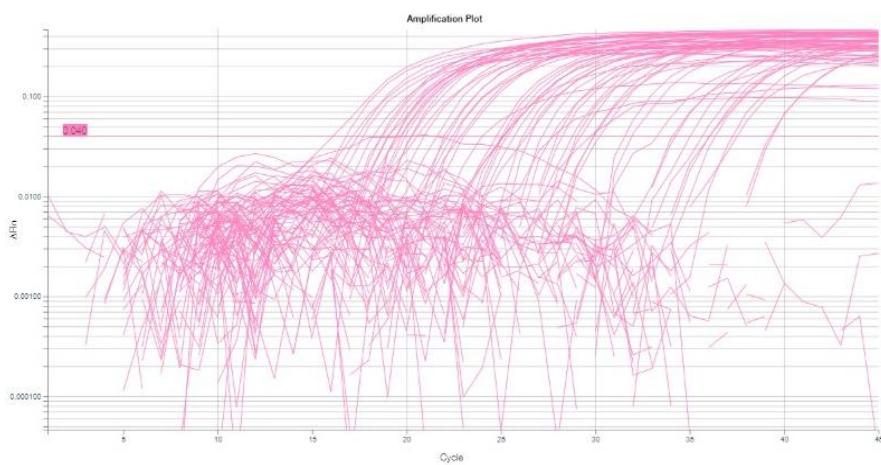
<b>Problemas técnicos</b>	<b>Possíveis soluções</b>
O tempo de duração de um sequenciamento é longo e complexo, sendo necessário alguns dias para a confirmação da linhagem da amostra e fornecimento de uma resposta correta para a Secretaria de Saúde;	Utilização de ensaios de inferência para predição preliminar de linhagem e envio de uma resposta prévia para a Secretaria de Saúde, devido seu menor tempo de execução e custo envolvido. Sendo posteriormente confirmado através de sequenciamento;
O alto valor dos reagentes e equipamentos se torna problemático para a quantidade de amostras a serem sequenciadas e o número de repetições necessárias;	Estimular o empreendedorismo em Biotecnologia no estado por meio da união entre governo, academia e setor privado. <i>Spin-offs</i> e <i>startups</i> em biotecnologia precisam ser estimulados, assim como parcerias entre academia e empresas para o desenvolvimento de processos e produtos para o setor no Estado.
<b>Problemas logísticos</b>	<b>Possíveis soluções</b>
O fluxo é grande e complexo, passando por várias pessoas e lugares, o que implica em dificuldades de organização e manuseio de amostras, além de falhas de comunicação frequentes;	Utilização do desenho de fluxograma desenvolvido como forma de comunicação entre diferentes setores e organização dos dados sobre as amostras. Além disso, pode ser utilizado para explicar de maneira prática todas as etapas envolvidas;
Tanto o processo como as técnicas utilizados são de difícil compreensão para novos membros e visitantes, sendo complicado explicar de forma clara e objetiva o serviço desenvolvido.	Estabelecer parcerias com o Bacharelado em Biotecnologia da UFC e cursos técnicos para treinamento de pessoal.

Fonte: Elaborado pelo autor.

## 5.5 Levantamento do percentual de ensaios de inferência confirmados através de sequenciamento

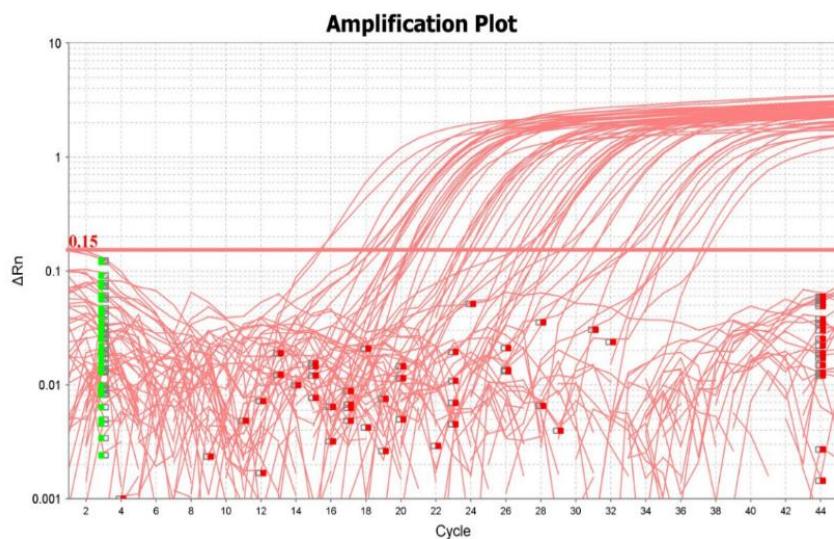
As Figuras 16 e 17 mostram as curvas de amplificação por RT-qPCR dos ensaios de inferência realizados baseado na inserção de quatro nucleotídeos no início da região do Gene N (28263 bp) e na deleção de nove nucleotídeos na região ORF1ab (11288 pb), respectivamente, característicos da VOC Gama. A Figura 18 mostra o ensaio de genotipagem para as amostras, retratando a discriminação dos alelos da mutação K417T.

Figura 16 - Curvas de amplificação de amostras analisadas por inferência baseados na amplificação de uma região de inserção de quatro nucleotídeos característico da VOC Gama (P.1).



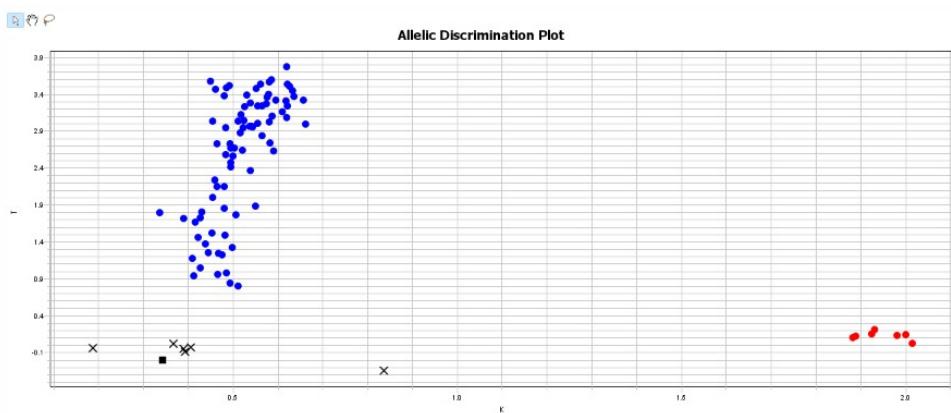
Fonte: Imagem do autor.

Figura 17 - Curvas de amplificação de amostras analisadas por inferência baseados na amplificação de uma região de deleção de nove nucleotídeos característico da VOC Gama (P.1).



Fonte: Imagem do autor.

Figura 18 - Ensaio de genotipagem medindo a intensidade da fluorescência da mutação K417T, baseado na detecção de alelos característicos da VOC Gama (P.1). Em vermelho: Homozigotos K/K (não mutantes); Em Azul: Homozigotos T/T (mutantes); Em preto: indeterminados.



Fonte: Imagem do autor.

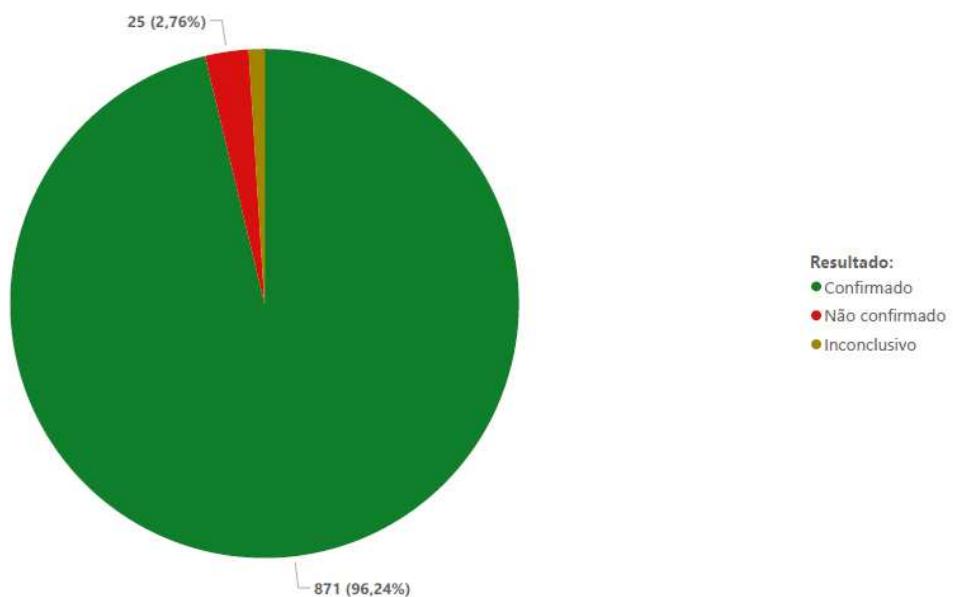
A Tabela 4 a seguir mostra o comparativo dos resultados de inferência para 20 lotes de sequenciamento analisados, com o percentual total dos dados para os 3 possíveis resultados (confirmado, não confirmado e inconclusivo) plotados em um diagrama circular (Figura 19).

Tabela 4 - Comparativo do resultado dos ensaios de inferência por lote de sequenciamento.

<b>Lotes</b>	<b>Número de testes</b>	<b>Número de testes</b>	<b>Número de testes</b>	<b>Número total de</b>
	<b>confirmados</b>	<b>não confirmados</b>	<b>inconclusivos</b>	<b>testes</b>
<b>01</b>	17	0	0	17
<b>02</b>	24	1	0	25
<b>03</b>	20	0	0	20
<b>04</b>	122	4	0	126
<b>05</b>	213	6	0	219
<b>06</b>	1	0	1	2
<b>07</b>	0	0	0	0
<b>08</b>	143	6	1	150
<b>09</b>	4	0	0	4
<b>10</b>	20	0	0	20
<b>11</b>	77	1	5	83
<b>12</b>	67	2	0	69
<b>13</b>	34	0	0	34
<b>14</b>	35	0	0	35
<b>15</b>	29	2	1	32
<b>16</b>	21	2	1	24
<b>17</b>	1	0	0	1
<b>18</b>	0	0	0	0
<b>19</b>	41	1	0	42
<b>20</b>	2	0	0	2
<b>Total</b>	<b>871</b>	<b>25</b>	<b>9</b>	<b>905</b>

Fonte: Elaborado pelo autor.

Figura 19 - Diagrama circular do percentual do resultado da validação do ensaio de inferência comparado com os dados de sequenciamento.



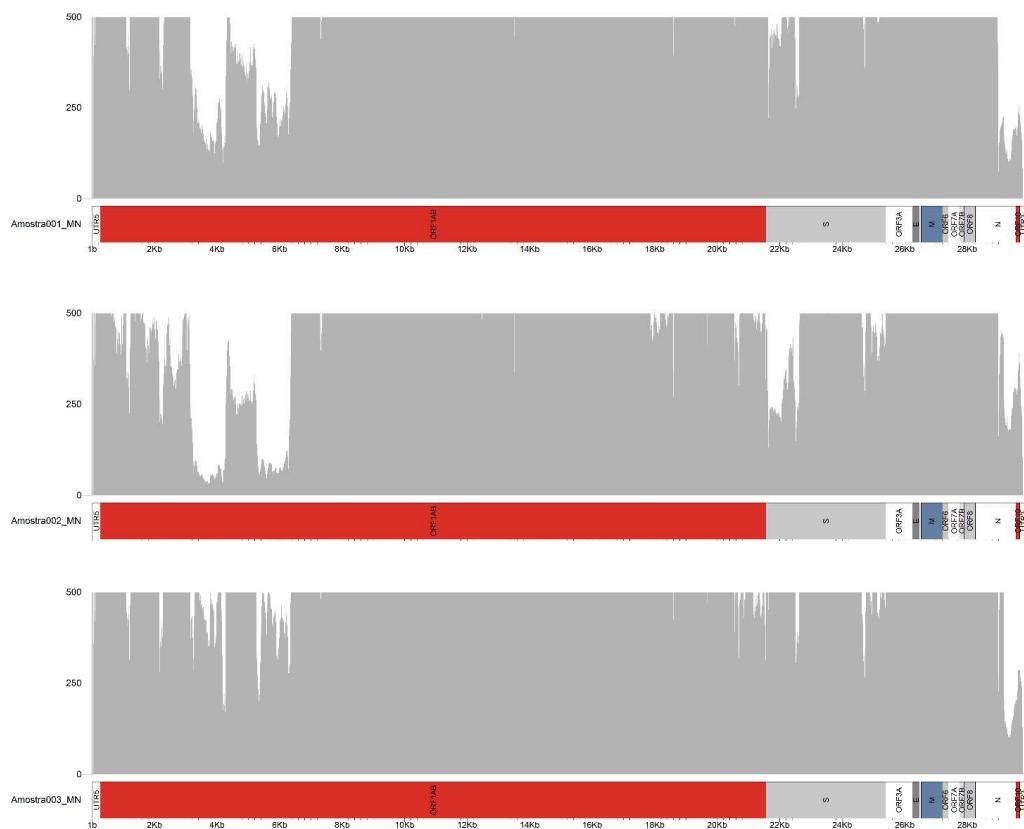
Fonte: Elaborado pelo autor.

O número de testes confirmados por sequenciamento representou 96,24% do total de testes analisados, com apenas 2,76% de discordância e menos de 1% de testes inconclusivos. Esses dados mostram a capacidade dos ensaios em detectar as mutações características da linhagem P.1. A utilização complementar do painel de genotipagem por RT-qPCR com estratégias de sequenciamento do genoma completo se mostrou promissora para o rastreamento mais abrangente e em tempo real das variantes de interesse do SARS-CoV-2 (HASAN et al., 2021). Esses ensaios apresentaram capacidade de serem utilizados como prognóstico preliminar de linhagens, a serem confirmados através de sequenciamento. Além disso, a abordagem baseada na reação em cadeia da PCR permite a rápida identificação de linhagens de SARS-CoV-2 em menos de 4 horas, através da presença ou ausência do painel de polimorfismo definido (deleção 9, inserção 4 e mutação K417T). Esses dados são condizentes com os achados de HALE et al., 2021, que expande o potencial de ensaios de inferência para a detecção de outras linhagens.

## 5.6 Comparação e otimização de diferentes protocolos de preparo de amostras para sequenciamento

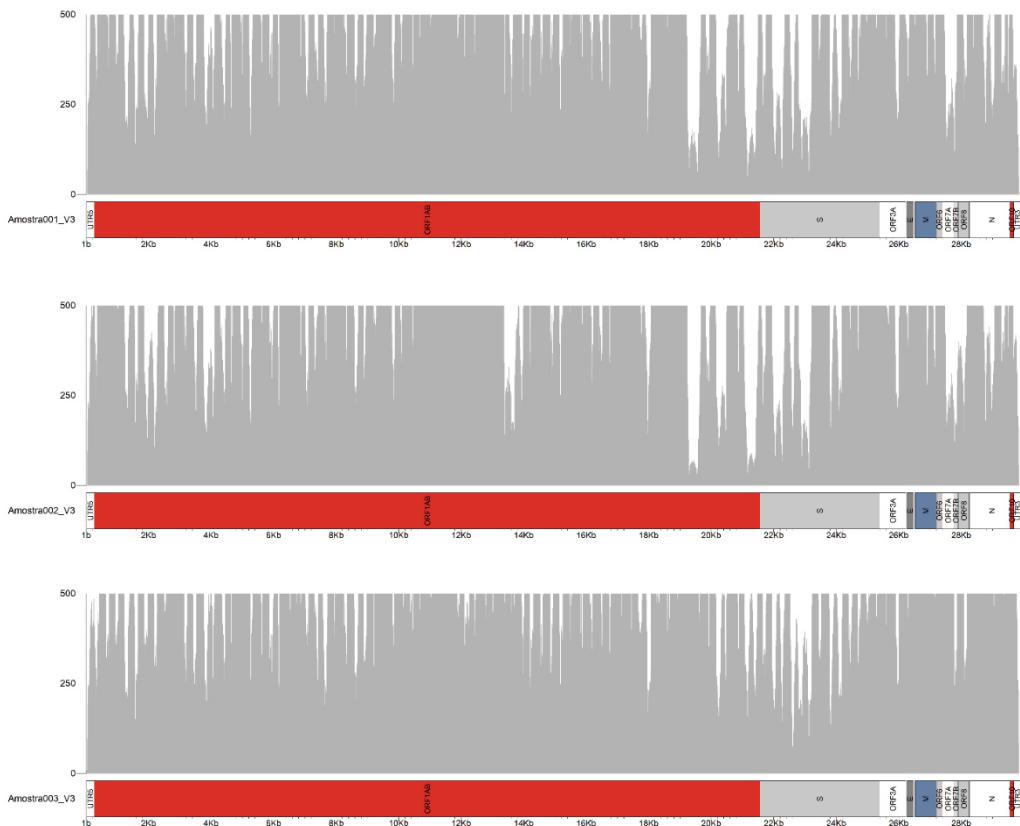
Os mapas de profundidade gerados no primeiro teste para amostras da variante Gama, utilizando os iniciadores do protocolo Midnight™ e ArticV3™, estão sendo mostrados nas Figuras 20 e 21 abaixo.

Figura 20 - Mapas de profundidade de amostras classificadas como variante Gama (P.1) sequenciadas utilizando os iniciadores Midnight™.



Fonte: Elaborado pelo autor.

Figura 21 - Mapas de profundidade de amostras classificadas como variante Gama (P.1) sequenciadas utilizando os iniciadores ArticV3™.



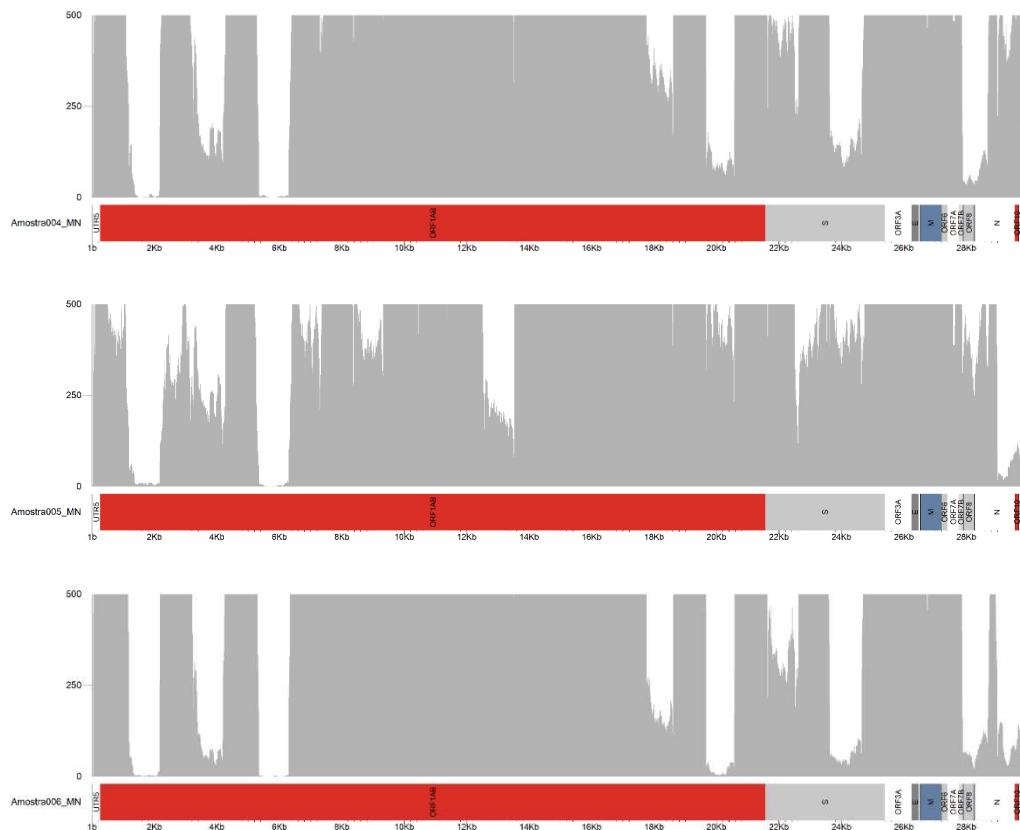
Fonte: Elaborado pelo autor.

Ao analisarmos a Figura 20, pode-se perceber que o protocolo utilizando o primeiro conjunto de iniciadores apresenta um caráter muito mais constante na média de profundidade para cada base ao longo do genoma, com “quedas” ocorrendo sempre em regiões específicas, mais precisamente nos intervalos de 4 kb, 6 kb, 22 kb, 25 kb e acima de 28 kb.

Quando se compara com os mapas de profundidade gerados na Figura 21, nota-se um padrão completamente diferente. Os mapas apresentam uma média de profundidade variando bastante, possuindo maiores dispersões ao longo do genoma. Existem regiões com uma densidade maior de “quedas” no final do gene da ORF1ab e gene S, entre 19 e 25 kb e em regiões mais terminais do genoma, acima de 26 kb.

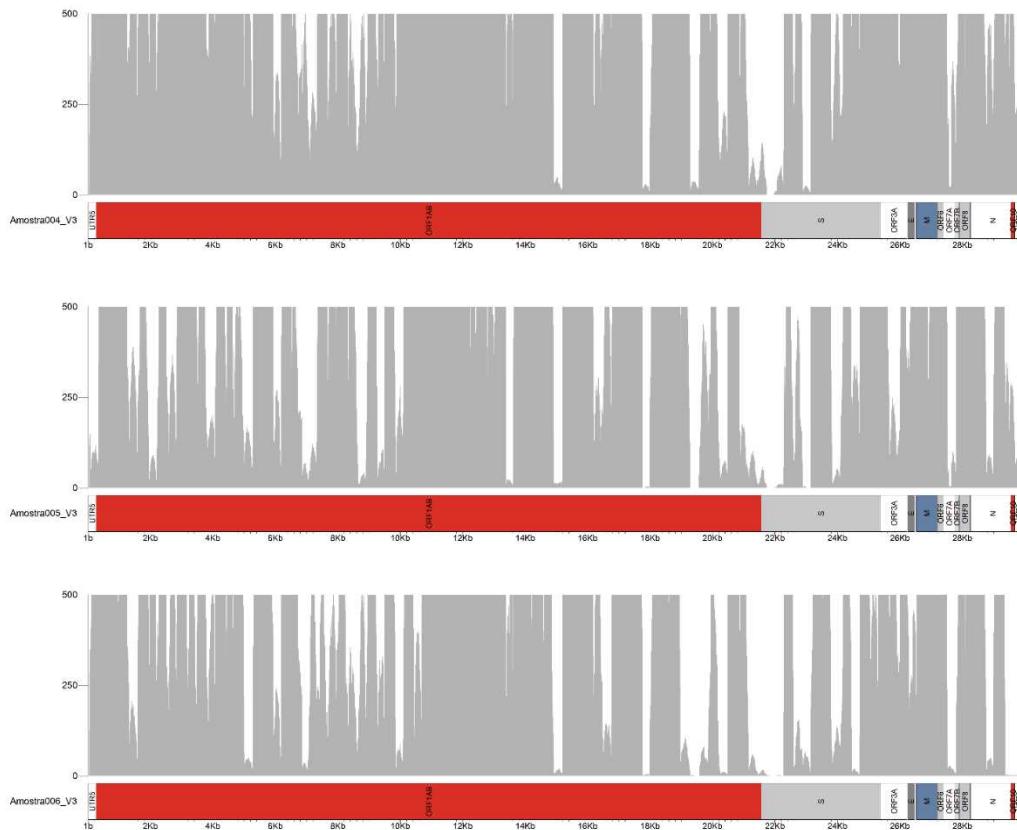
Os mapas de profundidade gerados no segundo teste para amostras da variante Delta utilizando os mesmos conjuntos de iniciadores estão sendo mostrados nas Figuras 22 e 23 abaixo.

Figura 22 - Mapas de profundidade de amostras classificadas como variante Delta (B.1.617.2) sequenciadas utilizando os iniciadores Midnight™.



Fonte: Elaborado pelo autor.

Figura 23 - Mapas de profundidade de amostras classificadas como variante Delta (B.1.617.2) sequenciadas utilizando os iniciadores ArticV3™.



Fonte: Elaborado pelo autor.

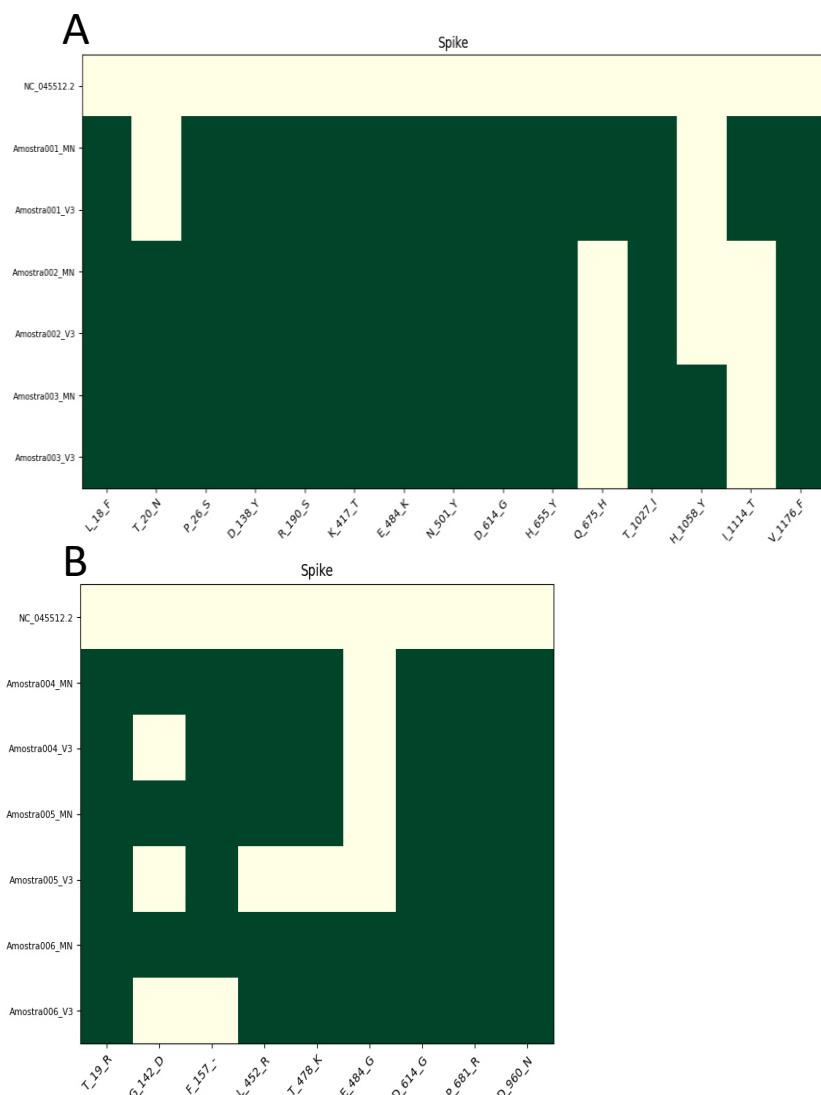
Nota-se um padrão diferente quando se utiliza o mesmo protocolo para preparo de amostras classificadas como Delta. Na Figura 22, foram visualizadas “quedas” na média de profundidade nas regiões de 2 kb, 4 kb, 6 kb, 18 kb, 20 kb, 22 kb, 24 kb e acima de 28 kb. Existem também regiões de baixa profundidade presentes apenas na amostra 005 em 7 kb, 9 kb e 13 kb. Esse padrão é diferente do visto anteriormente para as amostras Gama, possuindo novas regiões com baixa profundidade. O padrão semelhante (4 kb, 6kb e acima de 28 kb) se mostrou muito mais acentuado nas amostras Delta, com médias de profundidade chegando a quase zero nessas regiões.

A comparação dos mapas de profundidade gerados utilizando os iniciadores ArticV3™ na Figura 23, também se observa novamente um padrão diferente. As amostras continuam com

regiões de alta densidade de “quedas” ao final do gene ORF1ab e gene S, porém muito mais intensos do que nas amostras Gama. A figura também mostra regiões com média de profundidade próximas a zero em várias outras regiões do genoma, mostrando uma limitação de qualidade do protocolo para o sequenciamento de amostras Delta.

A seguir, a Figura 24 também apresenta os mapas das mutações de aminoácidos na região da proteína *Spike* (gene S) detectadas por ambos os protocolos nos 2 testes realizados.

Figura 24 - Mapa das mutações na região da proteína Spike (gene S), utilizando os iniciadores Midnight™ e ArticV3™, alinhadas com uma amostra referência de número de acesso NC\_045512.2; a) Amostras classificadas como Gama (P.1); b) Amostras classificadas como Delta (B.1.617.2); em branco) ausência da mutação; em verde) presença da mutação.



Fonte: Elaborado pelo autor.

Ambos os protocolos conseguiram detectar as mesmas mutações para as primeiras três amostras, tendo uma pequena variação entre elas. Segundo Naveca et al., 2021, a variante Gama possui 10 mutações definidoras de linhagem na região do gene S (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y e T1027I), estando presentes nas três amostras, com exceção da mutação T20N que não foi detectada na amostra 001 por nenhum dos protocolos. A substituição D614G também se encontra presente em ambos os sequenciamentos, sendo uma mutação prevalente desde a primeira evidência de alteração no material genético do SARS-CoV-2 em abril de 2020 (YANIV et al., 2021). Segundo os dados de KORBER et al., 2020, essa mudança ocorreu mesmo em epidemias locais em países mais isolados, realçando a consistência desse padrão como alteração significativa para a evolução do vírus. Além destas, a substituição V1176F foi detectada para as três amostras, as mutações Q675H e I1114T apenas na amostra 001 e H1058Y apenas na amostra 003. A substituição V1176F também é uma assinatura para a Gama, porém que contrasta com a versão ancestral da variante (linhagem B.1.1.28) (NAVECA et al., 2021), o que justifica sua presença nas amostras testadas. As demais substituições (Q675H, I1114T e H1058Y) aparecem ser mutações privadas, não comuns para sublinhagens da Gama.

Para as últimas três amostras, houve diferenças na detecção das mutações quando compararmos os dois protocolos. A variante Delta possui seis mutações presentes no seu gene S (T19R, L452R, T478K, D614G, P681R e D960N), com uma deleção na posição 157/158 (SHIEHZADEGAN et al., 2021). Destas, a mutação T19R, D614G, P681R e D960N foram detectadas em ambos os protocolos. As mutações L452R e T478K não foram identificadas na amostra 005 sequenciada com o protocolo ArticV3™, porém puderam ser identificadas pelo protocolo Midnight™. A deleção na posição 157/158 não pôde ser identificada na amostra 006 com o protocolo ArticV3™, porém foi identificada pelo protocolo Midnight™. Além disso, a mutação G142D também só foi identificada pelo protocolo Midnight™. A mutação E484G foi identificada apenas na amostra 006, o que indica ser uma mutação privada. A baixa média de profundidade presente na região de 22 kb nos mapas da Figura 23, se mostraram como uma falha do protocolo em sequenciar mutações nessa região.

De acordo com esses dados, observou-se uma limitação dos iniciadores ArticV3™ em identificar mutações da variante Delta. Os mapas se mostraram úteis em sinalizar essas falhas, podendo ser utilizados para outras linhagens e em pesquisas futuras.

## 6 PERSPECTIVAS

O impacto ocasionado pela emergência do SARS-CoV-2 é imenso. O novo coronavírus é um vírus sem precedentes na nossa história recente, e mesmo após quase 2 anos de pandemia ainda tem afetado uma grande parcela da população mundial, alterando de maneira significativa os protocolos e a forma como é monitorado doenças respiratórias e sua evolução. As variantes de preocupação do SARS-CoV-2 têm apresentado muitas características similares, mas cada uma também possui atributos ímpares e exclusivos, impactando de forma heterogênea, desigual e muitas vezes imprevisível as populações afetadas, até mesmos aquelas de uma mesma localidade ou região. Assim, estratégias em forma de redes de vigilância genômica decentralizadas, mas com foco regionalizado, como o estabelecido pela Rede Genômica da Fiocruz Ceará, são importantes modelos para atendimento de demandas locais e melhor entendimento da situação sanitária e epidemiológica. Ao mesmo tempo isto facilita a integração com outros centros, propiciando uma compreensão abrangente da realidade, por exemplo da dinâmica de introdução, transmissão, espalhamento e persistência de variantes dentro de um contexto geográfico, a exemplo de um país e além de suas fronteiras. Tais iniciativas, porém, requerem o contínuo aprimoramento de ferramentas e soluções analíticas para o aperfeiçoamento e efetividade dos serviços, bem como investimentos em materiais e recursos humano visando sua manutenção.

No presente estudo, é destacada a importância do planejamento e concepção de novas soluções para otimização dos fluxos de trabalho, particularmente quanto a triagem qualitativa de amostras representativas, inclusive com o uso de ensaios moleculares complementares, como de inferência por rt-qPCR, aprimoramento de protocolos e atualização de painéis de sequenciamento para maior robustez na obtenção dos dados e menor índice de falhas, maior objetividade e rapidez na classificação de linhagens e efetividade no reportamento dos resultados e notificações. Por exemplo, as análises realizadas neste trabalho sugerem que protocolos originais de sequenciamento podem apresentar severas limitações de aplicabilidade, dependendo do contexto epidemiológico geográfico e principalmente temporal, pois a dinâmica da evolução de VOCs tem sido implacável, significativamente alterando os cenários epidemiológicos ao longo de poucas semanas ou meses. Em função disso, mesmo para protocolos já referendados, novas implementações têm sido preconizadas, introduzindo-se aprimoramentos através do redesenho de novos iniciadores, substituição de alguns e customização de outros que possam suprir falhas, e ultimamente incrementar-se a profundidade

de leitura, o nível de cobertura genômico, e maior confiança da atribuição de uma mutação ou conjunta de assinaturas moleculares.

## 7 CONCLUSÃO

A contínua evolução do SARS-CoV2, com surgimento e expansão de novas sublinhagens das VOCs Delta e Ômicron têm alertado a OMS e os serviços de saúde e vigilância do mundo todo para a contínua manutenção de um modelo de monitoramento ativo da população viral e sua evolução. Os dados levantados por este trabalho reforçam a importância estratégica e o impacto da introdução do modelo de vigilância genômica estudado no estado do Ceará. Desta forma, recomenda-se a ampliação do nível de testagem, particularmente dos testes moleculares por rt-qPCR para se minimizar o impacto de subnotificações e otimizar-se a seleção e representatividade amostral, ao mesmo tempo estender-se a oferta de exames de sequenciamento para se incrementar a efetividade da vigilância genômica em identificar-se novos eventos de interesse e de impacto epidemiológico. Importante ressaltar a necessidade de contínuo aprimoramento dos fluxos de trabalho para triagem, controle de qualidade e integração de dados, visando maior celeridade e efetividade no retorno de relatórios técnicos, notificação de casos e compartilhamento de sequências genômicas para maior efetividade dos sistemas de vigilância do Brasil e do mundo.

## 8 REFERÊNCIAS

- Aksamentov, I., Roemer, C., Hodcroft, E., & Neher, R. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, 6(67), 3773. <https://doi.org/10.21105/joss.03773>.
- Brito, A. F., Semenova, E., Dudas, G., Hassler, G. W., Kalinich, C. C., Kraemer, M. U. G., ... Faria, N. R. (2021). Global disparities in SARS-CoV-2 genomic surveillance. *MedRxiv : The Preprint Server for Health Sciences*, 1–24. <https://doi.org/10.1101/2021.08.21.21262393>.
- BHAGAT, S. et al. (2021). Novel coronavirus (COVID-19) pandemic: current status and possible strategies for detection and treatment of the disease. Expert review of anti-infective therapy.
- Burki, T. K. (2021). Omicron variant and booster COVID-19 vaccines. *The Lancet Respiratory Medicine*, 2600(21), 2600. [https://doi.org/10.1016/s2213-2600\(21\)00559-2](https://doi.org/10.1016/s2213-2600(21)00559-2).
- CASCELLA, M. et al. (2021). Features, evaluation, and treatment of coronavirus (COVID19).
- Chavez, S., Long, B., Koyfman, A., & Liang, S. Y. (2021). Coronavirus Disease (COVID-19): A primer for emergency physicians. *American Journal of Emergency Medicine*, 44, 220–229. <https://doi.org/10.1016/j.ajem.2020.03.036>.
- Chiara, M., D'Erchia, A. M., Gissi, C., Manzari, C., Parisi, A., Resta, N., ... Pesole, G. (2021). Next generation sequencing of SARS-CoV-2 genomes: Challenges, applications and opportunities. *Briefings in Bioinformatics*, 22(2), 616–630. <https://doi.org/10.1093/bib/bbaa297>.
- Costa, Z. G. A., Romano, A. P. M., Elkhoury, A. N. M., & Flannery, B. (2011). Evolução histórica da vigilância epidemiológica e do controle da febre amarela no Brasil. *Revista Pan-Amazônica de Saúde*, 2(1), 11–26. <https://doi.org/10.5123/s2176-62232011000100002>.
- de Souza, F. S. H., Hojo-Souza, N. S., da Silva, C. M., & Guidoni, D. L. (2021). Second wave of COVID-19 in Brazil: younger at higher risk. *European Journal of Epidemiology*, 36(4), 441–443. <https://doi.org/10.1007/s10654-021-00750-8>.

Dezordi, Z., Filipe (2021), NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice. 1. *MedRxiv*, (165), 1–13.

Duchene, S. et al. (2020) ‘Temporal Signal and the Phylodynamic Threshold of SARS-CoV-2’, *Virus Evolution*, 6: veaa061.

Freed, N., Silander, O., & Freed, N. (2021). *SARS-CoV2 genome sequencing protocol (1200bp amplicon “midnight” primer set, using Nanopore Rapid kit)*. 2, 1–14. Retrieved from <https://www.protocols.io/view/sars-cov2-genome-sequencing-protocol-1200bp-amplic-btsrnnd6>.

Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S. G., Park, D. J., Kanneh, L., ... Lander, E. S. (2014). During the 2014 Outbreak. *Science*, 345(6202), 1369–1372.

Githinji, G., de Laurent, Z. R., Mohammed, K. S., Omuoyo, D. O., Macharia, P. M., Morobe, J. M., ... Agoti, C. N. (2021). Tracking the introduction and spread of SARS-CoV-2 in coastal Kenya. *Nature Communications*, 12(1), 1–10. <https://doi.org/10.1038/s41467-021-25137-x>.

Grada, A., & Weinbrecht, K. (2013). Next-generation sequencing: Methodology and application. *Journal of Investigative Dermatology*, 133(8), e11–4. <https://doi.org/10.1038/jid.2013.248>.

Hale, R., Crowley, P., Dervisevic, S., Coupland, L., Cliff, P. R., Ebie, S., ... Stanley, K. (2021). Development of a multiplex tandem pcr (Mt-pcr) assay for the detection of emerging sars-cov-2 variants. *Viruses*, 13(10). <https://doi.org/10.3390/v13102028>.

Harper, H., Burridge, A., Winfield, M., Finn, A., Davidson, A., Matthews, D., ... Barker, G. (2021). Detecting SARS-CoV-2 variants with SNP genotyping. *PLoS ONE*, 16(2 February 2021), 1–12. <https://doi.org/10.1371/journal.pone.0243185>.

Hasan, M. R., Kalikiri, M. K. R., Mirza, F., Sundararaju, S., Sharma, A., Xaba, T., ... Tang, P. (2021). Real-Time SARS-CoV-2 Genotyping by High-Throughput Multiplex PCR Reveals the Epidemiology of the Variants of Concern in Qatar. *International Journal of Infectious Diseases*, 112(June), 52–54. <https://doi.org/10.1016/j.ijid.2021.09.006>.

Hasoksuz, M., Kiliç, S., & Sarac, F. (2020). Coronaviruses and sars-cov-2. *Turkish Journal of Medical Sciences*, 50(SI-1), 549–556. <https://doi.org/10.3906/sag-2004-127>.

IBGE. Divisão do Brasil em mesorregiões e microrregiões geográficas. Rio de Janeiro. (1990). Slocum, T.A., McMaster R.B., Kessler, F.C e Howard, H.H. (2009). Thematic Cartography and

- Geovisualization. Prentice Hall.
- HUI, D. S. et al. (2020). The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in Wuhan, China. *Intl. Journal Infectology Disease*.
- Kandeel, M., Mohamed, M. E. M., Abd El-Lateef, H. M., Venugopala, K. N., & El-Beltagi, H. S. (2021). Omicron variant genome evolution and phylogenetics. *Journal of Medical Virology*, (December), 1–6. <https://doi.org/10.1002/jmv.27515>.
- Kannan, S. R., Spratt, A. N., Cohen, A. R., Naqvi, S. H., Chand, H. S., Quinn, T. P., ... Singh, K. (2021). Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. *Journal of Autoimmunity*, 124(July), 102715. <https://doi.org/10.1016/j.jaut.2021.102715>.
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalsterer, W., ... Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812-827.e19. <https://doi.org/10.1016/j.cell.2020.06.043>.
- Lauring, A. S., & Hodcroft, E. B. (2021). Genetic Variants of SARS-CoV-2 - What Do They Mean? *JAMA - Journal of the American Medical Association*, 325(6), 529–531. <https://doi.org/10.1001/jama.2020.27124>.
- Martins, A. F., Zavascki, A. P., Wink, P. L., Volpato, F. C. Z., Monteiro, F. L., Rosset, C., ... Barth, A. L. (2021). Detection of SARS-CoV-2 lineage P.1 in patients from a region with exponentially increasing hospitalisation rate, February 2021, Rio Grande do Sul, Southern Brazil. *Eurosurveillance*, 26(12), 0–4. <https://doi.org/10.2807/1560-7917.ES.2021.26.12.2100276>.
- Merhy, E.E. et al. (2007). O trabalho em Saúde: olhando e experienciando o SUS no cotidiano. 4<sup>a</sup> Ed. São Paulo: Hucitec.
- Mlcochova, P., Kemp, S., Dhar, M. S., Papa, G., Meng, B., Ferreira, I. A. T. M., ... Gupta, R. K. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, 599(November). <https://doi.org/10.1038/s41586-021-03944-y>.
- Naveca, F. G., Nascimento, V., de Souza, V. C., Corado, A. de L., Nascimento, F., Silva, G., ... Bello, G. (2021). COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nature Medicine*, 27(7), 1230–1238.

<https://doi.org/10.1038/s41591-021-01378-7>.

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., ... Rambaut, A. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, 7(2), 1–9. <https://doi.org/10.1093/ve/veab064>.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., ... Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5(11), 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>.

Sallas, J., Elidio, G. A., Rohlf, D. B., De Medeiros, A. C., & Guilhem, D. B. (2021). A vigilância genômica do SARS-CoV-2 no Brasil na resposta à pandemia da COVID-19. *Revista Panamericana de Salud Pública/Pan American Journal of Public Health*, 45, 1–2. <https://doi.org/10.26633/RPSP.2021.75>.

Shiehzadegan, S., Alaghemand, N., Fox, M., & Venketaraman, V. (2021). Analysis of the Delta Variant B.1.617.2 COVID-19. *Clinics and Practice*, 11(4), 778–784. <https://doi.org/10.3390/clinpract11040093>.

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13), 2–4. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.

Silva, J. D. O. da. (2020). *A Tempestade do Coronavírus*. 52.

SLACK et al. (2009). Administração da Produção.

Sluter, C. R., Iescheck, A. L., & Bravo, J. V. M. (2011). Mapas Coropléticos. *Cartografia Temática*, (Figura 1), 8. Retrieved from <http://www.atlasbrasil.org.br>.

TSANG, H. F. et al. An Update on COVID-19 Pandemic: The Epidemiology, Pathogenesis, Prevention and Treatment Strategies. Expert review of anti-infective therapy, 2021.

Walensky, R. P., Walke, H. T., & Fauci, A. S. (2021). SARS-CoV-2 Variants of Concern in the United States-Challenges and Opportunities. *JAMA - Journal of the American Medical Association*, 325(11), 1037–1038. <https://doi.org/10.1001/jama.2021.2294>.

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Dempsey, D. M., Dutilh, B. E., ... Davison, A. J. (2019). Changes to virus taxonomy and the International Code of Virus

Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). *Archives of Virology*, 164(9), 2417–2429. <https://doi.org/10.1007/s00705-019-04306-w>.

WU, J. T.; LEUNG, K.; LEUNG, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study.

Yaniv, K., Ozer, E., & Kushmaro, A. (2021). SARS-CoV-2 variants of concern, Gamma (P.1) and Delta (B.1.617), sensitive detection and quantification in wastewater employing direct RT-qPCR. *MedRxiv*, 2021.07.14.21260495. Retrieved from <http://medrxiv.org/content/early/2021/07/16/2021.07.14.21260495.abstract>.

## ANEXO A – INICIADORES DO PROTOCOLO MIDNIGHT™

Lista de iniciadores utilizados do protocolo Midnight™.

<b>Nome do primer</b>	<b>Sequência</b>	<b>Pool</b>	<b>Tm</b>	<b>Tamanho</b>
SARSCoV_1200_1_LEFT	ACCAACCAACTTCGATCTCTGT	1	60.69	24
SARSCoV_1200_1_RIGHT	GGTTGCATTCACTTGGTGACGC	1	61.49	22
SARSCoV_1200_3_LEFT	GGCTTGAAGAGAAGTTAACCGAAGGT	1	61.19	26
SARSCoV_1200_3_RIGHT	GATTGTCCCACTGCCGTCTG	1	61.5	22
SARSCoV_1200_5_LEFT	ACCTACTAAAAGGCTGGTGGC	1	60.55	22
SARSCoV_1200_5_RIGHT	AGCATCTTAGAGCAGGTGGA	1	61.16	22
SARSCoV_1200_7_LEFT	ACCTGGTGTATACGTTGCTTG	1	60.8	24
SARSCoV_1200_7_RIGHT	GCTGAAATCGGGGCCATTGTA	1	61.53	22
SARSCoV_1200_9_LEFT	AGAAGTTACTGGCGATAGTTGTAATAACT	1	60.59	29
SARSCoV_1200_9_RIGHT	TGCTGATATGTCCAAAGCACCA	1	60.29	22
SARSCoV_1200_11_LEFT	AGACACCTAAAGTATAAGTTGTCGCA	1	60.74	27
SARSCoV_1200_11_RIGHT	GCCCACATGGAAATGGCTTGAT	1	61.8	22
SARSCoV_1200_13_LEFT	ACCTCTAACACAGCAGCAAAC	1	61.55	23
SARSCoV_1200_13_RIGHT	CGTCCTTTCTTGGAAAGCGACA	1	61.38	22
SARSCoV_1200_15_LEFT	TTTTAAGGAATTACTTGTATGCTGCT	1	60.06	28
SARSCoV_1200_15_RIGHT	ACACACAAACAGCATCGTCAGAG	1	61.12	22
SARSCoV_1200_17_LEFT	TCAAGCTTTGCAGCAGAAACG	1	61.28	23
SARSCoV_1200_17_RIGHT	CCAAGCAGGGTTACGTGTAAGG	1	61.19	22
SARSCoV_1200_19_LEFT	GGCACATGGCTTGAGTTGACA	1	61.91	22
SARSCoV_1200_19_RIGHT	CCTGTTGCCATCAAAGTGTCCC	1	61.62	23
SARSCoV_1200_21_LEFT	TCTGTAGTTCTAAGGTTGCAAAGTGA	1	60.58	28
SARSCoV_1200_21_RIGHT	GCAGGGGTAATTGAGTTCTGG	1	60.95	22
SARSCoV_1200_23_LEFT	ACTTAGAGTCCAACCAACAGAAATCT	1	60.18	26
SARSCoV_1200_23_RIGHT	TGACTAGCTACACTACGTGCC	1	61.52	22
SARSCoV_1200_25_LEFT	TGCTGCTACTAAATGTCAGAGTGT	1	60.51	25
SARSCoV_1200_25_RIGHT	CATTCCAGCAAAGCCAAAGCC	1	61.45	22
SARSCoV_1200_27_LEFT	TGGATCACCGGTGGAATTGCTA	1	61.75	22
SARSCoV_1200_27_RIGHT	TGTCGTTAGGCCTGACAAGT	1	60.74	22
SARSCoV_1200_29_LEFT	TGAGGGAGCCTGAATACACCA	1	61.1	22
SARSCoV_1200_29_RIGHT	TAGGCAGCTCTCCCTAGCATTG	1	61.61	22
SARSCoV_1200_2_LEFT	CCATAATCAAGACTATTCAACCAAGGGT	2	61.27	28

SARSCoV_1200_2_RIGHT	ACAGGTGACAATTTGCCACCG	2	61.33	22
SARSCoV_1200_4_LEFT	GGAATTGGTGCCACTTGTGCT	2	61.66	22
SARSCoV_1200_4_RIGHT	CCTGACCCGGGTAAAGTGGTTAT	2	61.49	22
SARSCoV_1200_6_LEFT	ACTTCTATTAAATGGGCAGATAACAACTG	2	60.18	29
SARSCoV_1200_6_RIGHT	GATTATCCATTCCCTGCGCGTC	2	61.75	22
SARSCoV_1200_8_LEFT	CAATCATGCAATTGTTTTCAGCTATTTG	2	60.39	30
SARSCoV_1200_8_RIGHT	TGACTTTTGCTACCTGCGCAT	2	61.39	22
SARSCoV_1200_10_LEFT	TTTACCAGGAGTTTCTGTGGTGT	2	60.32	24
SARSCoV_1200_10_RIGHT	TGGGCCTCATAGCACATTGGTA	2	61.5	22
SARSCoV_1200_12_LEFT	ATGGTGCTAGGAGAGTGTGGAC	2	61.48	22
SARSCoV_1200_12_RIGHT	GGATTTCCCACAATGCTGATGC	2	60.48	22
SARSCoV_1200_14_LEFT	ACAGGCACTAGTACTGATGTCGT	2	61.12	23
SARSCoV_1200_14_RIGHT	GTGCAGCTACTGAAAAGCACGT	2	61.94	22
SARSCoV_1200_16_LEFT	ACAACACAGACTTATGAGTGTCTCT	2	60.18	26
SARSCoV_1200_16_RIGHT	CTCTGTCAGACAGCACTTCACG	2	61.17	22
SARSCoV_1200_18_LEFT	GCACATAAAGACAAATCAGCTCAATGC	2	62.03	27
SARSCoV_1200_18_RIGHT	TGTCTGAAGCAGTGGAAAAGCA	2	60.68	22
SARSCoV_1200_20_LEFT	ACAATTGATACTTATAACCTCTGGAACAC	2	60.15	30
SARSCoV_1200_20_RIGHT	GATTAGGCATAGCAACACCCGG	2	61.39	22
SARSCoV_1200_22_LEFT	GTGATGTTCTGTTAACAACTAAACGAACA	2	61.44	30
SARSCoV_1200_22_RIGHT	AACAGATGCAAATCTGGTGGCG	2	62.03	22
SARSCoV_1200_24_LEFT	GCTGAACATGTCAACAACTCATATGA	2	60.13	26
SARSCoV_1200_24_RIGHT	ATGAGGTGCTGACTGAGGGAAG	2	61.74	22
SARSCoV_1200_26_LEFT	GCCTTGAAAGCCCCTTTCTTA	2	60.29	22
SARSCoV_1200_26_RIGHT	AATGACCACATGGAACCGCGTAC	2	61.5	22
SARSCoV_1200_28_LEFT	TTTGTGCTTTAGCCTTCTGCT	2	60.14	24
SARSCoV_1200_28_RIGHT	GTTTGGCCTTGTGTTGGC	2	61.82	22

Fonte: Freed & Silander, 2021.

## ANEXO B – INICIADORES DO PROTOCOLO ARTICV3™

Lista de iniciadores utilizados do protocolo ArticV3™.

<b>Nome do primer</b>	<b>Seqüência</b>	<b>Pool</b>	<b>Tm</b>	<b>Tamanho</b>
nCoV-2019_1_LEFT	ACCAACCAACTTCGATCTCTTGT	1	60.69	24
nCoV-2019_1_RIGHT	CATCTTAAGATGTTGACGTGCCTC	1	60.45	25
nCoV-2019_3_LEFT	CGGTAAATAAAGGAGCTGGTGGC	1	61.32	22
nCoV-2019_3_RIGHT	AAGGTGTCTGCAATTCTAGCTCT	1	60.32	24
nCoV-2019_5_LEFT	TGGTGAAACTTCATGGCAGACG	1	61.39	22
nCoV-2019_5_RIGHT	ATTGATGTTGACTTCTCTTTGGAGT	1	60.17	28
nCoV-2019_7_LEFT	ATCAGAGGCTGCTCGTGTGTA	1	61.73	22
nCoV-2019_7_LEFT_alt0	CATTGCATCAGAGGCTGCTCG	1	62.44	22
nCoV-2019_7_RIGHT	TGCACAGGTGACAATTGTCCA	1	60.95	22
nCoV-2019_7_RIGHT_alt5	AGGTGACAATTGTCCACCGAC	1	61.07	22
nCoV-2019_9_LEFT	TCCCACAGAACGTGTTAACAGAGGA	1	61.18	24
nCoV-2019_9_LEFT_alt4	TTCCCACAGAACGTGTTAACAGAGG	1	60.44	24
nCoV-2019_9_RIGHT	ATGACAGCATCTGCCACAACAC	1	61.71	22
nCoV-2019_9_RIGHT_alt2	GACAGCATCTGCCACAACACAG	1	62.26	22
nCoV-2019_11_LEFT	GGAATTGGTGCACCTCTGCT	1	61.66	22
nCoV-2019_11_RIGHT	TCATCAGATTCAACTTGCATGGCA	1	61.35	24
nCoV-2019_13_LEFT	TCGCACAAATGTCTACTTAGCTGT	1	60.56	24
nCoV-2019_13_RIGHT	ACCACAGCAGTAAAACACCCCT	1	60.36	22
nCoV-2019_15_LEFT	ACAGTGCTTAAAAGTGTAAAAGTGCCT	1	61.32	27
nCoV-2019_15_LEFT_alt1	AGTGCTTAAAAGTGTAAAAGTGCCT	1	60.13	26
nCoV-2019_15_RIGHT	AACAGAAACTGTAGCTGGCACT	1	60.16	22
nCoV-2019_15_RIGHT_alt3	ACTGTAGCTGGCACTTGAGAGA	1	61.57	23
nCoV-2019_17_LEFT	CTTCTTCTTGAGAGAACGTGAGGACT	1	60.69	27
nCoV-2019_17_RIGHT	TTTGTGAGTGTAAACAATGCACT	1	60.11	25
nCoV-2019_19_LEFT	GCTGTTATGTACATGGCACACT	1	61.18	23
nCoV-2019_19_RIGHT	TGTCCAACCTAGGGTCAATTCTGT	1	60.40	25
nCoV-2019_21_LEFT	TGGCTATTGATTATAAACACTACACACCC	1	61.49	29
nCoV-2019_21_LEFT_alt2	GGCTATTGATTATAAACACTACACACCC	1	61.29	29
nCoV-2019_21_RIGHT	TAGATCTGTGGCCAACCTCT	1	60.83	22
nCoV-2019_21_RIGHT_alt0	GATCTGTGGCCAACCTCTTC	1	61.20	22
nCoV-2019_23_LEFT	ACAACTACTAACATAGTTACACGGTGT	1	60.26	27
nCoV-2019_23_RIGHT	ACCACTACAGTAGGTTGCAATAGT	1	60.57	25
nCoV-2019_25_LEFT	GCAATTGTTTCAGCTATTTGCAGT	1	60.73	27

nCoV-2019_25_RIGHT	ACTGTAGTGACAAGTCTCTCGCA	1	61.30	23
nCoV-2019_27_LEFT	ACTACAGTCAGCTTATGTGTCAACC	1	60.80	25
nCoV-2019_27_RIGHT	AATAACAAGCACCAAGGTACGG	1	61.13	22
nCoV-2019_29_LEFT	ACTTGTGTTCTTTGTTGCTGC	1	61.39	24
nCoV-2019_29_RIGHT	AGTGTACTCTATAAGTTGATGGTGT	1	60.69	29
nCoV-2019_31_LEFT	TTCTGAGTACTGTAGGCACGGC	1	62.03	22
nCoV-2019_31_RIGHT	ACAGAATAAACACCAGGTAAGAATGAGT	1	60.69	28
nCoV-2019_33_LEFT	ACTTTGAAGAAGCTGCGCTGT	1	61.58	22
nCoV-2019_33_RIGHT	TGGACAGTAAACTACGTCAAGC	1	61.08	25
nCoV-2019_35_LEFT	TGTCGCATTCAACCAGGACAG	1	61.39	22
nCoV-2019_35_RIGHT	ACTTCATAGCCACAAGGTTAAAGTCA	1	60.69	26
nCoV-2019_37_LEFT	ACACACACTGGTTTACTCAC	1	60.93	23
nCoV-2019_37_RIGHT	GTCCACACTCTCTAGCACCAT	1	61.48	22
nCoV-2019_39_LEFT	AGTATTGCCCTATTTCTTCATAACTGGT	1	61.00	29
nCoV-2019_39_RIGHT	TGTAACGGACACATTGAGCCC	1	60.55	22
nCoV-2019_41_LEFT	GTTCCCTTCCATCATATGCAGCT	1	60.75	23
nCoV-2019_41_RIGHT	TGGTATGACAACCATTAGTTGGCT	1	60.75	25
nCoV-2019_43_LEFT	TACGACAGATGTCTTGCTGC	1	60.93	22
nCoV-2019_43_RIGHT	AGCAGCATCTACAGCAAAGCA	1	61.14	22
nCoV-2019_45_LEFT	TACCTACAACTTGTGCTAATGACCC	1	60.57	25
nCoV-2019_45_LEFT_alt2	AGTATGTACAACACCTACAACCTGTGCT	1	60.94	29
nCoV-2019_45_RIGHT	AAATTGTTCTTCATGTTGGTAGTTAGAGA	1	60.01	30
nCoV-2019_45_RIGHT_alt7	TTCATGTTGGTAGTTAGAGAAAGTGTGTC	1	61.53	29
nCoV-2019_47_LEFT	AGGACTGGTATGATTTGTAGAAAACCC	1	61.42	28
nCoV-2019_47_RIGHT	AATAACGGTCAAAGAGAGTTAACCTCTC	1	60.06	28
nCoV-2019_49_LEFT	AGGAATTACTGTGTATGCTGCTGA	1	60.57	25
nCoV-2019_49_RIGHT	TGACGATGACTGGTTAGCATTAAATACA	1	61.05	28
nCoV-2019_51_LEFT	TCAATAGCCGCCACTAGAGGAG	1	61.34	22
nCoV-2019_51_RIGHT	AGTGCATTAACATTGGCCGTGA	1	61.14	22
nCoV-2019_53_LEFT	AGCAAAATGTTGGACTGAGACTGA	1	60.69	24
nCoV-2019_53_RIGHT	AGCCTCATAAAACTCAGGTTCCC	1	60.31	23
nCoV-2019_55_LEFT	ACTCAACTTACTTAGGAGGTATGAGCT	1	61.43	28
nCoV-2019_55_RIGHT	GGTGTACTCTCCTATTTGTACTTACTGT	1	60.54	29
nCoV-2019_57_LEFT	ATTCTACACTCCAGGGACCACC	1	61.16	22
nCoV-2019_57_RIGHT	GTAATTGAGCAGGGTCGCCAAT	1	61.26	22
nCoV-2019_59_LEFT	TCACGCATGATTTCATCTGCA	1	61.42	23
nCoV-2019_59_RIGHT	AAGAGTCCTGTTACATTTCAGCTTG	1	60.02	26

nCoV-2019_61_LEFT	TGTTTATCACCCCGCGAAGAAAGC	1	61.50	22
nCoV-2019_61_RIGHT	ATCACATAGACAACAGGTGCGC	1	61.25	22
nCoV-2019_63_LEFT	TGTTAACGCGTGTGACTGGACT	1	60.16	22
nCoV-2019_63_RIGHT	ACAAAATGCCACCACATACAACC	1	61.85	22
nCoV-2019_65_LEFT	GCTGGCTTAGCTTGTTGGGTTT	1	61.92	22
nCoV-2019_65_RIGHT	TGTCAGTCATAGAACAAACACCAATAGT	1	60.90	28
nCoV-2019_67_LEFT	GTTGTCCAACAATTACCTGAAACTTACT	1	60.43	28
nCoV-2019_67_RIGHT	CAACCTTAGAAACTACAGATAAAATCTTGGG	1	60.40	30
nCoV-2019_69_LEFT	TGTCGAAAATATACTCAACTGTGTCA	1	61.43	27
nCoV-2019_69_RIGHT	TCTTATAGCCACGGAACCTCCA	1	61.14	23
nCoV-2019_71_LEFT	ACAAATCCAATTCAAGTGTCTCTATT	1	60.54	29
nCoV-2019_71_RIGHT	TGGAAAAGAAAGGTAAGAACAAAGTCCT	1	60.80	27
nCoV-2019_73_LEFT	CAATTTGTAATGATCCATTGGGTGT	1	60.29	29
nCoV-2019_73_RIGHT	CACCAGCTGTCCAACCTGAAGA	1	62.45	22
nCoV-2019_75_LEFT	AGAGTCCAACCAACAGAACATCTATTGT	1	60.24	26
nCoV-2019_75_RIGHT	ACCACCAACCTTAGAATCAAGATTGT	1	60.69	26
nCoV-2019_77_LEFT	CCAGCAACTGTTGTGGACCTA	1	60.75	22
nCoV-2019_77_RIGHT	CAGCCCCATTAAACAGCCTGC	1	61.59	22
nCoV-2019_79_LEFT	GTGGTGATTCAACTGAATGCAGC	1	60.92	23
nCoV-2019_79_RIGHT	CATTCATCTGTGAGCAAAGGTGG	1	60.62	24
nCoV-2019_81_LEFT	GCACTTGGAAAACCTCAAGATGTGG	1	61.24	25
nCoV-2019_81_RIGHT	GTGAAGTTCTTCTTGCGAGGG	1	60.73	24
nCoV-2019_83_LEFT	TCCTTGCAACCTGAATTAGACTCA	1	60.46	25
nCoV-2019_83_RIGHT	TTTGACTCCTTGAGCACTGGC	1	61.33	22
nCoV-2019_85_LEFT	ACTAGCACTCTCCAAGGGTGTT	1	61.03	22
nCoV-2019_85_RIGHT	ACACAGTCTTTACTCCAGATTCCC	1	60.51	25
nCoV-2019_87_LEFT	CGACTACTAGCGTGCCTTGTA	1	60.16	22
nCoV-2019_87_RIGHT	ACTAGGTTCCATTGTTCAAGGAGC	1	60.81	24
nCoV-2019_89_LEFT	GTACCGCGTCCATGTGGTCATT	1	61.50	22
nCoV-2019_89_LEFT_alt2	CGCGTTCATGTGGTCATTCAA	1	62.01	22
nCoV-2019_89_RIGHT	ACCTGAAAGTCAACGAGATGAAACA	1	60.91	25
nCoV-2019_89_RIGHT_alt4	ACGAGATGAAACATCTGTTGTCACT	1	60.74	25
nCoV-2019_91_LEFT	TCACTACCAAGAGTGTGTTAGAGGT	1	60.93	25
nCoV-2019_91_RIGHT	TTCAAGTGAGAACCAAAAGATAATAAGCA	1	60.03	29
nCoV-2019_93_LEFT	TGAGGCTGGTCTAAATCACCCA	1	61.59	23
nCoV-2019_93_RIGHT	AGGTCTTCCTGCCATGTTGAG	1	60.55	22
nCoV-2019_95_LEFT	TGAGGGAGCCTGAATACACCA	1	61.10	22

nCoV-2019_95_RIGHT	CAGTACGTTTGCCGAGGCCTT	1	61.95	22
nCoV-2019_97_LEFT	TGGATGACAAAGATCCAATTCAAAGA	1	60.22	28
nCoV-2019_97_RIGHT	ACACACTGATTAAAGATTGCTATGTGAG	1	60.17	28
nCoV-2019_2_LEFT	CTGTTTACAGGTTCGCGACGT	2	61.67	22
nCoV-2019_2_RIGHT	TAAGGATCAGTGCAAGCTCGT	2	61.74	22
nCoV-2019_4_LEFT	GGTGTATACTGCTGCCGTGAAC	2	61.56	22
nCoV-2019_4_RIGHT	CACAAGTAGTGGCACCTCTTAGT	2	60.97	25
nCoV-2019_6_LEFT	GGTGTGTTGGAGAAGGTTCCG	2	61.64	22
nCoV-2019_6_RIGHT	TAGCGGCCTCTGTAAAACACG	2	61.18	22
nCoV-2019_8_LEFT	AGAGTTCTTAGAGACGGTGGGA	2	61.00	24
nCoV-2019_8_RIGHT	GCTTCAACAGCTTCACTAGTAGGT	2	60.56	24
nCoV-2019_10_LEFT	TGAGAAGTGCTCGCCTATACAGT	2	61.12	24
nCoV-2019_10_RIGHT	TCATCTAACCAATCTCTTCTGCTCT	2	60.31	27
nCoV-2019_12_LEFT	AAACATGGAGGAGGTGTTGCAG	2	61.08	22
nCoV-2019_12_RIGHT	TTCACTCTCATTCCAAAAAGCTTGA	2	60.36	27
nCoV-2019_14_LEFT	CATCCAGATTCTGCCACTCTTGT	2	60.62	23
nCoV-2019_14_LEFT_alt4	TGGCAATCTCATCCAGATTCTGC	2	61.47	24
nCoV-2019_14_RIGHT	AGTTTCCACACAGACAGGCATT	2	60.42	22
nCoV-2019_14_RIGHT_alt2	TGCGTGTCTCTGCATGTGC	2	62.76	22
nCoV-2019_16_LEFT	AATTGGAAGAAGCTGCTCGGT	2	60.82	22
nCoV-2019_16_RIGHT	CACAACTTGCGTGTGGAGGTTA	2	61.32	22
nCoV-2019_18_LEFT	TGGAAATACCCACAAGTTAATGGTTAAC	2	60.69	29
nCoV-2019_18_LEFT_alt2	ACTTCTATTAAATGGGCAGATAACAACGT	2	61.38	30
nCoV-2019_18_RIGHT	AGCTTGTACACCACACGTACAAGG	2	61.51	24
nCoV-2019_18_RIGHT_alt1	GCTTGTACACCACACGTACAAGG	2	60.30	23
nCoV-2019_20_LEFT	ACAAAGAAAACAGTTACACAACACCA	2	60.68	27
nCoV-2019_20_RIGHT	ACGTGGCTTATTAGTTGCATTGTT	2	60.28	25
nCoV-2019_22_LEFT	ACTACCGAAGTTAGGAGACATTATACT	2	61.25	29
nCoV-2019_22_RIGHT	ACAGTATTCTTGCTATAGTAGTCGGC	2	60.73	27
nCoV-2019_24_LEFT	AGGCATGCCCTTCTACTGTACTG	2	60.37	23
nCoV-2019_24_RIGHT	ACATTCTAACCATAGCTGAAATCGGG	2	61.19	26
nCoV-2019_26_LEFT	TTGTGATACATTCTGTGCTGGTAGT	2	60.28	25
nCoV-2019_26_RIGHT	TCCGCACTATACCAACATCAG	2	60.42	22
nCoV-2019_28_LEFT	ACATAGAAGTTACTGGCGATAGTTGT	2	60.13	26
nCoV-2019_28_RIGHT	TGTTAGACATGACATGAACAGGTGT	2	60.91	26
nCoV-2019_30_LEFT	GCACAACTAATGGTACTTTTGCA	2	61.19	25
nCoV-2019_30_RIGHT	ACCACTAGATACACAAACACCAG	2	60.30	26

nCoV-2019_32_LEFT	TGGTGAATACAGTCATGTAGTTGCC	2	61.09	25
nCoV-2019_32_RIGHT	AGCACATCACTACGCAACTTAGA	2	60.56	24
nCoV-2019_34_LEFT	TCCCATCTGGTAAAGTTGAGGGT	2	61.02	23
nCoV-2019_34_RIGHT	AGTGAAATTGGGCCTCATAGCA	2	60.03	22
nCoV-2019_36_LEFT	TTAGCTTGGTTGACGCTGCTG	2	61.44	22
nCoV-2019_36_RIGHT	GAACAAAGACCATTGAGTACTCTGGA	2	60.74	26
nCoV-2019_38_LEFT	ACTGTGTTATGTATGCATCAGCTGT	2	60.86	25
nCoV-2019_38_RIGHT	CACCAAGAGTCAGTCTAAAGTAGCG	2	61.13	25
nCoV-2019_40_LEFT	TGCACATCAGTAGTCTTACTCTCAGT	2	61.25	26
nCoV-2019_40_RIGHT	CATGGCTGCATCACGGTCAAAT	2	62.09	22
nCoV-2019_42_LEFT	TGCAAGAGATGGTTGTTCCC	2	61.08	22
nCoV-2019_42_RIGHT	CCTACCTCCCTTGTGTTGTGTT	2	60.69	23
nCoV-2019_44_LEFT	TGCCACAGTACGTCTACAAGCT	2	61.66	22
nCoV-2019_44_LEFT_alt3	CCACAGTACGCTCTACAAGCTGG	2	60.67	22
nCoV-2019_44_RIGHT	AACCTTCCACATACCGCAGAC	2	60.87	22
nCoV-2019_44_RIGHT_alt0	CGCAGACGGTACAGACTGTGTT	2	62.77	22
nCoV-2019_46_LEFT	TGTCGCTTCCAAGAAAAGGACG	2	61.38	22
nCoV-2019_46_LEFT_alt1	CGCTTCCAAGAAAAGGACGAAGA	2	61.35	23
nCoV-2019_46_RIGHT	CACGTTCACCTAACGTTGGCGTA	2	60.86	22
nCoV-2019_46_RIGHT_alt2	CACGTTCACCTAACGTTGGCGTAT	2	61.17	23
nCoV-2019_48_LEFT	TGTTGACACTGACTAACAAAGCCT	2	61.09	25
nCoV-2019_48_RIGHT	TAGATTACCAGAACGCAGCGTGC	2	60.74	22
nCoV-2019_50_LEFT	GTTGATAAGTACTTGATTGTTACGATGGT	2	60.59	30
nCoV-2019_50_RIGHT	TAACATGTTGCCAACCAACCA	2	60.95	22
nCoV-2019_52_LEFT	CATCAGGAGATGCCACAACTGC	2	61.83	22
nCoV-2019_52_RIGHT	GTTGAGAGCAAAATTGAGGTCC	2	60.62	25
nCoV-2019_54_LEFT	TGAGTTAACAGGACACATGTTAGACA	2	60.18	26
nCoV-2019_54_RIGHT	AACCAAAAATTGTCATTAGCACA	2	60.11	25
nCoV-2019_56_LEFT	ACCTAGACCACCACTAACCGA	2	60.49	22
nCoV-2019_56_RIGHT	ACACTATGCGAGCAGAAGGGTA	2	61.21	22
nCoV-2019_58_LEFT	TGATTTGAGTGTGTCATGCCAGA	2	61.44	25
nCoV-2019_58_RIGHT	CTTTCTCCAAGCAGGGTACGT	2	61.06	23
nCoV-2019_60_LEFT	TGATAGAGACCTTATGACAAGTTGCA	2	60.53	27
nCoV-2019_60_RIGHT	GGTACCAACAGCTCTAGTAGC	2	60.44	24
nCoV-2019_62_LEFT	GGCACATGGCTTGAGTTGACA	2	61.91	22
nCoV-2019_62_RIGHT	GTTGAAACCTTCTACAAGCCGC	2	60.35	22
nCoV-2019_64_LEFT	TCGATAGATATCCTGCTAATTCCATTGT	2	60.11	28

nCoV-2019_64_RIGHT	AGTCTTGAAAGTGTCCAGAGGT	2	60.10	25
nCoV-2019_66_LEFT	GGGTGTGGACATTGCTGCTAAT	2	61.21	22
nCoV-2019_66_RIGHT	TCAATTCCATTGACTCCTGGGT	2	60.45	24
nCoV-2019_68_LEFT	ACAGGTTCATCTAAGTGTGTGT	2	60.14	24
nCoV-2019_68_RIGHT	CTCCTTATCAGAACAGCACCA	2	60.31	23
nCoV-2019_70_LEFT	ACAAAAGAAAATGACTCTAAAGAGGGTTT	2	60.13	29
nCoV-2019_70_RIGHT	TGACCTTCTTTAAAGACATAACAGCAG	2	60.27	28
nCoV-2019_72_LEFT	ACACGTGGTGTATTACCCGTGAC	2	61.04	24
nCoV-2019_72_RIGHT	ACTCTGAACTCACTTCCATCCAAC	2	60.97	25
nCoV-2019_74_LEFT	ACATCACTAGGTTCAAACTTACTTGC	2	60.68	28
nCoV-2019_74_RIGHT	GCAACACAGTTGCTGATTCTCTTC	2	60.85	24
nCoV-2019_76_LEFT	AGGGCAAACCTGGAAAGATTGCT	2	60.76	22
nCoV-2019_76_LEFT_alt3	GGGCAAACCTGGAAAGATTGCTGA	2	61.87	23
nCoV-2019_76_RIGHT	ACACCTGTGCCTGTTAAACCATTGA	2	60.42	22
nCoV-2019_76_RIGHT_alt0	ACCTGTGCCTGTTAAACCATTGA	2	60.69	23
nCoV-2019_78_LEFT	CAACTTACTCCTACTTGGCGTGT	2	60.55	23
nCoV-2019_78_RIGHT	TGTGTACAAAAACTGCCATATTGCA	2	60.22	25
nCoV-2019_80_LEFT	TTGCCTTGGTGATATTGCTGCT	2	60.89	22
nCoV-2019_80_RIGHT	TGGAGCTAAGTTGTTAACAGCG	2	60.02	24
nCoV-2019_82_LEFT	GGGCTATCATCTTATGTCCTCCCT	2	61.52	25
nCoV-2019_82_RIGHT	TGCCAGAGATGTCACCTAAATCAA	2	60.02	24
nCoV-2019_84_LEFT	TGCTGTAGTTGCTCAAGGGCT	2	61.61	22
nCoV-2019_84_RIGHT	AGGTGTGAGTAAACTGTTACAACAAAC	2	60.36	27
nCoV-2019_86_LEFT	TCAGGTGATGGCACAACAAGTC	2	61.07	22
nCoV-2019_86_RIGHT	ACGAAAGCAAGAAAAAGAAGTACGC	2	61.01	25
nCoV-2019_88_LEFT	CCATGGCAGATTCCAACGGTAC	2	61.58	22
nCoV-2019_88_RIGHT	TGGTCAGAATAGGCCATGGAGT	2	61.40	23
nCoV-2019_90_LEFT	ACACAGACCATTCCAGTAGCAGT	2	61.58	23
nCoV-2019_90_RIGHT	TGAAATGGTGAAATTGCCCTCGT	2	60.82	22
nCoV-2019_92_LEFT	TTTGTGCTTTAGCCTTCTGCT	2	60.14	24
nCoV-2019_92_RIGHT	AGGTTCTGGCAATTAATTGTAAGG	2	60.53	27
nCoV-2019_94_LEFT	GGCCCCAAGGTTACCCAATAA	2	60.56	22
nCoV-2019_94_RIGHT	TTTGGCAATGTTGTCCTGAGG	2	60.18	23
nCoV-2019_96_LEFT	GCCAACAACAACAAGGCCAAC	2	61.82	22
nCoV-2019_96_RIGHT	TAGGCTCTGTTGGTGGGAATGT	2	61.36	22
nCoV-2019_98_LEFT	AACAATTGCAACAATCCATGAGCA	2	60.50	24
nCoV-2019_98_RIGHT	TTCTCCTAAGAAGCTATTAACATGG	2	60.01	30