



Instituto de Biologia do Exército

Divisão de Ensino e Pesquisa

Aline Lima de Moura Natarelli

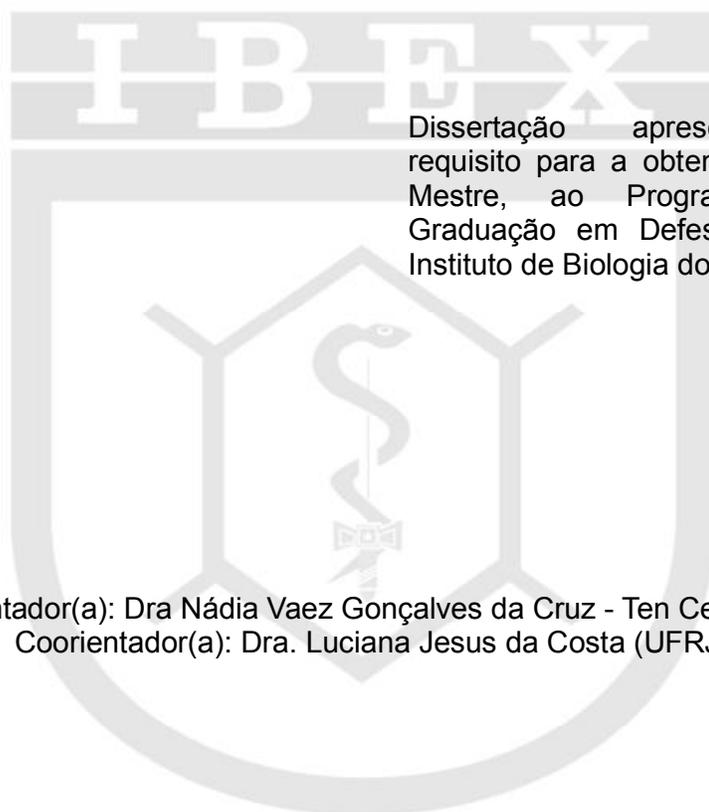
TÍTULO: ANÁLISE GENÔMICA DE SARS-COV-2 NO ESTADO DO RIO DE JANEIRO EM AMOSTRAS DIAGNOSTICADAS NO INSTITUTO DE BIOLOGIA DO EXÉRCITO E EM GENOMAS DEPOSITADOS NO GISAID, DURANTE A PANDEMIA DA COVID-19

Rio de Janeiro

2023

Ten Cel ALINE LIMA DE MOURA NATARELLI

TÍTULO: ANÁLISE GENÔMICA DE SARS-COV-2 NO ESTADO DO RIO DE JANEIRO EM AMOSTRAS DIAGNOSTICADAS NO INSTITUTO DE BIOLOGIA DO EXÉRCITO E EM GENOMAS DEPOSITADOS NO GISAID, DURANTE A PANDEMIA DA COVID-19



Dissertação apresentada como requisito para a obtenção do grau de Mestre, ao Programa de Pós-Graduação em Defesa Biológica, do Instituto de Biologia do Exército.

Orientador(a): Dra Nádia Vaez Gonçalves da Cruz - Ten Cel (IBEx)
Coorientador(a): Dra. Luciana Jesus da Costa (UFRJ)

Rio de Janeiro

2023

FICHA CATALOGRÁFICA

N273 Natarelli, Aline Lima da Moura

Análise genômica de SARS-CoV-2 no Estado do Rio de Janeiro em amostras diagnosticadas no Instituto de Biologia do Exército e em genomas depositados no GISAID, durante a pandemia da COVID-19/Aline Lima de Moura Natarelli – Rio de Janeiro, 2023.

128 folhas.

Orientador (a): Dra Nádia Vaez Gonçalves da Cruz

Coorientadores: Dra Luciana Jesus da Costa

Dissertação (Mestrado) – Instituto de Biologia do Exército, Divisão de Ensino e Pesquisa, Pós-Graduação em Defesa Biológica, 2023.

Referências: f. 102-116.

1. SARS-CoV-2 2. MUTAÇÕES. 3. ANÁLISE GENÔMICA 4. COVID-19. 5. SEQUENCIAMENTO POR NANOPOROS. I. Dra Nádia Vaez Gonçalves da Cruz. II. Instituto de Biologia do Exército. III. Título.

CDD 543.08

Ten Cel ALINE LIMA DE MOURA NATARELLI

TÍTULO: ANÁLISE GENÔMICA DE SARS-COV-2 NO ESTADO DO RIO DE JANEIRO EM AMOSTRAS DIAGNOSTICADAS NO INSTITUTO DE BIOLOGIA DO EXÉRCITO E EM GENOMAS DEPOSITADOS NO GISAID, DURANTE A PANDEMIA DA COVID-19

Dissertação apresentada como requisito para a obtenção do grau de Mestre, ao Programa de Pós-Graduação em Defesa Biológica, do Instituto de Biologia do Exército.

Data de Aprovação:

Banca Examinadora:

Dra.Nádia Vaez Gonçalves da Cruz – Ten Cel
Instituto de Biologia do Exército – IBEx
Presidente/Orientadora

Dr. Marcio da Costa Cipitelli – 1º Ten
Instituto de Biologia do Exército – IBEx
Membro

Dra. Aline Rosa Vianna de Souza
Universidade Federal Fluminense
Membro

Rio de Janeiro
2023

DEDICATÓRIA

À minha mãe, meu esposo e a minha filha
uma homenagem como recompensa pela
execução deste trabalho diante da
grandeza de suas pessoas.

AGRADECIMENTOS

Agradeço à minha orientadora, Dra. Nádia Vaez Gonçalves da Cruz, pela atenção e apoio constantes, pela grata convivência, por sua doçura e tranquilidade, mesmo nos momentos mais difíceis. Expresso minha admiração pela excelência do seu trabalho como professora e pesquisadora e por todo o conhecimento adquirido, que constituem uma grande fonte de inspiração profissional para mim.

À minha coorientadora, Dra. Luciana Jesus da Costa, que se mostrou colaboradora, atenciosa e paciente durante essa jornada.

Ao grupo de trabalho do Laboratório de Metabolismo Macromolecular Firmino Torres de Castro, da UFRJ, chefiado pela Dra Rosane Silva, pela colaboração no sequenciamento e na pesquisa com os genomas diagnosticados no IBEx.

Aos professores Dra. Aline Rosa Vianna de Souza e Dr Marcio da Costa Cipitelli, por terem aceitado o convite para compor a banca examinadora e a Dra Vanessa Santos, pelas importantes contribuições na revisão deste trabalho.

Agradeço a Ten Andressa Duarte e aos demais integrantes da Divisão de Ensino e Pesquisa do IBEx, por terem me auxiliado nas atividades de laboratório, sempre que necessário.

Ao Diretor do Instituto de Biologia do Exército, Cel André Luiz Meriano Figueiredo e ao Chefe da Divisão Técnica, Ten Cel Marcos Dornelas Ribeiro, por disponibilizarem meios e apoio necessários à realização deste trabalho. A todos os meus colegas do IBEx, que de forma direta ou indireta, prestaram importantes contribuições.

Agradeço imensamente aos meus amados pais Jorge (*in memoriam*) e Maryland por terem tornado tudo possível desde o início, ao me incentivarem nos estudos e concederem, com muito esforço, uma boa educação. Obrigada por me apoiarem sempre, com amor e compreensão.

Meu agradecimento especial ao meu amor Cesar Augusto, esposo e pai dedicado, cujo apoio foi imprescindível para que eu atingisse essa conquista pessoal e profissional. Muito obrigada por toda compreensão, paciência e por me ajudar sempre, inclusive com um bom vinho em sua companhia. É um privilégio ter você ao meu lado.

Agradeço à minha amada filha Luiza pelas demonstrações de afeto e carinho, tão revigorantes nessa árdua caminhada, verdadeira alegria do meu viver.

Ao final, expresso minha gratidão a Deus, pela vida, pelas oportunidades e pela força na superação dos desafios.

Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse uma gota.

(Madre Teresa de Calcutá)

RESUMO

NATARELLI, Aline Lima de Moura. *Análise genômica de SARS-CoV-2 no Estado do Rio de Janeiro em amostras diagnosticadas no Instituto de Biologia do Exército e em genomas depositados no GISAID, durante a pandemia da COVID-19*. 2023. XXX f. Dissertação (Mestrado em Defesa Biológica) – Divisão de Ensino e Pesquisa, Instituto de Biologia do Exército, Rio de Janeiro, 2023.

No final de dezembro de 2019, casos de pneumonia grave causada por agente viral não identificado foram relatados em Wuhan, China. Tratava-se do SARS-CoV-2, cuja rápida disseminação geográfica pelo mundo fez com que a Organização Mundial da Saúde (OMS) elevasse a sua classificação para pandemia, em 11 de março de 2020. Com o objetivo de monitorar a rápida evolução viral, pesquisadores de todo o mundo utilizaram diversas plataformas de sequenciamento de alta performance para caracterizar o genoma completo do SARS-CoV-2, dentre elas, as tecnologias de nanoporos (MinION-Nanopore) e de semicondutores (ION S5-ThermoFisher), utilizadas no sequenciamento de nossas amostras. Esse esforço global permitiu, desde 2020, a obtenção de dados que auxiliam na compreensão da evolução do vírus e de sua dinâmica de transmissão. Neste estudo, sequenciamos 20 genomas completos de SARS-CoV-2 dos anos de 2020 (19) e 2021 (1) obtidos a partir de swab nasofaríngeo de pacientes positivos diagnosticados no IBEx por RT-PCR em tempo real. Os genomas eram oriundos de 03 municípios do estado do Rio de Janeiro e tiveram suas mutações gênicas e as respectivas linhagens analisadas. Também selecionamos outros genomas completos de SARS-CoV-2, disponíveis no GISAID nos períodos concomitantes às amostras selecionadas para análise genômica comparativa, e genomas adicionais referentes aos períodos de dezembro de 2021 a janeiro de 2022 (entrada da variante Ômicron) e ao 1º semestre de 2023, visando a obter um panorama mais atual das linhagens circulantes e respectivas mutações. Ao todo, 1521 genomas completos foram analisados. As frequências de linhagens e mutações nas amostras de 2020 foram semelhantes às encontradas no GISAID e também por outros autores. Observou-se ao longo dos anos o aumento constante na média de substituições de aminoácidos (de 7 em 2020 para 64 em 2023), com especial destaque para o gene que codifica a proteína S. Em outubro de 2021, observamos que a frequência da linhagem AY.99.2 (Delta) representava 88,07% dos genomas, sendo reduzida rapidamente para 36,71% entre dezembro de 2021 a janeiro de 2022, com a entrada da variante Ômicron. Neste período, circulavam no estado do Rio de Janeiro pelo menos 26 diferentes tipos de linhagens de SARS-CoV-2, das variantes Gamma, Delta e Ômicron, além da recombinante XZ, dentre os genomas GISAID. Já no 1º semestre de 2023, todos os genomas avaliados pertenceram à variante Ômicron, ressaltando-se que as 10 linhagens de Ômicron que foram observadas entre dezembro de 2021 e janeiro de 2022 não permaneceram em 2023, dando lugar a 36 novas linhagens, muitas delas recombinantes, ocorrendo mesmo com a queda do número de casos e mortes no estado e no país. Seguindo a tendência esperada, as mutações em S continuaram a ocorrer com bastante frequência. Dentre os genomas de 2023, foram observadas 40 novas mutações em S em comparação às linhagens de SARS-CoV-2 do período anterior. Interessante destacar que a mutação D614G, observada logo no início da pandemia, manteve-se conservada até as linhagens atuais da variante Ômicron. O gene E, ao contrário de S, manteve-se conservado, até apresentar em 2023 a mutação T9I em 99,2% dos genomas avaliados. Esses resultados destacam a importância da análise genômica e a necessidade do contínuo sequenciamento genômico seguido por estudos, para rastreamento de novas variantes de preocupação, entendimento da dinâmica de transmissão, permitindo o desenvolvimento de novas técnicas de diagnóstico, vacinas e antivirais.

Palavras-chave: SARS-CoV-2, mutações, análise genômica

ABSTRACT

NATARELLI, Aline Lima de Moura. *Análise genômica de SARS-CoV-2 no Estado do Rio de Janeiro em amostras diagnosticadas no Instituto de Biologia do Exército e em genomas depositados no GISAID, durante a pandemia da COVID-19*. 2023. XXX f. Dissertação (Mestrado em Defesa Biológica) – Divisão de Ensino e Pesquisa, Instituto de Biologia do Exército, Rio de Janeiro, 2023.

In late December 2019, cases of severe pneumonia caused by an unidentified viral agent were reported in Wuhan, China. This agent was further identified as SARS-CoV-2, agent of COVID-19. The rapid spread of the virus around the world led the WHO to classify the disease as a pandemic on March 11th, 2020. To monitor viral genome evolution, researchers from different parts of the world used a several high-performance sequencing platforms, including MinION and ION S5 (nanopore and semiconductor based technologies). This global effort has allowed collection of massive data which helped understanding viral evolution and transmission dynamics of COVID-19. In this study, we sequenced 20 genomes of SARS-CoV-2 and analyzed comparatively with sequences available at GISAID database. Viral genomes were obtained from nasopharyngeal swabs of patients living on Rio de Janeiro state, previously diagnosed with COVID-19 at IBEX, by RT-PCR. Sequenced genomes were isolated mainly in 2020 (19 samples). We selected sequences at GISAID from periods of the most important phases of COVID-19 spread in Rio de Janeiro state: March to June of 2020 (298 sequences) and October of 2021 (788 sequences). We also performed genomic analysis in additional genomes from GISAID-RJ collected from December 2021 to January 2022 (158 sequences, representing the entry of Ômicron variant) and January to June 2023 (257 sequences, to obtain a more current view of circulating lineages and mutations). A total of 1521 complete genomes were analyzed, focusing on detection of mutations and identification of circulating lineages. The lineage frequencies and mutations in the 2020 samples were similar to those found in GISAID and also by other authors. Over the years, a constant increase in the average number of amino acid substitutions was observed (from 7 in 2020 to 64 in 2023), with special emphasis on the gene S. In October 2021, the frequency of the lineage AY.99.2 (Delta) represented 88.07% of the genomes, being quickly reduced to 36.71% between December 2021 and January 2022, with the entry of the Ômicron variant in the state. During this period, at least 26 different SARS-CoV-2 strains were circulating in the state, including lineages from Gamma, Delta, Ômicron variants and the Ômicron's recombinant XZ lineage. In the 1st semester of 2023, all genomes evaluated belonged to the Ômicron variant, highlighting that the 10 Ômicron lineages observed between December 2021 and January 2022 did not remain in 2023, giving way to 36 new lineages, some of them recombinants, occurring even with the drop in the number of cases and deaths in the state and country. Following the expected trend, mutations in S continued to occur quite frequently. Among the 2023 genomes, 40 new mutations were observed in S compared to SARS-CoV-2 lineages from the previous period. We also observed that the D614G mutation, found at the beginning of the pandemic, remained preserved until the current lineages of the Ômicron variant. Unlike gene S, the E gene remained conserved until 2023, when it started to present the T9I mutation in 99.2% of the genomes. These results highlight the importance of genomic analysis and the need for continuous genomic sequencing, to track new variants of concern, understand transmission dynamics, allowing the development of new diagnostic techniques, vaccines and antivirals.

Keywords: SARS-CoV-2, mutations, genomic analysis

LISTA DE QUADROS

Quadro 1- Epidemiologia da COVID-19 no Estado do Rio de Janeiro.....	29
Quadro 2- Informações das amostras sequenciadas.....	46
Quadro 3 – Quadro informativo da quantidade de amostras sequenciadas e de genomas obtidos no GISAID para análise genômica.....	55
Quadro 4- Informações de qualidade e cobertura das amostras sequenciadas.....	56
Quadro 5- Frequência de linhagens de SARS-CoV-2 em genomas sequenciados no IBEx e obtidos no GISAID durante os anos de 2020 a 2023.....	62
Quadro 6- Substituições de aminoácidos nas amostras sequenciadas em 2020 (1 a 19) e 2021(20), por região do genoma.....	69
Quadro 7 - Frequências das principais substituições de nucleotídeos e aminoácidos nas 19 amostras sequenciadas de 2020.....	71
Quadro 8- Comparativo de número de substituições de aminoácidos entre as amostras IBEx e sequências GISAID 2020 a 2023, por região do genoma.....	73
Quadro 9- Comparativo de médias substituições de nucleotídeos, aminoácidos, inserções, deleções e mudanças de quadro de leitura no universo total de genomas.....	74
Quadro 10- Principais substituições de aminoácidos no gene S observadas entre 2020 e 2023, de acordo com a variante (denominação OMS) e clado.....	76

Quadro 11- Principais substituições de aminoácidos em ORF1a observadas entre 2020 e 2023, de acordo com a variante (denominação OMS) e clado.....	78
Quadro 12- Mapa de calor com as frequências de todas as mutações em S encontradas nos genomas de 2020 a 2023.....	79
Quadro 13- Informações de qualidade do genoma e principais mutações observadas na recombinante XZ.....	82
Quadro 14-Frequência de inserções nas principais linhagens em cada período de estudo.....	83
Quadro 15- Frequências das deleções nas principais linhagens de SARS-COV-2 nos genomas GISAID RJ entre 2020 e 2023.....	85
Quadro 16-Frequências das mudanças de quadro de leitura observadas nas principais linhagens de SARS-CoV-2 em cada período de estudo.....	87

LISTA DE FIGURAS

Figura 1- Total de sequenciamentos genômicos de SARS-CoV-2 no Estado do Rio de Janeiro por mês de coleta, no período de abril de 2020 a abril de 2023.....	28
Figura 2- Representação esquemática da estrutura e do genoma do SARS-CoV-2.....	33
Figura 3– Representação esquemática do ciclo de replicação do SARS-CoV-2.....	34
Figura 4– Figura representativa da filogenia da pandemia por coronavírus no Brasil com base em dados do GISAID.....	38
Figura 5 - Etapas e ferramentas utilizadas para a avaliação da qualidade das sequências, visualização e agrupamento dos genomas e alinhamento ao genoma referência.....	53
Figura 6 - Ferramentas e softwares utilizados para a chamada de variantes, atribuição de clados, análise de polimorfismos e análise filogenética.....	54
Figura 7- Figura representativa dos dados de qualidade de alguns dos 24 genomas sequenciados em 2022.....	57
Figura 8- Figura representativa dos dados de qualidade do sequenciamento da amostra “Barcode11”. (Nº20-7842/2021 IBEx)	59
Figura 9- Figura representativa dos dados dos escores de qualidade das leituras da amostra “Barcode11” (Nº20-7842/2021 IBEx)	59
Figura 10- Figura representativa dos dados do escore médio de qualidade de leitura da amostra “Barcode11” (Nº20-7842/2021 IBEx).....	60
Figura 11- Figura representativa dos dados de cobertura dos Pools de Primers A e B da amostra “Barcode11” (Nº20-7842/2021 IBEx)...	60

Figura 12 – Frequência de linhagens de SARS-CoV-2 em genomas sequenciados no IBEx e obtidos no GISAID durante os anos de 2020 a 2023.....	61
Figura 13 - Frequência de linhagens nas 19 amostras IBEx e no universo total nos meses de Abril e Junho de 2020.....	64
Figura 14 - Frequência de linhagens na amostra IBEx e nos 788 genomas GISAID do Estado do Rio de Janeiro. Amostras coletadas entre 1º e 31 de Outubro de 2021.....	65
Figura 15 - Frequência de variantes nos 158 genomas GISAID do Estado do Rio de Janeiro. Amostras coletadas entre 1º de Dezembro e 31 de Janeiro de 2022.....	66
Figura 16 – Dados consolidados, mostrando as frequências das diversas linhagens dos genomas GISAID RJ agrupadas por tipo de variante.....	66
Figura 17 – Frequência de variantes nos 257 genomas GISAID do Estado do Rio de Janeiro. Amostras coletadas entre 1º de Janeiro e 30 de Junho de 2023.....	67
Figura 18 – Frequência de variantes nos 66 genomas Ômicron 2021/2022 e nos 257 genomas da variante em 2023, GISAID do Estado do Rio de Janeiro.....	68
Figura 19 - Figura representativa dos SNP'S e deleções observados nas amostras sequenciadas 2020 e 2021.....	71
Figura 20 – Figura representativa das substituições de nucleotídeos (44) e de aminoácidos (33) apresentadas pela amostra 7842/2021 IBEx.....	72
Figura 21 – Figura representativa das deleções (vermelho) observadas nas amostras sequenciadas 1 a 5 (2020) e 20 (2021).....	84
Figura 22- Figura representativa dos 5 genomas IBEx sequenciados que apresentaram deleção de nucleotídeos entre as posições 27701 e 27759 (sinalizada em preto) e mudança de quadro de leitura em ORF 7a (faixa de códons 104 a 122).....	87

- Figura 23 - Árvore filogenética genomas das amostras sequenciadas, genomas de 2020 (representando o início da pandemia) e 2021 GISAID, além do genoma de referência Wuhan..... 89
- Figura 24 - Árvore filogenética genomas das amostras sequenciadas e genomas de linhagens de maior frequência entre os anos de 2020 e 2023, além do genoma de referência Wuhan..... 90

SUMÁRIO

1	INTRODUÇÃO	27
1.1	CENÁRIO DA PANDEMIA NO MUNDO E NO BRASIL	27
1.2	CLASSIFICAÇÃO, CARACTERÍSTICAS E CICLO REPLICATIVO DO SARS-COV-2	30
	1.2.1. Características gerais e classificação taxonômica	30
	1.2.2 Características estruturais e genômicas	31
	1.2.3. Ciclo Replicativo	33
1.3	MUTAÇÕES E O SURGIMENTO DE NOVAS VARIANTES	35
1.4	LINHAGENS CIRCULANTES NO BRASIL E NO MUNDO	37
1.5	TECNOLOGIAS DE SEQUENCIAMENTO DO GENOMA COMPLETO	40
1.6	A VIGILÂNCIA GENÔMICA E SUA IMPORTÂNCIA PARA A BIODEFESA	43
2.0	OBJETIVOS	45
2.1	OBJETIVO GERAL	45
2.2	OBJETIVOS ESPECÍFICOS	45
3	MATERIAL E MÉTODOS	46
3.1	REGULAMENTAÇÃO ÉTICA	46
3.2	AMOSTRAGEM	46
3.3	SELEÇÃO DE GENOMAS NO GISAID	48
3.4	EXTRAÇÃO DE RNA	48
3.5	REALIZAÇÃO DE PCR EM TEMPO REAL	49
3.6	PREPARO DA BIBLIOTECA GENÔMICA E SEQUENCIAMENTO DE AMOSTRAS COLETADAS NO IBEX	49
3.7	ANÁLISE DA QUALIDADE DE LEITURA E QUALIDADE DAS SEQUÊNCIAS	52
3.8	ALINHAMENTO, CHAMADA DE VARIANTES, ATRIBUIÇÃO DE CLADOS E ANÁLISE DE POLIMORFISMOS	52
3.9	ANÁLISE FILOGENÉTICA	54
4	RESULTADOS	55
4.1	UNIVERSO DE SEQUÊNCIAS ANALISADAS	55

4.2	QUALIDADE DAS SEQUÊNCIAS E COBERTURA DOS GENOMAS	55
4.3	FREQUÊNCIA DE VARIANTES	61
4.4	SUBSTITUIÇÕES DE NUCLEOTÍDEOS E DE AMINOÁCIDOS	67
	4.4.1. No universo de amostras sequenciadas	67
	4.4.2. Nas sequências do Estado do Rio de Janeiro obtidas no GISAID entre 2020 e 2023.	73
4.5	INSERÇÕES E DELEÇÕES	82
4.6	MUDANÇAS DE QUADRO DE LEITURA (FRAMESHIFTS)	87
4.7	ANÁLISE FILOGENÉTICA	88
5	DISCUSSÃO	91
	REFERÊNCIAS	102
	ANEXO A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	118
	ANEXO B- ESQUEMA DE PRIMERS V4	122

LISTA DE ABREVIATURAS E SIGLAS

Ac	Anticorpo
ACE2	Receptor da enzima conversora da Angiotensina 2 (do inglês: <i>angiotensin-converting enzyme 2</i>)
cDNA	DNA complementar
dsDNA	DNA dupla fita
COVID-19	Doença do Coronavírus-19 (do inglês <i>Coronavirus disease 19</i>)
DNA	Ácido desoxirribonucleico
dNTP	Desoxirribonucleotídeo trifostato
dsDNA	DNA de dupla fita
EB	Exército Brasileiro
GISAID	Iniciativa global no compartilhamento de dados de Influenza (do inglês, <i>Global Initiative on Sharing All Influenza Data</i>)
MAFFT	Programa de alinhamento múltiplo para sequências de nucleotídeos ou aminoácidos, do inglês, “ <i>Multiple alignment program for amino acid or nucleotide sequences</i> ”)
nsp5	Proteínas não-estruturais individuais
NGS	Sequenciamento de Nova Geração (do inglês <i>Next Generation Sequencing</i>)
OMS	Organização Mundial da Saúde
ORF	Janela aberta de leitura (do inglês <i>Open Reading Frame</i>)
pb	Pares de base
PCR	Reação em Cadeia da Polimerase (do inglês <i>Polimerase Chain Reaction</i>)
qPCR	PCR em tempo real quantitativo
RBD	Domínio de ligação ao receptor (do inglês <i>Receptor binding domain</i>)
RBM	Motivo de ligação ao receptor (do inglês <i>Receptor binding motif</i>)
RNA	Ácido ribonucleico
rpm	Rotação por minuto
SARS-CoV-2	Coronavírus 2 da síndrome respiratória aguda grave
SNPs	Polimorfismos de nucleotídeo único (do inglês <i>Single Nucleotide Polymorphism</i>).
UTRs	Regiões não traduzidas do genoma (do inglês: <i>untranslated regions</i>)
VOI	Variante de interesse (do inglês, <i>Variants of Interest</i>)

VOC	Variante de preocupação (do inglês, <i>Variants of Concern</i>)
VUM	Variante sob monitoramento (do inglês, <i>Variant under monitoring</i>)

1 INTRODUÇÃO

1.1 CENÁRIO DA PANDEMIA NO MUNDO E NO BRASIL

No final de dezembro de 2019, pacientes com pneumonia devido a um agente viral não identificado foram relatados em Wuhan, China. Um novo coronavírus foi posteriormente identificado como o patógeno causador, provisoriamente chamado de novo coronavírus de 2019 (2019-nCoV) (LU et al., 2020). Em março de 2020, a rápida disseminação geográfica do vírus fez com que a Organização Mundial da Saúde (OMS) elevasse a classificação do novo coronavírus (SARS-CoV-2) para pandemia (LIU et al., 2020). Segundo a OMS, pandemia é a disseminação mundial de uma nova doença e o termo passa a ser usado quando uma epidemia, surto que afeta uma região, se espalha por diferentes continentes com transmissão sustentada de pessoa para pessoa. Ainda de acordo com a OMS, até 21 de setembro de 2023, mais de 770 milhões de casos de COVID-19 tinham sido confirmados mundialmente, incluindo mais de 6,9 milhões de mortes (OMS, 2023). Embora a pandemia seja sempre motivo de preocupação e monitoramento, a redução de casos e de mortes foi expressiva a partir do início da vacinação, que ocorreu mundialmente em 08 de dezembro de 2020, e no Brasil em 17 de janeiro de 2021. Até 17 de setembro de 2023, 13.504.973.753 de doses de vacina tinham sido administradas no mundo (OMS, 2023).

No Brasil, dados do Painel Coronavirus (Ministério da Saúde, 2023) mostravam que até o dia 19 de setembro de 2023, foram registrados e confirmados 37.778.122 casos e 705.645 óbitos por COVID-19. A alta incidência e mortalidade pela doença ocorridas no Brasil indicam a grande circulação do vírus nos anos de 2020 a 2023. De acordo com o site oficial da OMS, os dados consolidados até 12 de setembro de 2023 informavam que os Estados Unidos da América eram o país com maior número de casos no mundo. O Brasil ocupava a 6ª colocação em número de casos confirmados (OMS, 2023).

Na cidade do Rio de Janeiro, foram observados, desde o ano de 2020 até março de 2022, 5 ondas da COVID-19, segundo dados do Boletim Epidemiológico dos casos de COVID-19 no Estado do Rio de Janeiro 2020-2021 (Rio de Janeiro, 2022). Este Boletim informa que, na primeira onda, que ocorreu entre os meses de

abril e maio de 2020, houve predominância da cepa inicial que chegou ao Brasil, a variante B.1.1.33, que já se encontrava dispersa mundialmente (RESENDE et al., 2021). Durante essa onda, foram registrados 99.760 casos, 24.174 internações e 11.270 óbitos. A segunda onda ocorreu entre os meses de novembro de 2020 e janeiro de 2021, com predomínio da variante P.2 (Zeta), considerada uma variante de interesse (VOI) e identificada primeiro no Rio de Janeiro e depois em outros estados do Brasil (VOLOCH et al., 2021). No período, foram registrados 282.339 casos, 27.778 internações e 10.621 óbitos. A Figura 1 mostra os sequenciamentos ocorridos no Estado do Rio de Janeiro no período de abril de 2020 a abril de 2023, evidenciando que a vigilância genômica no Estado não foi capaz de gerar dados robustos nas duas primeiras ondas da pandemia (Painel Saúde, RJ, 2023).

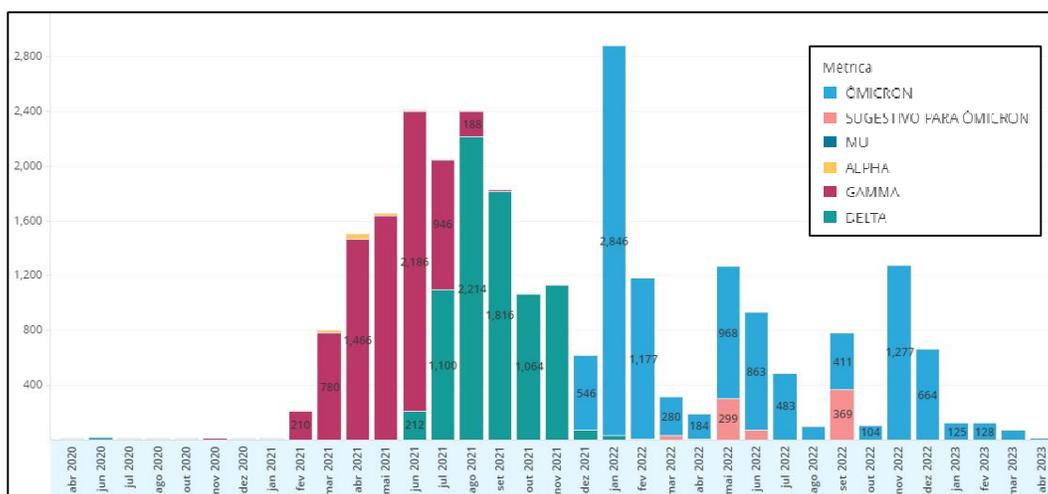


FIGURA 1 –Total de sequenciamentos genômicos de SARS-CoV-2 no Estado do Rio de Janeiro por mês de coleta, no período de abril de 2020 a abril de 2023, mostrando as variantes predominantes em cada período – Fonte: Painel Saúde, RJ, 2023.

A terceira onda teve início em fevereiro de 2021, com a entrada da variante P.1 (Gama) no estado. Ela foi detectada inicialmente em Manaus, disseminando-se posteriormente para o restante do país (FUJINO et al., 2021). Esta foi a onda com a maior duração, ocorrendo até meados do mês de junho (28/02/2021 a 12/06/2021). Ao todo, foram 358.980 casos notificados, 55.150 internações e 20.079 óbitos (Rio de Janeiro, 2022). Em agosto de 2021, o estado entrou na 4ª onda com a chegada da variante A.Y.99.2 (Delta) ao país (ROMANO et al., 2022). Esta onda se destacou

por ter ocorrido após o início da campanha de vacinação (janeiro de 2021), tendo sido menos transmissível que as outras, registrando, no período, 71.053 casos, 6.772 internações e 2.524 óbitos. A quinta onda ocorreu com a entrada da variante BA.1 (Ômicron) no Brasil, entre dezembro de 2021 e janeiro de 2022. A partir de janeiro de 2022, a exemplo do que ocorreu no restante do país, a Ômicron e suas subvariantes passaram a dominar o cenário da COVID-19 no Rio de Janeiro. A Quadro 1 mostra o número de casos confirmados, casos graves, óbitos, e as taxas de incidência e de mortalidade por 100.000 habitantes nos anos de 2020 a 2023, e evidencia uma expressiva redução dessas taxas a partir de 2022, quando a maior parte da população já se encontrava vacinada contra a COVID-19 (Rio de Janeiro, 2022).

QUADRO 1 – Epidemiologia da COVID-19 no Estado do Rio de Janeiro

	2020	2021	2022	2023
CASOS CONFIRMADOS	222.975	309.694	762.615	57.247
CASOS GRAVES	42.587	43.315	7.748	922
ÓBITOS	18.888	16.405	2.858	214
TAXA DE INCIDÊNCIA (por 100mil hab)	3.347,3	4.649,1	11.448,3	859,4
TAXA DE MORTALIDADE (por 100 mil hab)	283,5	246,2	42,9	3,2

Fonte: Observatório Epidemiológico do Rio de Janeiro, disponível no site oficial da Plataforma ArcGis (dados atualizados até 24/09/23).

Finalmente, após mais de três anos depois de ter sido decretada pandemia, em 5 de maio de 2023, a OMS declarou o fim da Emergência de Saúde Pública Internacional (ESPI II) da pandemia da COVID-19, o mais alto título de alerta da Organização, com base na tendência decrescente nas mortes por COVID-19, na queda no número de hospitalizações e internações devido à infecção e no alto nível de imunização da população. O decreto não significa, entretanto, o fim da pandemia, uma vez que a doença da COVID-19 ainda mantém sua disseminação global, com transmissão sustentada (OMS, 2023).

1.2 CLASSIFICAÇÃO, CARACTERÍSTICAS E CICLO REPLICATIVO DO SARS-CoV-2

1.2.1. Características gerais e classificação taxonômica

A COVID-19 representa a terceira e maior disseminação de coronavírus nos últimos 20 anos, após a Síndrome Respiratória Aguda Grave (*Severe Acute Respiratory Syndrome* - SARS) e Síndrome Respiratória do Oriente Médio (MERS) ocorridas em 2002 e 2012, respectivamente (GIOVANETTI et al., 2021). É causada pelo vírus SARS-CoV-2, e tem sua transmissão principalmente por meio de pequenos aerossóis, gerados pela boca ou nariz de uma pessoa infectada enquanto ela tosse, espirra, fala, canta ou respira. As evidências disponíveis sugerem que o vírus se dispersa principalmente entre pessoas que estão em contato próximo umas com as outras, normalmente dentro da distância de 1 metro (curto alcance). Entretanto, o vírus também pode alcançar distâncias maiores que 1 metro de distância (longo alcance) e se espalhar em ambientes internos mal ventilados e/ou com aglomerações, uma vez que os aerossóis permanecem suspensos no ar (OPAS, 2023). Uma pessoa pode ser infectada quando aerossóis ou gotículas contendo o vírus são inalados ou entram em contato direto com os olhos, nariz ou boca (OPAS, 2023).

Os principais sintomas apresentados pelos indivíduos infectados são febre, cansaço e tosse seca. Outros sintomas menos comuns e que podem afetar alguns pacientes são: perda de paladar ou olfato, congestão nasal, conjuntivite, dor de garganta, dor de cabeça, dores nos músculos ou juntas, diferentes tipos de erupção cutânea, náusea ou vômito, diarreia, calafrios ou tonturas (OPAS, 2023).

Os coronavírus (CoVs) são um grupo de vírus envelopados, pertencentes à família *Coronaviridae*. Os vírus desta família foram assim denominados por possuírem um conjunto de estruturas proteicas que lembram o formato de uma coroa. Apresentam genoma de RNA fita simples, não segmentado, de sentido positivo, de aproximadamente 30.000 bases. (ABBASIAN et al., 2023). As partículas virais possuem a forma esférica e tamanho do virion variando de 70 a 90nm (GRUDLEWSKA-BUDA et al., 2021). Causam principalmente infecções do trato

respiratório e gastrointestinal e são geneticamente classificados em quatro gêneros principais: *Alphacoronavirus* (α CoV), *Betacoronavirus* (β CoV), *Gammacoronavirus* (γ CoV) e *Deltacoronavirus* (δ CoV). O SARS-CoV-2 pertence ao gênero *Betacoronavirus*, subfamília *Coronavirinae*, família *Coronaviridae*, ordem *Nidovirales* e reino *Riboviria*. (GORBALENYA et al., 2020). (GUSEV et al., 2022). Os morcegos servem como hospedeiros evolutivos dos *Alfacoronavírus* e dos *Betacoronavírus*. O sequenciamento completo do genoma e a análise filogenética classificaram o SARS-CoV-2 como *Betacoronavírus* do subgênero *Sarbecovirus*, que também inclui o SARS-CoV-1 (RASTOGI et al., 2020). O SARS-CoV-2 compartilha semelhança com duas cepas de Coronavírus derivadas de morcegos bat-SL-CoVZC45 e bat-SL-CoVZXC21 (LU et al., 2020)

1.2.2 Características estruturais e genômicas

O genoma do SARS-CoV-2, possui uma fita simples de RNA de sentido positivo, que serve como modelo tanto para a replicação quanto para a síntese de proteínas. Ele codifica quatro proteínas estruturais: espícula (do inglês *spike*) (S), envelope (E), membrana (M) e nucleocapsídeo (N), além de diversas proteínas não estruturais e acessórias (Figura 2) (KIM et al., 2020). As quatro proteínas estruturais são responsáveis pela replicação e estruturação viral, ligação do vírus aos receptores celulares, bem como pela patogenicidade do vírus. (GUSEV et al., 2022). Dentre as proteínas estruturais, a proteína S promove a adesão às células hospedeiras através da ligação chave-fechadura com diversos receptores, sendo o principal deles o receptor celular da Enzima Conversora de Angiotensina-2 (do inglês ACE-2). A proteína E é a menor proteína estrutural, composta por 76-109 aminoácidos e altamente conservada entre os coronavírus. Está presente no vírion em menor quantidade que as demais proteínas estruturais e atua desencadeando o processo de montagem da partícula viral e brotamento, funcionando como uma viroporina canalizadora de íons, que auxilia no rompimento das membranas do hospedeiro, permitindo a liberação viral. (GUSEV et al., 2022). O envelope é composto por uma bicamada lipídica e envolve o nucleocapsídeo e nele estão ancoradas as proteínas de espícula (S), membrana (M) e envelope (E). (HELMY et al., 2020; LI et al., 2020). A proteína M é a mais abundante, atuando no transporte

de nutrientes, na montagem das partículas virais, além de determinar o sítio de brotamento do vírus e a forma do envelope viral. A proteína N empacota o genoma viral formando um nucleocapsídeo de formato helicoidal, que desempenha um papel na replicação do genoma viral (RASTOGI et al., 2020). A interação da proteína M com a proteína N estabiliza o nucleocapsídeo. (MALIK, 2020).

Na extremidade 5' do genoma, o RNA apresenta o "Cap". Há, tanto na extremidade 5' quanto na 3', poli-adeninas (Poli-A) e regiões não traduzidas (UTRs, do inglês: *Untranslated regions*) do genoma. Na extremidade 5', em seguida da UTR, o RNA genômico apresenta duas grandes ORFs (do inglês: *Open Reading Frames* - quadros de leitura abertos), a ORF1a e ORF1b, a maioria delas envolvida na transcrição e replicação do vírus, formando um complexo denominado replicase-transcriptase. As duas ORFs compreendem aproximadamente 67% do genoma e codificam as poliproteínas pp1a e pp1ab (ABBASIAN et al., 2023); (SIMAS et al., 2023), (RASTOGI et al., 2020) precursoras de 16 Nsps (proteínas não estruturais) (nsp1 a nsp16), enquanto as demais ORFS codificam proteínas acessórias e estruturais (KUMAR et al., 2020); (GUSEV et al., 2022). A proteína pp1a codificada por ORF1a é clivada proteoliticamente em 10 proteínas não estruturais maduras (NSP1-10) e a proteína pp1ab expressa por ORF1b é processada em 16 Nsps (NSP1-16) (RASTOGI et al., 2020). A porção do genoma próxima à extremidade 3' apresenta ORFs que codificam proteínas estruturais e proteínas acessórias (ORFs 3a, 3b, 3c, 3d, 6, 7a, 7b, 8a, 8b, 9b, 9c e ORF10), que estão distribuídas entre os quatro principais genes de proteínas estruturais (S, E, M e N). Além do RNA genômico, vários mRNAs subgenômicos também são produzidos. (ABBASIAN et al., 2023) (GUSEV et al., 2022). Alguns domínios funcionais e estruturais de Nsps estão bem definidos, sendo NSP3 uma protease semelhante à papaína (PL2pro), a NSP5 uma protease semelhante a 3C (3CLpro ou Mpro - que cliva os polipeptídeos pp1a e pp1ab); a NSP12 uma RNA polimerase RNA-dependente (RdRp) (que em complexo com a NSP7 e NSP8 como cofatores, forma um complexo de replicase para replicação e transcrição de RNA genômico viral), NSP13 é uma enzima helicase (Hel) e NSP14 atua como uma exoribonuclease de revisão (ExoN), sendo estas enzimas críticas que facilitam a replicação e transcrição do RNA viral (KUMAR et al., 2020). A função e a expressão das proteínas acessórias ainda não estão amplamente conhecidas (ABBASIAN et al., 2023) (GUSEV et al., 2022).

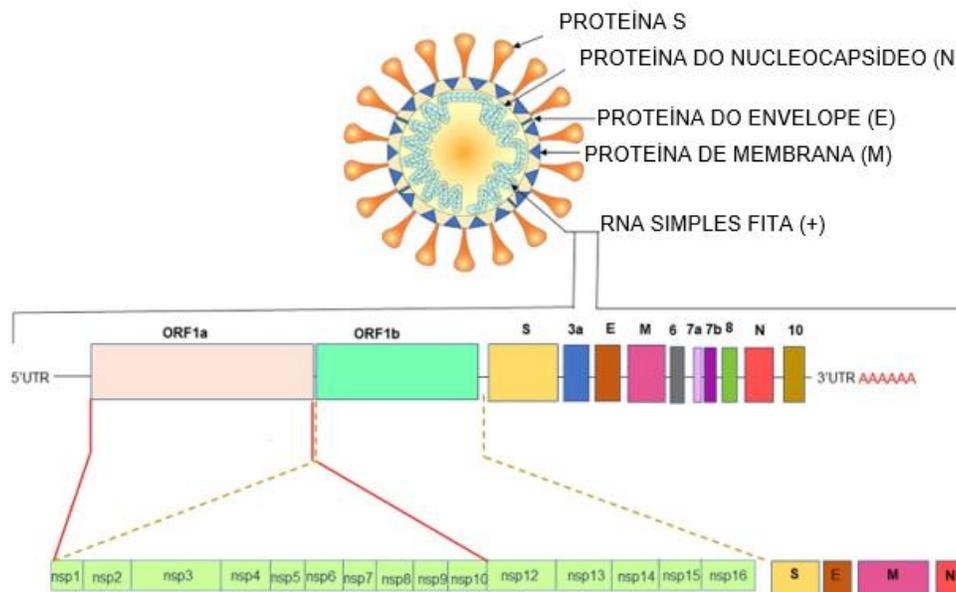


FIGURA 2- Representação esquemática da estrutura e do genoma do SARS-CoV-2, mostrando as proteínas estruturais, não estruturais e acessórias. Observa-se que a região das ORFs1a e 1b corresponde a mais da metade do genoma. Fonte: Figura adaptada de Rastogi et al., 2020.

1.2.3. Ciclo Replicativo

O SARS-CoV-2 invade a célula hospedeira, por meio de uma ligação da proteína S do vírus com o receptor de ACE-2 na célula humana (Figura 3), além da utilização de enzimas do hospedeiro, como a serinoproteases transmembranares do tipo 2 (do inglês TMPRSS2), que auxiliam na entrada do vírus e fusão com a membrana celular ou endossomal (V'KOVSKI et al., 2021). No geral, a entrada do SARS-CoV-2 nas células hospedeiras é um processo complexo e multifatorial, envolvendo muitas moléculas auxiliares no processo (proteínases, co-receptores e ativadores de sua expressão). A presença de co-receptores, em particular, permite que o SARS-CoV-2 infecte células com pouca expressão de ACE-2 nas membranas. Os diversos tropismos do SARS-CoV-2 aos tecidos humanos pode explicar a facilidade de contágio, bem como a invasão viral de órgãos internos durante a progressão da COVID-19 (GUSEV et al., 2022).

A Proteína S é na realidade uma glicoproteína transmembrana, com arranjo

em trímero, localizada na superfície viral, dividida nas regiões S1 (subunidade de ligação ao receptor) e a subunidade S2 que abrange a membrana C-terminal, necessária para a fusão celular (subunidade de fusão de membranas) (JACKSON et al., 2021). A região S1 é exposta à superfície e compreende o domínio N-terminal (NTD) e o domínio de ligação ao receptor (do inglês *Receptor Binding Domain - RBD*), que vai mediar a ligação do vírus ao receptor ACE-2 na célula hospedeira, determinando seu tropismo e patogenicidade. O RBD é, portanto, um domínio de grande importância, por estar relacionado à eficácia da resposta imune, do tratamento com anticorpos monoclonais e das vacinas. A região S2, transmembrana, é composta pelo peptídeo de fusão, que após extensos rearranjos conformacionais medeia a fusão do vírus com a membrana celular (V'KOVSKI et al., 2021), e também por regiões de repetição heptadas HR1 e HR2 (padrões de repetição de sete aminoácidos) e um sítio de clivagem, que é clivado por proteases do hospedeiro, como TMPRSS2 (LU et al., 2020; JACKSON et al., 2021).

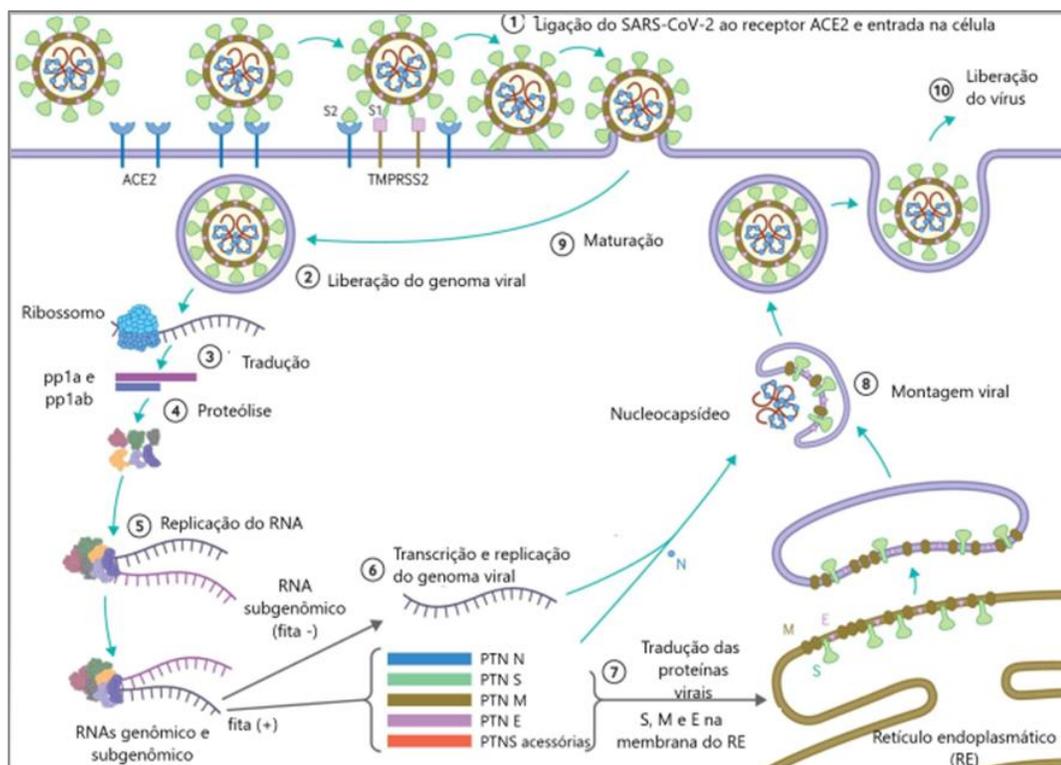


FIGURA 3 – Representação esquemática do ciclo de replicação do SARS-CoV-2. Fonte: adaptado de CEVIK et al., 2020.

Na infecção por SARS-CoV-2, o domínio RBD primeiro se liga ao seu receptor ACE2, para formar um complexo RBD/ACE2. Esta ligação desencadeia mudanças conformacionais na proteína S, levando à fusão da membrana do vírus (COUTARD et al., 2020), mediada por HR1 e HR2. Esse processo culmina na entrada do vírus nas células-alvo humanas. A presença de sítio de clivagem da furina no limite entre S1 e S2 é importante para infecciosidade e virulência. A clivagem da furina facilita a mudança conformacional necessária para a ligação da proteína S ao receptor ACE2 (V'KOVSKI et al., 2021).

Após a entrada na célula, o RNA genômico é liberado e são traduzidas as regiões ORF1a e ORF1b, que resultam nas poliproteínas pp1a e pp1ab (Figura 3). Estas são processadas e transformadas em proteínas não-estruturais individuais (nsps), que formam o complexo de replicação e transcrição viral. Ao mesmo tempo que as nsps são expressas, as organelas de replicação viral, que consistem nas vesículas de membrana dupla perinuclear (DMVs – do inglês *perinuclear double-membrane vesicles*), membranas convolutas (CMs- do inglês *convoluted membranes*) e pequenas esférulas abertas de membrana dupla (DMSs) criam um microambiente protetor para replicação de RNA genômico viral e transcrição de mRNAs subgenômicos (sg mRNAs). As proteínas estruturais traduzidas translocam-se em membranas do retículo endoplasmático (ER) e transitam pelo ER ao compartimento intermediário do complexo de Golgi (ERGIC), onde o RNA genômico associa-se com a proteína N, ambos recém produzidos. (V'KOVSKI et al, 2021). Uma vez montados, os vírus são transportados por vesículas até a membrana da célula hospedeira e liberados por exocitose. Ao ser transportada para a superfície celular, a proteína S permite que as células infectadas e saudáveis sejam fundidas, resultando na formação de grandes células multinucleadas que espalham o vírus no organismo hospedeiro (GUSEV et al., 2022).

1.3 MUTAÇÕES E O SURGIMENTO DE NOVAS VARIANTES

Mutações e alterações no genoma fazem parte do processo normal de replicação e evolução dos vírus, sendo muito mais frequentes em vírus com o genoma RNA, uma vez que são mais susceptíveis à erros de replicação mediados pela RNA polimerase e também em virtude do tamanho significativo da população viral, que possui uma taxa de replicação mais alta. Embora a maioria das mutações

seja neutra e não resulte em alterações favoráveis ao patógeno viral, à medida que a transmissão é contínua e em grande escala, a seleção das mutações que favorecem o "fitness" viral é facilitada e são amplificadas em rodadas subsequentes de infecção na população (GUSEV et.al, 2022).

Os coronavírus e outros membros da ordem Nidovirales possuem os maiores e mais complexos genomas conhecidos entre os vírus de RNA. A replicação do coronavírus faz uso da atividade de revisão da exonuclease do nsp14 para evitar o acúmulo de mutações deletérias. Apesar desta alta fidelidade de replicação única, uma série de mutações no genoma do SARS-CoV-2 foram observadas ao longo do curso atual da pandemia de COVID-19 (JACKSON et al., 2021). Mutações na proteína S capazes de aumentar a afinidade do vírus pelo receptor celular ou a quantidade de vírus produzida durante o curso da infecção podem estar associadas ao aumento da transmissão do vírus (MICHELON, 2021). Devido ao seu papel na ligação do vírus aos receptores da superfície celular do hospedeiro, sabe-se que a proteína S é o alvo principal dos anticorpos neutralizantes gerados após a infecção por SARS-CoV-2 (HARVEY et al., 2021; ABBASIAN et al., 2023), além de ser a proteína cuja produção é induzida pelas vacinas baseadas em mRNA. Desta forma, mutações relacionadas à proteína S que afetem sua antigenicidade são particularmente importantes alvos de estudo e têm um papel importante do ponto de vista clínico-epidemiológico (AL-ZYOUD et al., 2021; ABBASIAN et al., 2023).

Mutações capazes de melhorar a aptidão viral foram detectadas logo nos primeiros meses da pandemia, como por exemplo, a alteração de aminoácidos D614G da proteína S em abril de 2020 (KORBER et al., 2020); (HARVEY et al., 2021), que estabilizou a associação S1/S2. Para HOFFMANN e colaboradores, a estabilidade entre S1 e S2 melhora a eficiência da clivagem entre S1 e S2, modulando a infecção por SARS-CoV-2 e permitindo assim que sejam selecionadas as mutações que aumentaram a clivagem S1/S2. Para PLANTE e colaboradores, a mutação D614G modifica a região RBD para uma conformação "para cima", facilitando a ligação com o receptor ACE2, levando a uma maior infectividade do vírion. Independente do mecanismo de ação, a mutação D614G foi mantida até as linhagens da variante Ômicron observadas no ano de 2023 (ABBASIAN et al., 2023).

A compreensão das regiões do genoma com altas taxas de mutação e regiões conservadas do SARS-CoV-2 torna-se, portanto, de extrema importância,

uma vez que regiões conservadas podem ser consideradas alvos promissores para o desenvolvimento de vacinas, para o desenvolvimento de ferramentas de diagnóstico e para fortalecer a sensibilidade de detecção para reduzir resultados falso-negativos em testes de diagnóstico, como o RT-PCR. O teste, recomendado pela OMS e amplamente aplicado em diversos países para o diagnóstico da COVID-19, pode apresentar resultados falso-negativos que são considerados um desafio importante para o controle da pandemia. Isto porque com as diversas mutações no genoma do SARS-CoV-2, podem ocorrer mudanças nos locais de ligação dos primers e no reconhecimento das sondas. (ABBASIAN et al., 2023).

1.4 LINHAGENS CIRCULANTES NO BRASIL E NO MUNDO

De acordo com a OMS (OMS, 2021), a definição de Variante de Preocupação (VOC) é feita com base em uma avaliação comparativa, sendo aquela que demonstra estar associada a uma ou mais das seguintes alterações em um grau de significância para a saúde pública global: aumento da transmissibilidade ou alteração prejudicial na epidemiologia da COVID-19; ou aumento da virulência ou mudança na apresentação clínica da doença; ou diminuição da eficácia das medidas sociais e de saúde pública ou diagnósticos, vacinas e terapias disponíveis (OMS, 2021). Já a Variante de Interesse (VOI) é considerada como tal se o seu genoma contiver, em comparação com a variante original, mutações que alterem o fenótipo do vírus e se tiver sido identificada como causadora de transmissão comunitária, de múltiplos casos ou de clusters (agrupamentos de casos) de COVID-19 ou tiver sido detectada em vários países; ou ainda se for de outra forma avaliada como uma VOI pela OMS em consulta com o Grupo de Trabalho de Evolução do Vírus SARS-CoV-2 (OMS, 2021).

No Brasil, o primeiro caso confirmado de coronavírus ocorreu em 26 de fevereiro de 2020. Inicialmente, a epidemia foi impulsionada principalmente pelas linhagens B.1.1.28 e B.1.1.33, que foram as mais prevalentes até outubro de 2020 (Figura 4) (MICHELON, 2020; CANDIDO et al. 2020). De acordo com informações disponíveis na plataforma GISAID (*Global Initiative on Sharing Influenza Data* - <https://gisaid.org/>), após outubro, mereceram destaque as variantes P.1 (B.1.1.28.1 -

Gamma) e P.2 (B.1.1.28.2 - Zeta), duas variantes de origem nacional, originadas da linhagem B.1.1.28, que dominaram o cenário durante quase todo o primeiro semestre de 2021 (Figura 4), sendo que a P.1 se tornou a principal variante no território brasileiro a partir de Janeiro de 2021 (GISAID, 2023).

No 2º semestre de 2021, aproximadamente nos meses de agosto e setembro, a VOC Delta passou a prevalecer sobre a Gamma (ROMANO et al., 2022). Em janeiro de 2022, mesmo após grande parte da população brasileira já ter sido imunizada, observou-se um surto da variante de preocupação (VOC) Ômicron, que permaneceu como VOC prevalente durante os anos de 2022 e 2023. Ainda no primeiro semestre de 2023, a VOC Ômicron e suas sublinhagens têm sido as variantes globalmente majoritárias (OMS, 2023).

A VOC Ômicron é composta por várias sublinhagens, sendo que, inicialmente, a BA.1, BA.1.1 (clado 21K) e BA.2 (clado 21L) eram as mais comuns mundialmente. A partir de março de 2023, a proporção de sequências relatadas globalmente designadas como BA.2 foi aumentando em relação a BA.1, o que pôde ser observado também no Brasil no ano de 2022. A partir de maio até o mês de junho de

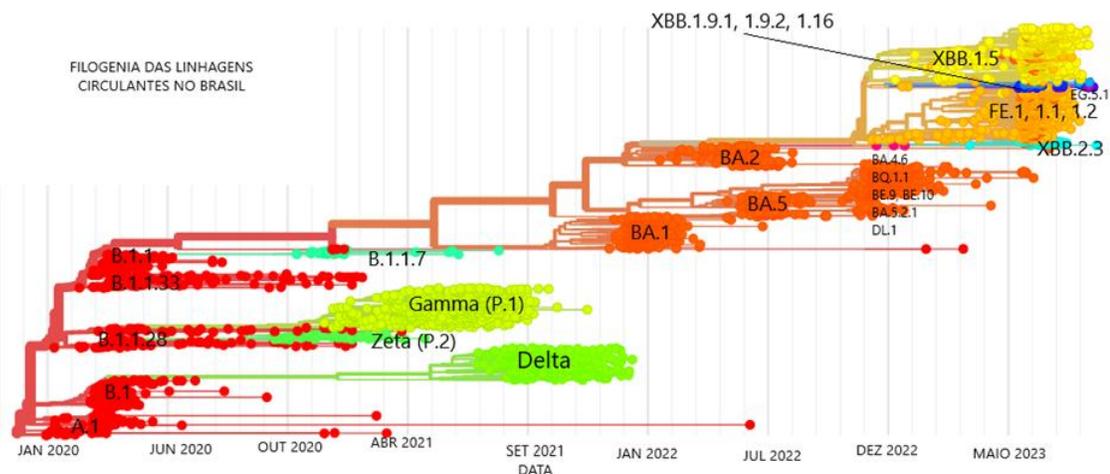


FIGURA 4 – Figura representativa da filogenia da pandemia por coronavírus no Brasil com base em dados do GISAID - 3970 genomas coletados entre Fevereiro de 2020 e Agosto de 2023. Fonte: adaptado de <https://gisaid.org/phylogenetics/brazil/> (acesso em 16/09/2023).

2022, a maioria das amostras corresponderam a linhagem BA.2. De acordo com o Informe da Rede Corona-Ômica BR-MCT– Brasília (13 de dezembro de 2022), os

meses de julho e agosto de 2022 mostraram o aumento na detecção do clado 22B, com destaque para BA.5.2, BA.5.2.1 e BA.5.1. Dados da mesma Rede de outubro e novembro/2022 confirmaram a tendência da circulação das sublinhagens do clado 22B, especialmente a BA.5.3 que foi detectada em ambos os meses. Ao final de outubro de 2022, o Informe da Rede Corona-Ômica BR-MCT– Brasília (13 de dezembro de 2022) relatou que a linhagem BQ.1.1 foi detectada em um paciente com histórico de viagem para a Europa. Tanto a BQ.1 quanto BQ 1.1 que circularam no Brasil no início do ano de 2023 são sublinhagens da linhagem BA.5. (ITO et al., 2023).

Os dados sobre as variantes de circulação no Brasil são constantemente atualizados na página da Rede Genômica Fiocruz, incluindo os genomas sequenciados pela própria Rede e depositados na plataforma GISAID por outras instituições a partir de amostras brasileiras. Os dados sobre as mudanças nas frequências de linhagens e sublinhagens de SARS-CoV-2 podem embasar autoridades sanitárias na tomada de decisões a respeito de medidas de prevenção e controle da doença, ressaltando a importância da constante vigilância genômica do vírus.

A Organização Pan-Americana da Saúde (OPAS/OMS) divulgou, em 23 de agosto de 2023, uma nota técnica com uma atualização sobre as sublinhagens Ômicron recentemente identificadas, devido ao surgimento e aumento de casos globalmente e também em alguns países pan-americanos, incluindo Canadá, Colômbia, Costa Rica, República Dominicana e Estados Unidos, das linhagens EG.5 e BA.2.86 (OPAS, 2023). A linhagem EG.5 é descendente de XBB.1.9.2 (uma sublinhagem de Ômicron), relatada pela primeira vez em fevereiro de 2023. Foi classificada como variante de interesse (VOI) em 9 de agosto de 2023. No entanto, nenhuma mudança na gravidade da doença foi relatada até o momento. A BA.2.86 foi inicialmente relatada em uma amostra colhida na Dinamarca no final de julho de 2023. Desde então, foi detectada em Israel, no Reino Unido e nos Estados Unidos, mas apenas um total de sete sequências foram relatadas. Foi classificada como variante sob monitoramento (VUM) pela OMS em 17 de agosto de 2023 por apresentar um número considerável de mutações no gene S. O risco para a saúde pública representado pelo EG.5 foi avaliado como baixo e é semelhante ao de outras

variantes preocupantes em circulação, conforme relato da OPAS na atualização. Com relação à BA.2.86, as informações são ainda limitadas e a avaliação inicial do risco será gerada posteriormente (OPAS, 2023).

Devido à prevalência mundial da variante Ômicron e suas sublinhagens, a OMS atualizou em 2023 seu sistema de rastreamento e definições de trabalho para variantes do SARS-CoV-2, de forma a avaliar independentemente sublinhagens Ômicron em circulação e classificar novas variantes com mais clareza. Especialistas do Grupo Técnico Consultivo da OMS sobre a Evolução do Vírus SARS-CoV-2 (TAG-VE) chegaram ao consenso de que a variante Ômicron, em comparação com as variantes anteriores, é a variante com mais mutações/divergências até o momento, com evolução genética e antigênica e uma variedade crescente de sublinhagens, com impacto nos testes de detecção, sistemas de tratamento e na epidemiologia da COVID-19 (WHO, 2023).

Desde fevereiro de 2022, a variante Ômicron tem representado mais de 98% das sequências publicamente disponíveis no mundo, sendo então o padrão genético do qual provavelmente surgirão novas variantes de SARS-CoV-2 (WHO, 2022). No entanto, ainda é possível o surgimento de variantes derivadas de COV's circulantes anteriormente ou de variantes completamente novas. O sistema anterior de classificação de VOC's da OMS classificava todas as sublinhagens da Ômicron como parte da VOC Ômicron (WHO, 2023). Desta forma, não havia a precisão e detalhamento necessários para comparar novas linhagens descendentes, com fenótipos alterados, com as linhagens parentais da Ômicron BA.1, BA.2, BA.4/BA.5. A partir de 15 de março de 2023, o sistema de rastreamento de variantes da OMS passou a considerar a classificação de sublinhagens Ômicron independentemente, como variantes sob monitoramento (VUMs), VOIs ou VOCs (WHO, 2023). Periodicamente, a OMS publica um relatório de atualização onde indica as variantes de interesse (VOI's) e sob monitoramento (VUM's).

1.5 TECNOLOGIAS DE SEQUENCIAMENTO DO GENOMA COMPLETO

Embora o sequenciamento de DNA e RNA tenha uma história que se estende por cinco décadas, o sequenciamento massivo paralelo em grande escala, ou

sequenciamento de nova geração (NGS), só tornou-se disponível comercialmente há menos de duas décadas (MCCOMBIE et al., 2019). NGS é o nome dado ao conjunto das mais modernas técnicas para a identificação das sequências de ácidos nucleicos (DNA e RNA) das mais variadas amostras biológicas, que se sobrepuseram ao modelo de Sanger após o Projeto Genoma Humano e que surgiram da necessidade de se produzir sequenciamentos mais rápidos, econômicos e de qualidade (SCHUSTER, 2008). Diferentes estratégias baseadas em NGS têm sido utilizadas com sucesso no rastreamento de origens e também na compreensão da propagação, transmissão e evolução dos agentes infecciosos. A utilização de métodos de sequenciamento eficientes e rápidos para reconstruir a sequência genômica do SARS-CoV-2, agente etiológico da COVID-19, tem sido fundamental para o desenho de testes moleculares de diagnóstico e elaborar medidas e estratégias eficazes para mitigar a difusão da pandemia (CHIARA et al., 2021).

A maioria das plataformas de sequenciamento massivo paralelo baseia-se no conceito de sequenciamento por síntese, ou seja, os métodos permitem a incorporação de nucleotídeos utilizando-se de uma variedade de enzimas e esquemas de detecção que permitem que a plataforma do instrumento correspondente colete dados em sincronia com a síntese enzimática. Geralmente o DNA é fragmentado em trechos de, em média, 150 a 300 pares de base, sequenciado e em seguida as leituras (“reads”) são remontadas. A grande diferença entre as tecnologias está no tamanho dos fragmentos, no mecanismo de detecção da incorporação de nucleotídeos e a forma como as sequências são lidas (MCCOMBIE et al., 2019).

Dentre as principais plataformas de sequenciamento, destacamos o primeiro método de sequenciamento de segunda geração lançado pela plataforma 454 Life Sciences, por pirosequenciamento. Diferente do método de Sanger, que utilizava a tecnologia com nucleotídeos marcados (dideoxinucleotídeos), essa tecnologia utiliza a detecção de pirofosfato (um subproduto da incorporação de nucleotídeos), que ao ser incorporado no DNA de interesse, emite luz. Ao ser detectada, ocorre a identificação do fragmento sequenciado. Esta tecnologia foi substituída por outras que surgiram em seguida (RONAGHI et al., 1998). Já o Ion Torrent é uma plataforma de sequenciamento considerada como uma transição entre a segunda e a terceira

geração de sequenciadores, lançada em 2010. Diferente do método utilizado por Sanger e Illumina (que utilizam nucleotídeos marcados com fluorocromos para detecção da fita sintetizada no sequenciamento), a tecnologia utiliza como base o sequenciamento em semi-condutor. O processo de sequenciamento é realizado em um *chip*, que possui em sua superfície micro-sensores de pH, que detectam o processo de amplificação do DNA a partir da liberação de íons H⁺ que são liberados no processo de incorporação de base (RUSK, 2011).

Outra tecnologia de NGS, o Illumina, é a plataforma mais popular de sequenciamento de nova geração pois é capaz de realizar grandes sequenciamentos com alta qualidade de “*reads*” por um valor mais acessível. Ela utiliza o método de sequenciamento por síntese. O processo identifica simultaneamente os nucleotídeos enquanto os incorpora em uma cadeia de ácido nucléico. Os nucleotídeos, modificados quimicamente, se ligam à fita molde de DNA por complementaridade. Eles possuem um marcador fluorescente e um terminador reversível, que bloqueia a incorporação de uma próxima base. O sinal fluorescente indica qual nucleotídeo foi adicionado e o terminador é clivado para que a próxima base possa se ligar e assim completar o sequenciamento do fragmento. À medida que o equipamento reconhece o nucleotídeo, vai se formando a sequência do fragmento que está sendo estudado (METZKER, 2010).

No ano de 2009, a PacBio, lançado pela Pacific Biosciences, fez parte da terceira geração de sequenciadores de DNA, revolucionando o método de sequenciamento pelo modo de operação em tempo real, conhecido por um processo de sigla SMRT (*Single-molecule real time* - observação em tempo real dos nucleotídeos adicionados no fragmento presente no sequenciamento), método que até então nenhum tipo de sequenciador era capaz de realizar. Após a fragmentação do DNA, as extremidades de ambas as fitas dos fragmentos são unidas, formando uma única fita circular com auto-pareamento. A circularização permite que um mesmo fragmento de DNA seja lido por completo nas duas fitas. O processo de sequenciamento é realizado em nano-poços, sendo em cada um fixada uma enzima DNA polimerase. A cada ciclo, uma base diferente é adicionada pelo equipamento. No casos em que há incorporação de bases, um fluoróforo é liberado e os sensores conseguem detectar em tempo real a reação da enzima e a velocidade com a qual ela realiza a incorporação (cinética enzimática), permitindo a detecção de alterações,

como metilações em bases específicas, visto que a afinidade da enzima é alterada neste casos (METZKER, 2020; LIU et al., 2012).

Outra ferramenta de sequenciamento paralelo massivo de terceira geração muito utilizada para sequenciar o genoma completo de SARS-CoV-2 é o MinION, tendo sido, inclusive, utilizado para sequenciar o primeiro genoma completo do vírus descrito no Brasil, no início da pandemia. Este equipamento foi lançado em 2014 pela Oxford Nanopore Technologies (ONT), foi o primeiro sequenciador comercial usando tecnologia de nanoporos. O princípio da tecnologia baseia-se na medição de mudanças na condutividade elétrica gerada à medida que as fitas de DNA passam por um poro biológico, identificando, assim, as bases de DNA. Seu tamanho reduzido, portabilidade, acessibilidade e velocidade na produção de dados o tornam adequado para aplicações em tempo real. O dispositivo pode ainda ser usado como uma rápida ferramenta de diagnóstico para gestão de surtos de várias doenças devido à rápida recuperação dos dados, sendo interessante em ações de emprego nas áreas de biossegurança, bioproteção e biodefesa. Na pandemia por COVID-19, os pesquisadores têm utilizado o sequenciamento de nanoporos para gerar dados que vão desde a caracterização inicial do genoma do vírus SARS-CoV-2 até a rápida identificação de variantes (PATER et al., 2021).

Os tipos de NGS continuarão a evoluir, proporcionando novas oportunidades de pesquisa e aplicações em uma variedade de campos, desde a medicina até a biologia evolutiva e a conservação da biodiversidade. À medida que essas tecnologias se tornam mais acessíveis e poderosas, elas desempenham um papel fundamental na compreensão da genética e na resolução de questões complexas relacionadas à biologia e à saúde (CHIARA et al., 2021).

1.6 A VIGILÂNCIA GENÔMICA E SUA IMPORTÂNCIA PARA A BIODEFESA

Desde o início da pandemia, o compartilhamento de sequências de SARS-CoV-2 em plataformas de dados como a GISAID (*Global Initiative on Sharing All Influenza Data*) tem auxiliado na identificação de variantes emergentes, na epidemiologia e definição de políticas de saúde pública no Brasil e no mundo. A capacidade do vírus de evoluir e disseminar de forma rápida torna o rastreamento

genômico e a análise de variantes ferramentas importantes na contenção e resposta à pandemia. A alta incidência e mortalidade pela doença ocorridas no Brasil indicam a grande circulação do vírus em especial nos anos de 2020 a 2022, conseqüentemente, favorecendo a ocorrência de altas taxas de mutações e possibilidade do surgimento de novas variantes (LAMARCA et al., 2022). Portanto, a vigilância genômica, mediante o sequenciamento e análises do genoma completo do SARS-CoV-2, tem papel fundamental no rastreamento dessas mutações, ampliando o conhecimento sobre evolução e diversidade genética, e contribuindo para a melhor eficácia de tratamentos e vacinas (PATTER et al., 2021; WHERRY et al., 2022).

. A rápida identificação de substituições, inserções ou deleções de nucleotídeos e aminoácidos que venham a impactar nas propriedades virais é também fundamental no monitoramento da pandemia (DA SILVA et al., 2021; HARVEY et al., 2021). O acompanhamento das mutações do vírus ao longo do tempo pode auxiliar no rastreamento da propagação do patógeno e facilitar a compreensão das dinâmicas da transmissão. A análise filogenética permite a reconstrução da história evolutiva do vírus (OMS, 2021). Diante deste cenário, o Exército Brasileiro (EB) vem estimulando a capacitação de seu pessoal e instalações físicas, para rapidamente identificar outros agentes biológicos, além do SARS-CoV-2, de interesse à Defesa Biológica. O domínio e incorporação de tecnologias de sequenciamento de nova geração (NGS), especialmente aquelas de tamanho reduzido e que apresentem portabilidade e velocidade rápida de produção de dados para aplicações em tempo real, como a Plataforma de sequenciamento por Nanoporos, é certamente importante, por ser uma rápida ferramenta de diagnóstico para gestão de surtos de várias doenças, sendo de fundamental importância em ações de emprego nas áreas de biossegurança, bioproteção e biodefesa.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Realizar a análise genômica do SARS-CoV-2, em amostras positivas para COVID-19 diagnosticadas por RT-PCR no Instituto de Biologia do Exército (IBEx), bem como em genomas completos do estado do Rio de Janeiro compartilhados na plataforma GISAID, para identificar as variantes e as alterações gênicas relevantes presentes, contribuindo com o monitoramento e o maior conhecimento do processo evolutivo do vírus.

2.2 OBJETIVOS ESPECÍFICOS

- Obter um total de 20 sequências de genomas completos de SARS-CoV-2, utilizando plataformas de sequenciamento de nova geração, a partir de amostras positivas diagnosticadas por RT-PCR no IBEx ;
- Analisar sequências genômicas completas de SARS-CoV-2 a partir do GISAID, de períodos concomitantes aos da amostragem do IBEx e demais períodos relevantes (“ondas”) da pandemia;
- Comparar os genomas completos obtidos no IBEx e no GISAID, por meio de alinhamento com a sequência do genoma referência de Wuhan NC_045512.2, identificando as variantes presentes, mutações, inserções e deleções relevantes.

3 MATERIAL E MÉTODOS

3.1 REGULAMENTAÇÃO ÉTICA

Os estudos desenvolvidos neste trabalho foram previamente aprovados pelo Comitê de Ética em Pesquisa do Centro de Capacitação Física do Exército (CCFEx), sob o código, CAAE 58840522.4.0000.9433. A participação nos estudos teve caráter voluntário e todos os participantes assinaram o Termo de Consentimento Livre e Esclarecido (ANEXO A).

3.2. AMOSTRAGEM

Para a atividade de sequenciamento de genomas completos, foram selecionadas aleatoriamente amostras de swab nasofaríngeo de pacientes, usuários do Fundo de Saúde do Exército (FUSEx). Esta população era composta por militares e seus dependentes que realizaram o teste de RT-PCR em tempo real e testaram positivo para COVID-19 no IBEx. As amostras foram provenientes de indivíduos com diferentes faixas etárias, em atendimentos ambulatoriais ou hospitalares, em diferentes localidades no Estado do Rio de Janeiro (Quadro 2). As coletas foram realizadas nos Postos de Coleta do IBEx no Rio de Janeiro (Vila Militar, Policlínica Militar da Praia Vermelha, Policlínica Militar do Rio de Janeiro e Policlínica Militar de Niterói) e de coletas realizadas em hospitais militares (Hospital Central do Exército, Hospital Geral do Rio de Janeiro, Hospital Militar de Resende) entre os anos de 2020 e 2021.

QUADRO 2 – Informações das amostras sequenciadas – Abril a Junho de 2020 + amostra Nr 20 (Outubro de 2021).

Índice	Nº AMOSTRA	DATA DA COLETA	IDADE	GÊNERO	ORIGEM	CT
1	58301	1/6/2020	9	FEM	RJ (IBEX)	13,80
2	59977	9/6/2020	9	FEM	RJ (IBEX)	15,70
3	57817	27/5/2020	44	MASC	RJ (IBEX)	16,30
4	58306	1/6/2020	43	FEM	RJ (IBEX)	18,70

(continua)

(continuação)

Índice	Nº AMOSTRA	DATA DA COLETA	IDADE	GÊNERO	ORIGEM	CT
5	59980	9/6/2020	7	MASC	RJ (IBEX)	18,20
6	32317	30/4/2020	82	FEM	RJ (HGERJ)	17,10
7	58271	1/6/2020	62	MASC	RJ (IBEX)	15,00
8	11784	29/4/2020	45	MASC	NITERÓI (PoMN)	15,40
9	57722	27/5/2020	3	FEM	RJ (IBEX)	33,90
10	1019	15/4/2020	36	FEM	RJ (HCE)	14,70
11	11785	29/4/2020	86	FEM	NITERÓI (PoMN)	11,60
12	11786	29/4/2020	19	MASC	NITERÓI (PoMN)	14,80
13	1640	29/4/2020	50	FEM	RJ (HCE)	14,00
14	2215	17/4/2020	31	MASC	RJ (HGERJ)	17,00
15	2517	29/5/2020	52	MASC	RJ (HCE)	14,50
16	32321	30/4/2020	71	FEM	RJ (HGERJ)	17,80
17	57437	25/5/2020	35	FEM	RJ (IBEX)	20,40
18	57721	27/5/2020	35	FEM	RJ (IBEX)	28,90
19	9339	1/6/2020	23	MASC	RESENDE (HMR)	21,80
20	7842 (Barcode 11)	11/10/21	84	FEM	RJ (HCE)	24,48

As amostras diagnosticadas positivas utilizadas no estudo foram identificadas e armazenadas à temperatura de - 80°C até o momento da reextração do RNA para o sequenciamento. Após a extração, foi feita uma nova reação de RT-PCR em tempo real com o objetivo de que fossem utilizadas preferencialmente amostras com Ct (*cycle threshold*) entre 18 e 35, para a melhor eficiência no sequenciamento de acordo com o protocolo escolhido.

Foram sequenciados com boa cobertura e elevado escore de qualidade 20 genomas completos de SARS-CoV-2 de amostras do IBEx. Destas, 19 amostras foram coletadas entre 15 de abril e 09 de junho de 2020, e sequenciadas no Sistema Ion 5 Gene Studio S5, do Laboratório de Metabolismo Macromolecular Firmino Torres de Castro, chefiado pela Dra Rosane Silva, como parte de uma colaboração para a vigilância da COVID-19 no Rio de Janeiro entre o IBEx e o grupo de pesquisadores da UFRJ e, inclusive, já foi publicado um artigo baseado na análise de 5 genomas (SIMAS et al., 2023). O equipamento utiliza a tecnologia de

semicondutores, onde é medida a mudança de pH ocorrida sempre que é adicionado um dNTP a um polímero de DNA.

A amostra coletada em 11 de outubro de 2021 (7842/2021) foi sequenciada no próprio IBEx, no sequenciador de nova geração MinION Mk1B (ONT), que utiliza a tecnologia de nanoporos para o sequenciamento. Outras 24 amostras coletadas em 2021 e 2022 foram sequenciadas por esta metodologia entretanto não foi possível obter boa cobertura e bom escore de qualidade para compor a presente análise deste trabalho.

3.3. SELEÇÃO DE GENOMAS NO GISAIID

Para ampliar o presente estudo, além das amostras positivas coletadas no IBEx, foram obtidas sequências genômicas de SARS-CoV-2 do Estado do Rio de Janeiro no principal repositório mundial de sequências, o GISAIID, aplicando-se os filtros “*complete*” e “*high coverage*”. Por definição da Plataforma GISAIID (<https://gisaid.org/>), as sequências consideradas completas foram aquelas com genomas com tamanho >29,000 pares de bases de comprimento. Com relação à cobertura, as sequências consideradas com alta cobertura foram as que apresentavam menos que 1% de nucleotídeos não determinados e menos que 0.05% de mutações únicas de aminoácidos (<https://gisaid.org/>) (SHU et al., 2017). Selecionamos genomas dos mesmos períodos de isolamento das amostras do IBEx, com o objetivo de compararmos a diversidade genômica entre elas (2020 – 1ª onda da pandemia e 2021 – entre a 4ª e 5ª ondas), além de cobrir períodos importantes da pandemia, como a 5ª onda de COVID-19 no Estado, causada pela chegada da variante Ômicron, onde houve aumento significativo no número de casos de COVID-19. O 1º semestre de 2023 foi também avaliado, de forma a obter um panorama mais atual das linhagens circulantes e mutações.

3.4 EXTRAÇÃO DE RNA

As extrações de RNA das amostras positivas, armazenadas a -80°C no IBEx, foram todas realizadas por metodologia de “beads” magnéticas, utilizando o kit de

extração Extracta Kit Fast DNA e RNA viral, seguindo as instruções do fabricante Locus, no equipamento de DNA e RNA (Extracta 32 – Locus®), em processo automatizado. A extração baseou-se no princípio de adsorção, onde a camada de silício que reveste as *beads* magnéticas, adsorve partículas carregadas negativamente, promovendo sua ligação com as moléculas de ácidos nucleicos presentes na amostra. Neste processo a separação ocorre por hastes magnetizadas com alto poder de atração em etapas sucessivas de lavagem e eluição.

3.5 REALIZAÇÃO DE RT-PCR EM TEMPO REAL

A reação de RT-PCR em tempo real da amostra sequenciada no IBEx foi feita no Laboratório de Biologia Molecular, utilizando o Kit BioMol One Step, IBMP, que detecta regiões-alvo nos genes do nucleocapsídeo (N) e da Open Read Frame OFR1ab. Um terceiro alvo é detectado no gene da RNaseP humana, para controle interno da reação, de acordo com o *Center of Disease Control and Prevention* (CDC, Atlanta, GA), utilizando o equipamento Step One Plus. São consideradas positivas as amostras que apresentam amplificação acima do “*threshold*” para os dois alvos virais, bem como para o controle interno, com $Ct \leq 37$. (MARTY et al., 2020).

3.6 PREPARO DA BIBLIOTECA GENÔMICA E SEQUENCIAMENTO DE AMOSTRAS COLETADAS NO IBEx

A amostra coletada e diagnosticada no IBEx, referente ao ano de 2021 (7842/2021) foi sequenciada no sequenciador de nova geração MinION Mk1B (Nanopore). Cada corrida de sequenciamento foi feita com a biblioteca de amostras preparada de acordo com o Protocolo de com o protocolo de Josh Quick et al. 2019, descrito pelo grupo do Projeto ARTIC Network nCoV-2019 (<https://artic.network/2-protocols.html>), que tem como objetivo desenvolver sistemas para processar amostras de surtos virais para gerar informações epidemiológicas em tempo real que sejam interpretáveis e acionáveis por órgãos de saúde pública (<https://artic.network/1-about.html>).

A partir de 11 μ L de amostra (previamente diluída em água estéril de acordo com os valores de Ct (sem diluição para valores de Ct entre 18 e 35; diluição 1:10

para valores de Ct entre 15 e 17; 1:100 para valores de Ct entre 12 e 14), foi realizada uma transcrição reversa usando a enzima SuperScript IV Reverse Transcriptase (SSIV, *Invitrogen by Life Technologies, Carlsbad, EUA*), bem como primers não específicos, para o preparo do DNA complementar (cDNA), de acordo os parâmetros de ciclagem e instruções do fabricante. Para a obtenção do DNA dupla fita foi utilizado o conjunto de primers de PCR multiplex v4, (198 oligos 100µM diluídos previamente 1:10 – ANEXO B) separados em dois pools (v4 Pools 1 e 2), juntamente com o mix para PCR *Q5 Hot Start High Fidelity Mastermix 2X (New England Biolabs)* e água nuclease free. O conjunto de primers v4 é específico para o genoma do SARS Cov2, desenvolvido de forma a cobrir todo o genoma (<https://artic.network/ncov-2019>). Para cada um dos pools, 2,5 µL de cDNA sintetizado na etapa anterior foi usado como fita molde. Nesta reação foram gerados amplicons de 400 pb, lado a lado, com sobreposições de vinte pb (não incluindo primers), sendo empregado 1 ciclo para a ativação e extensão final e 30 ciclos para as demais etapas, com seguintes parâmetros: ativação: 98°C, 30"; desnaturação: 98°C, 15 "; anelamento: 65°C, 5'; extensão: 72°C, 10"; extensão final: 72°C, 10' e final 4°C. Após a reação de PCR, os amplicons gerados foram quantificados no Fluorômetro Quantus™, utilizando-se o kit Quantifluor® dsDNA System (Promega), de acordo com as instruções do fabricante. A etapa de quantificação foi realizada para que fosse possível definir as os volumes de Pool 1 e de Pool 2 de cada amostra a serem associados em um só tubo, para seguir para as etapas seguintes (para amostras com 100ng de DNA foram utilizados 5µL do Pool; entre 101 e 149ng, 4µL, acima de 200ng as amostras foram diluídas para chegar nos intervalos citados). Após a associação dos Pools, o volume de cada amostra foi completado com água nuclease q.s.p. 50µL e as amostras foram tratadas com o Mix de Reparo (*New England Biolabs - Ultra II End Prep Reaction Buffer, Ultra II End Prep Enzyme Mix e água nuclease free*), com o objetivo de criar extremidades compatíveis nos amplicons para receber os códigos de barras (*barcodes*). Na reação de reparo, é adicionada uma cauda de Poli Adeninas (cauda "Poli A"), e as enzimas de reparo das terminações são inativadas (PATTER et al., 2021). Os "*barcodes*" são sequências de nucleotídeos previamente conhecidas que servem para a identificação das amostras, uma vez que o sequenciamento se dá por meio de um pool contendo todas amostras. As amostras foram multiplexadas usando os kits de

código de barras *Native Barcoding Expansion* da ONT (EXP-NBD104 e EXP-NBD114), seguido de etapas de limpeza com *beads* magnéticas *Pronex Magnetic beads* (Promega) e quantificação no Fluorômetro *Quantus* (Promega).

A biblioteca foi preparada usando *Ligation Sequencing Kit* SQK-LSK109, da ONT. Mais etapas de limpeza com *beads* magnéticas foram realizadas, para a remoção do excesso de adaptadores de sequenciamento. O DNA presente no pool de amostras foi novamente quantificado, da mesma forma que nas etapas anteriores. Uma quantidade total de no mínimo 15 ng desta biblioteca final foi sequenciada em uma célula de fluxo FLO-MIN106 (R9.4.1), multiplexando até vinte e quatro amostras por execução. Todos os processos de preparo da biblioteca e sequenciamento ocorreram nos Laboratórios de Biologia Molecular e de Genética do Instituto de Biologia do Exército. O sequenciamento foi acompanhado por intermédio do Software MinKNOW até que o número de dados suficientes tivesse sido coletado (em torno de 12 a 18 horas). A chamada de bases (“*basecalling*”), a demultiplexação e a remoção dos adaptadores das sequências foram realizados utilizando o programa MinKNOW (ONT), o qual detecta o sinal bruto proveniente do MinION, avalia a mudança de sinal com a passagem do DNA pelo nanoporo e fornece arquivos nos formatos Fasta e Fastq com informações das bases, as quais podem ser observadas em tempo real.

O sequenciamento dos 19 genomas feito pelo grupo de pesquisadores da UFRJ foi feito utilizando o Ion AmpliSeq SARS-CoV-2 Research Panel (ThermoFisher Scientific, Carlsbad, CA, EUA), por meio de dois conjuntos de pools de primers, assim como no IBEx, que cobrem o genoma do vírus. A reação de amplificação multiplex foi realizada usando 10 μ L do cDNA de acordo com as instruções do fabricante para 21 ciclos dos primers SARS-CoV-2 específicos para RT-PCR multiplex do painel (SIMAS et al., 2023). A quantificação das bibliotecas de cada amostra foi obtida usando o kit de quantificação Ion Library TaqMan™ (ThermoFisher Scientific, Carlsbad, CA, EUA). O conjunto de bibliotecas em uma concentração de 50 pM foi executado no sistema Ion Chef™ (ThermoFisher Scientific, Carlsbad, CA, EUA) usando os kits Ion 510, 520 e 530 (ThermoFisher Scientific, Carlsbad, CA) (SIMAS et al., 2023). As leituras geradas foram mapeadas para um genoma de referência SARS-CoV-2 Wuhan (NCBI GenBank ID MN908947) usando o software Ion Browser incluído no Torrent Suite 5.10.1 e visualizadas no Integrative Genomic Viewer (IGV 2.6.3) (Broad Institute, Cambridge, MA, EUA). O sequenciamento NGS

gerou arquivos em formato fasta, que foram posteriormente avaliados usando a ferramenta online Nextclade (desenvolvedores Nextstrain).

3.7 ANÁLISE DA QUALIDADE DE LEITURA E QUALIDADE DAS SEQUÊNCIAS

Os genomas sequenciados pelo grupo de pesquisadores da UFRJ e aqueles obtidos no GISAID já se encontravam com análise de qualidade da leitura realizada.

Para as amostras sequenciadas no MinION, foram utilizadas apenas as sequências que passaram no filtro de qualidade (fastq_pass) do MinKNOW. Os arquivos no formato Fastq foram convertidos em formato Fasta. As sequências convertidas em formato Fasta foram avaliadas quanto à qualidade de acordo com o Protocolo “FastqQC + Artic QC + Nextclade” utilizando-se o software Epi2MeDesktopAgent vs 3.5.6., onde foram fornecidos dados de qualidade da sequência (“AVG quality score”), número de “reads”, rendimento, bem como a atribuição dos clados, qualidade das sequências, classificação da linhagem e mutações (<https://labs.epi2me.io/quality-scores/>). A ferramenta Nextclade, disponível nas formas integrada ao protocolo do Epi2Me e também online (<https://clades.nextstrain.org/>) foi empregada para a avaliação das sequências IBEx e das obtidas no GISAID. Seu algoritmo utilizou-se de regras individuais de controle de qualidade, que combinadas, fizeram parte do cálculo geral de qualidade (“Overall QC score”). Dentre os índices de qualidade, foram calculados automaticamente e individualmente escores para Dados Ausentes (N ou “Missing Data”), Locais Mistos (M ou “Mixed Sites”), Mutações Privadas (P ou “Private Mutations”), Clusters de Mutação (C ou “Mutation Clusters”), Códon de Parada (S ou “Stop códons”) e Mudança de quadro de leitura (F ou “Frameshifts”). Ao final, foi calculado o escore geral de qualidade (“Boa”: de 0 a 29; “Mediana”: de 30 a 99; “Ruim”: ≥ 100) para cada sequência.

3.8. ALINHAMENTO, CHAMADA DE VARIANTES, ATRIBUIÇÃO DE CLADOS E ANÁLISE DE POLIMORFISMOS

Os arquivos em formato FASTA, tanto das amostras do IBEx quanto das obtidas no GISAID-RJ de cada período, avaliados previamente segundo o

protocolo FastQC + Protocolo Fastq QC + ARTIC + NextClade (aplicativo Epi2Me Desktop Agent) foram agrupados utilizando-se o software BioEdit vs 7.2.5. (Figura 5). O alinhamento ao genoma de referência de Wuhan-Hu-1 (NC_045512.2) foi realizado no software MAFFT (*Multiple alignment program for amino acid or nucleotide sequences*) vs 7 (<https://mafft.cbrc.jp/alignment/server/>).

Após o alinhamento, a chamada de mutação/variantes (processo pelo qual identificamos as variantes a partir dos dados das sequências), atribuição de clados, as verificações de qualidade foram feitas por meio da ferramenta online *Nextclade* (Figura 6), que possibilita que os dados sejam baixados no formato de planilha do Microsoft Excel. A partir de arquivos com múltiplas sequências alinhadas, foram analisados os polimorfismos de nucleotídeo único (*Single nucleotide polymorphisms* - SNP's), de forma a identificar variações nos genomas. Também foi utilizada a ferramenta *Base By Base*, disponível na página eletrônica <https://4virology.net/virology-ca-tools/base-by-base/> do *Viral Bioinformatics Research Centre* (VBRC), para análise de polimorfismos tais como SNP's, e inserções e deleções (INDEL's).

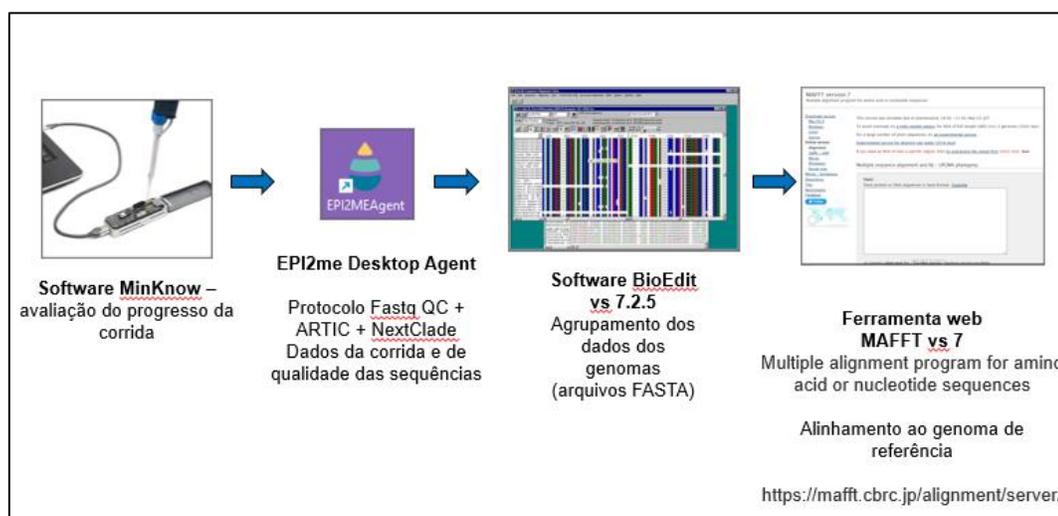


FIGURA 5 - Etapas e ferramentas utilizadas para a avaliação da qualidade das sequências, visualização e agrupamento dos genomas e alinhamento ao genoma referência.

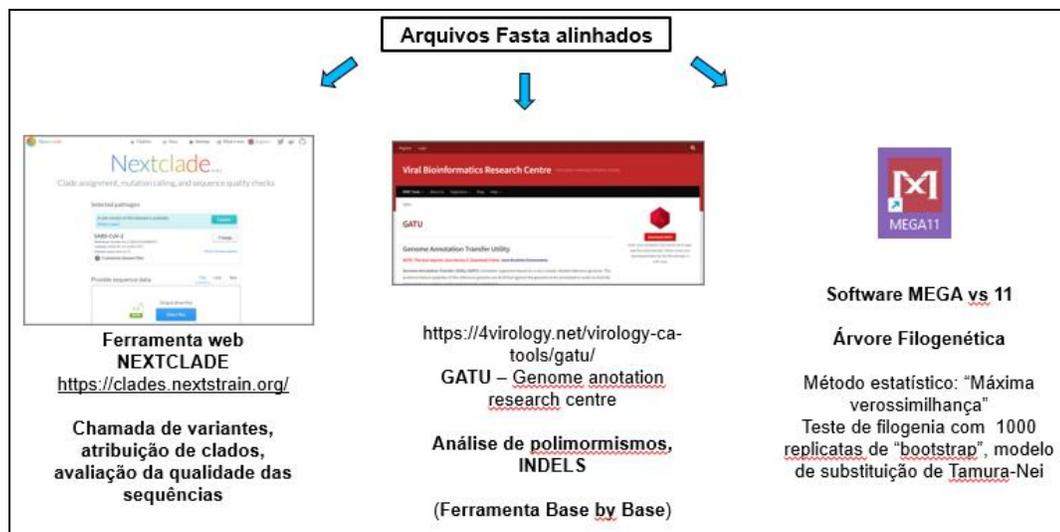


FIGURA 6 - Ferramentas e softwares utilizados para a chamada de variantes, atribuição de clados, análise de polimorfismos e análise filogenética.

3.9 ANÁLISE FILOGENÉTICA

As árvores filogenéticas foram construídas pelo método estatístico de "Máxima verossimilhança", com teste de filogenia com 1000 replicatas de "bootstrap", e modelo de substituição Tamura-Nei, no software MEGA 11. Foram construídas duas árvores filogenéticas, sendo a primeira composta pelos 20 genomas das amostras sequenciadas e genomas GISAID-RJ de Janeiro de 2020 (representando o início da pandemia) e de Outubro de 2021 (representando o período da amostra sequenciada), além do genoma referência de Wuhan-Hu-1 (NC_045512.2), evidenciando as distâncias entre os genomas destes períodos. A segunda árvore foi construída utilizando-se os 20 genomas sequenciados 2020/2021, o genoma referência os genomas das linhagens mais frequentes nos anos de 2021 (P.1 e AY.99.2), 2022 (BA.1, BA.1.1, AY.99.1 e AY.99.2) e 2023 (XBB.1.5, BQ.1.1 e XBB.1.5.15).

4. RESULTADOS

4.1. UNIVERSO DE SEQUÊNCIAS ANALISADAS

Em nosso trabalho, foram avaliadas 1521 sequências (Quadro 3). Neste número estão representadas apenas as que obtiveram escores de qualidade considerados bons pela plataforma interativa Nextclade (<https://clades.nextstrain.org/>), que realiza atribuição de clados, chamada de mutação e verificações de qualidade de sequência (AKASAMENOV et al., 2021). Os genomas que obtiveram os escores de qualidade global “ruim” ou “mediano”, segundo classificação gerada pela plataforma Nextclade e detalhada no item 3.7, foram desconsiderados.

QUADRO 3 – Quadro informativo da quantidade de amostras sequenciadas e de genomas obtidos no GISAID para análise genômica. .

Fonte	Quantidade de amostras/genomas	Período de coleta da amostra
IBEx 2020	19 amostras	Abril a Junho de 2020 (1ª onda)
GISAID 2020	298 genomas	Abril a Junho de 2020 (1ª onda)
IBEx 2021	01 amostra	Outubro de 2021 (4ª onda)
GISAID 2021	788 genomas	Outubro de 2021 (4ª onda)
GISAID 2021/2022	158 genomas	Dezembro de 2021 a Janeiro de 2022 (5ª onda)
GISAID 2023	257 genomas	Janeiro a Junho de 2023
TOTAL	1521	

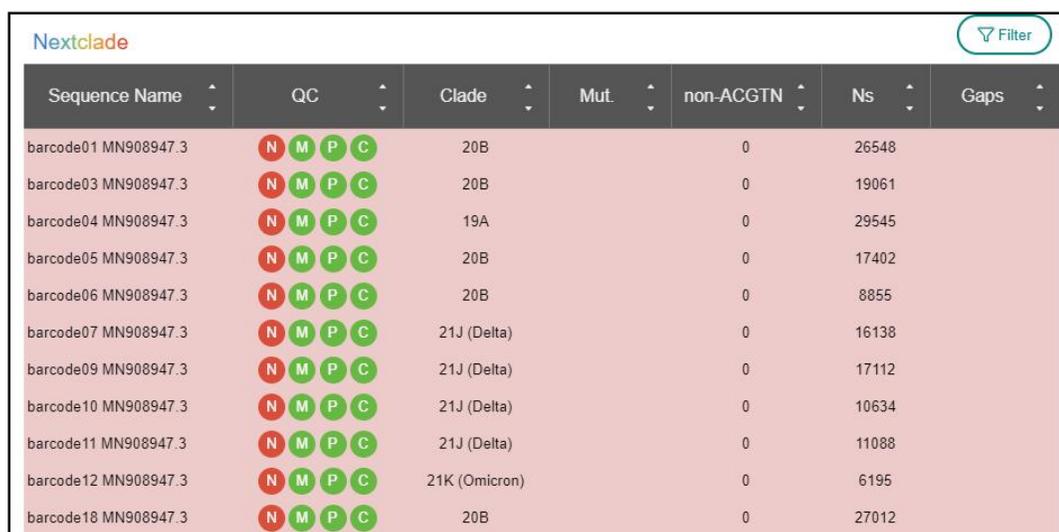
4.2 QUALIDADE DAS SEQUÊNCIAS E COBERTURA DOS GENOMAS

Todas as 20 amostras sequenciadas apresentaram status geral de qualidade (“qc overall status”) bom e cobertura do genoma acima de 97%, conforme mostra a Quadro 4.

QUADRO 4- Informações de qualidade e cobertura das amostras sequenciadas.

Índice	Amostra	Linhagem atribuída pelo Nextclade (Pango)	Clado	Status geral de qualidade (“qc.overall status”)	Cobertura (%)	Total de substituições	Total de inserções de aminoácidos	Quantidade de Mudanças de quadro de leitura e região de ocorrência
01	58301/2020	B.1.1.33	20B	bom	99,79	13	0	(1) ORF7a:104-122
02	59977/2020	B.1.1.33	20B	bom	99,79	14	0	(1) ORF7a:104-122
03	57817/2020	B.1.1.33	20B	bom	99,77	14	0	(1) ORF7a:104-122
04	58306/2020	B.1.1.33	20B	bom	99,77	12	0	(1) ORF7a:104-122
05	59980/2020	B.1.1.33	20B	bom	99,77	14	0	(1) ORF7a:104-122
06	32317/2020	B.1.1.33	20B	bom	99,79	10	0	-----
07	58271/2020	B.1.1.33	20B	bom	99,81	12	0	-----
08	11784/2020	B.1.1.33	20B	bom	99,83	14	0	-----
09	57722/2020	B.1.1.33	20B	bom	99,77	12	0	-----
10	1019/2020	B.1.1.33	20B	bom	99,78	15	0	-----
11	11785/2020	B.1.1.33	20B	bom	99,78	11	0	-----
12	11786/2020	B.1.1.33	20B	bom	99,78	14	0	-----
13	1640/2020	B.1.1.33	20B	bom	99,75	11	0	-----
14	2215/2020	B.1.1.33	20B	bom	99,84	9	0	-----
15	2517/2020	B.1.1.33	20B	bom	99,81	18	0	-----
16	32321/2020	B.1.1.33	20B	bom	99,81	14	0	-----
17	57437/2020	B.1.1.33	20B	bom	99,8	17	0	-----
18	57721/2020	B.1.1.33	20B	bom	99,78	15	0	-----
19	9339/2020	B.1.1.28	20B	bom	99,85	17	0	-----
20	7842/2021	AY.99.2	21J (Delta)	bom	97,08	44	0	-----

Outras 24 amostras coletadas em 2021 e 2022 foram sequenciadas no IBEx pela tecnologia de sequenciamento por Nanoporos, entretanto não foi possível obter sequências completas (com tamanho >29,000 pares de bases de comprimento), com boa cobertura (com menos que 1% de nucleotídeos não determinados e menos que 0.05% de mutações únicas de aminoácidos) e consequentemente com bom escore de qualidade para compor a presente análise (Figura 7). Este fato deve-se, provavelmente, a fatores tais como a possível degradação do RNA, (instável a oscilações de temperatura que podem ter ocorrido na manipulação ou pelo longo tempo de armazenamento), baixa amplificação de amostras nas etapas envolvendo pool A e B (onde há amplificação com primers específicos para o SARS-CoV-2 para os dois lados da fita de DNA), e também pelo fato da célula de fluxo apresentar contagem de poros próximo ao mínimo indicado (o ideal, segundo o fabricante, é a presença ≥ 800 poros ativos).



Sequence Name	QC	Clade	Mut.	non-ACGTN	Ns	Gaps
barcode01 MN908947.3	N M P C	20B		0	26548	
barcode03 MN908947.3	N M P C	20B		0	19061	
barcode04 MN908947.3	N M P C	19A		0	29545	
barcode05 MN908947.3	N M P C	20B		0	17402	
barcode06 MN908947.3	N M P C	20B		0	8855	
barcode07 MN908947.3	N M P C	21J (Delta)		0	16138	
barcode09 MN908947.3	N M P C	21J (Delta)		0	17112	
barcode10 MN908947.3	N M P C	21J (Delta)		0	10634	
barcode11 MN908947.3	N M P C	21J (Delta)		0	11088	
barcode12 MN908947.3	N M P C	21K (Omicron)		0	6195	
barcode18 MN908947.3	N M P C	20B		0	27012	

FIGURA 7- Figura representativa dos dados de qualidade de alguns dos 24 genomas sequenciados em 07/11/2022 (corrida 361592). A letra "N" em vermelho indica Status geral de qualidade ruim- muitos dados perdidos ("Overall QC status: bad/Missing Data: bad"). Na coluna "Ns" podem ser vistos os números de dados ausentes

Fonte: Epi2Me Desktop Agent (Nanopore).

Ainda assim, para grande parte das amostras, foi possível, mesmo com baixa cobertura, a atribuição de clados pelo Nextclade com as regiões sequenciadas, como mostrado da Figura 7.

Da mesma forma que nas amostras sequenciadas, desconsideramos na análise das sequências GISAID RJ de 2020 a 2023 o total de 130 genomas completos e com alta qualidade, embora a cobertura fosse satisfatória, por apresentarem escores gerais de qualidade “ruim” ou “mediano” segundo o Nextclade. Ainda em relação à qualidade dos genomas, observamos um genoma da variante XZ coletado na UFRJ em Janeiro de 2022, início da circulação de Ômicron (genoma Nr 89- RJ-COVID-19-UFRJ-62795/2022-01-17|EPI_ISL_16754208|2022-01-17), que apresentou cobertura de 99,67% e escores individuais de qualidade bons para Dados Ausentes (N ou “*Missing Data*”), Locais Mistos (M ou “*Mixed Sites*”), Clusters de Mutação (C ou “*Mutation Clusters*”), Códon de Parada (S ou “*Stop códons*”) e Mudança de quadro de leitura (F ou “*Frameshifts*”), mas não para Mutações Privadas (P ou “*Private Mutations*”) Embora o escore geral de qualidade tenha sido “ruim” devido à pontuação alta para “Mutações privadas”, optamos por considerá-lo na análise, uma vez que pela definição de “Mutações privadas” do Nextclade (mutações que representam diferenças para o nó mais próximo na árvore de referência), a pontuação ruim pode ocorrer quando o genoma refere-se a uma linhagem recombinante ainda não designada (Nextclade, 2023). Mutações privadas podem ainda apresentar escore ruim quando houver boas sequências, sem representação na árvore de referência ou quando a sequência é ruim e está contaminada com outra variante ou referência (opção que foi descartada, devido aos resultados obtidos nos demais escores individuais de qualidade). As mutações privadas do genoma XZ recombinante serão mostradas no item 4.4.2 (Quadro 13).

Os dados de qualidade do sequenciamento realizado no IBEx (amostra coletada em 11/10/21-Nº20-7842/2021) são mostrados nas Figuras 7 e 8. Neste sequenciamento, foram analisadas 18.470 "reads". Os valores de escore médio de qualidade acima de 7 são considerados bons no programa Epi2Me Agent, da Nanopore (Figuras 9 e 10), indicando boa qualidade das leituras.

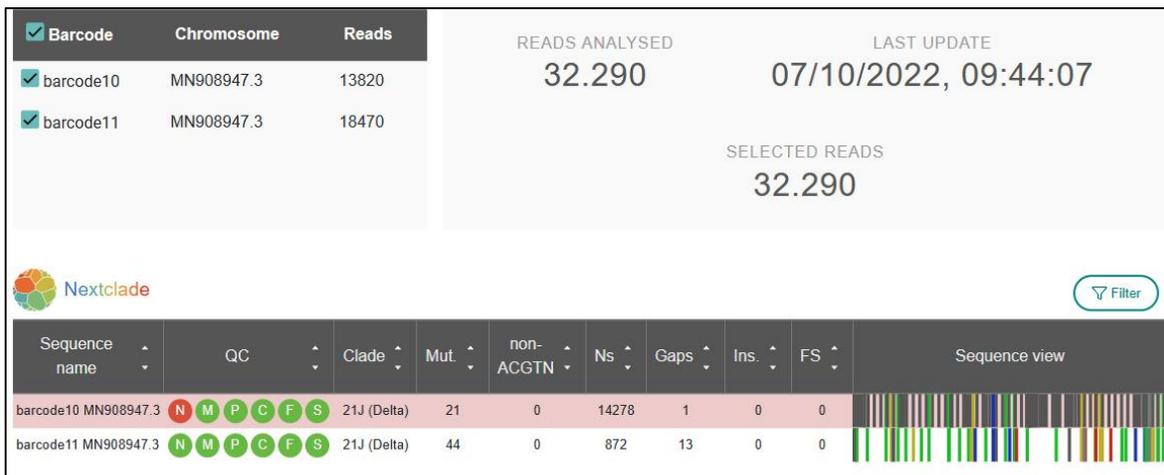


FIGURA 8- Figura representativa dos dados de qualidade do sequenciamento da amostra "Barcode11" (Nº20-7842/2021 IBEx), Corrida 355787 sequenciada em 07/10/2022, onde "N"- Dados ausentes, "M": Locais mistos, "P": Mutações privadas, "C": Clusters de Mutação, "F": Mudanças de quadro, "S": Stop códon. A sinalização em verde indica boa qualidade. Em "Sequence view" é fornecida uma visão geral do genoma, com os dados perdidos ou não identificados (preto), as mutações (em cores) e o restante do genoma sequenciado (branco).
Fonte: Epi2Me Desktop Agent (Nanopore).



FIGURA 9- Figura representativa dos dados dos escores de qualidade das leituras da amostra "Barcode11" (Nº20-7842/2021 IBEx), Corrida 355787 sequenciada em 07/10/2022.
Fonte: Epi2Me Desktop Agent (Nanopore).

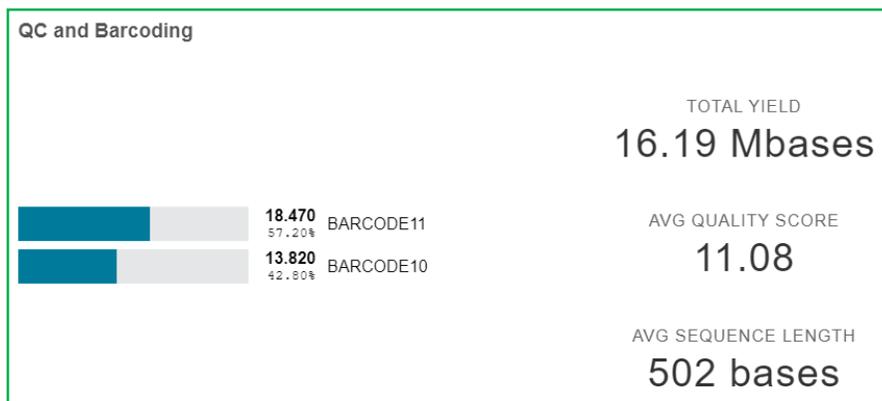


FIGURA 10- Figura representativa dos dados do escore médio de qualidade de leitura da amostra “Barcode11” (Nº20-7842/2021 IBEx), Corrida 355787 sequenciada em 07/10/2022, com o rendimento total, o escore de qualidade e o comprimento das “reads”.

Fonte: Epi2Me Desktop Agent (Nanopore).

A Figura 11 mostra a cobertura obtida para os Pools de primers A e B da amostra “Barcode11” (Nº20-7842/2021 IBEx), evidenciando que o Pool A de primers obteve cobertura maior que o Pool B e que ambos foram capazes de cobrir em torno de 100 vezes todo o genoma do SARS-CoV-2.



FIGURA 11- Figura representativa dos dados de cobertura dos Pools de Primers A e B da amostra “Barcode11” (Nº20-7842/2021 IBEx),

Fonte: Corrida 355787 - Epi2Me Desktop Agent (Nanopore)

4.3. FREQUÊNCIA DE VARIANTES

As frequências de variantes/linhagens observadas durante os anos de 2020 a 2023, incluindo amostras sequenciadas e genomas GISAID RJ são mostradas na Figura 12 e na Quadro 5. De acordo com o gráfico, observa-se o aumento da diversidade de linhagens ao longo do período em estudo. Em 2020, circulavam no Estado do Rio de Janeiro entre Abril e Junho 7 linhagens do vírus, com prevalência (90,85%) da linhagem B.1.1.33 (linhagem sem denominação OMS). Em Outubro de 2021, havia 17 linhagens, sendo 88,07% AY.99.2 (denominação OMS: Delta). Entre Dezembro de 2021 e Janeiro de 2022, o número de linhagens circulantes passou a 26, iniciando-se a queda da frequência da linhagem mais prevalente AY.99.2 para 36%, devido à entrada da variante BA.1 (denominação OMS: Ômicron) no Estado. Entre Janeiro e Junho de 2023, foram observadas 36 diferentes linhagens, todas da variante Ômicron, única VOC detectada neste último período. As frequências individuais em cada período avaliado são mostradas nas Figuras 13,14, 15 e 17.

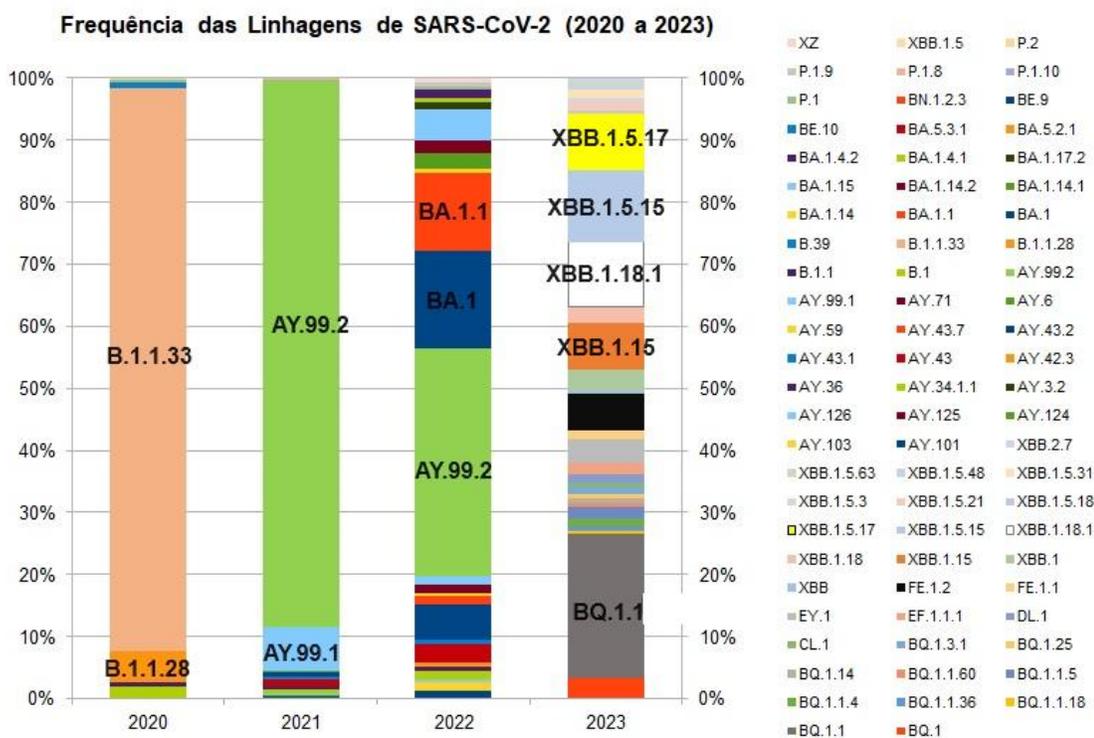


FIGURA 12 – Frequência de linhagens de SARS-CoV-2 em genomas sequenciados no IBEx e obtidos no GISAID durante os anos de 2020 a 2023, com destaque para as principais linhagens

QUADRO 5 - Frequência de linhagens de SARS-CoV-2 em genomas sequenciados no IBEx e obtidos no GISAID durante os anos de 2020 a 2023, evidenciando as linhagens de maior frequência (amarelo) ao longo dos anos. .

Linhagem	Frequência da linhagem/ano (%)			
	2020	2021	2022	2023
AY.101	----	0,38	1,27	----
AY.103	----	----	1,27	----
AY.124	----	0,13	----	----
AY.125	----	0,13	----	----
AY.126	----	0,13	0,63	----
AY.3.2	----	0,13	----	----
AY.34.1.1	----	0,51	1,27	----
AY.36	----	0,63	0,63	----
AY.42.3	----	----	0,63	----
AY.43	----	1,14	3,16	----
AY.43.1	----	0,38	0,63	----
AY.43.2	----	0,63	5,70	----
AY.43.7	----	0,13	1,27	----
AY.59	----	----	0,63	----
AY.6	----	0,13	----	----
AY.71	----	----	1,27	----
AY.99.1	----	6,98	1,27	----
AY.99.2	----	88,07	36,71	----
B.1	1,89	----	----	----
B.1.1	0,63	----	----	----
B.1.1.28	5,05	----	----	----
B.1.1.33	90,85	----	----	----
B.39	0,95	----	----	----
BA.1	----	----	15,82	----
BA.1.1	----	----	12,66	----
BA.1.14	----	----	0,63	----
BA.1.14.1	----	----	2,53	----
BA.1.14.2	----	----	1,90	----
BA.1.15	----	----	5,06	----
BA.1.17.2	----	----	1,27	----
BA.1.4.1	----	----	0,63	----
BA.1.4.2	----	----	1,27	----
BA.5.2.1	----	----	----	0,39
BA.5.3.1	----	----	----	0,39

(continua)

(continuação)

Linhagem	Frequência da linhagem/ano (%)			
	2020	2021	2022	2023
BE.10	----	----	----	2,33
BE.9	----	----	----	5,06
BN.1.2.3	----	----	----	0,39
BQ.1	----	----	----	1,95
BQ.1.1	----	----	----	14,01
BQ.1.1.18	----	----	----	0,39
BQ.1.1.36	----	----	----	0,39
BQ.1.1.4	----	----	----	0,78
BQ.1.1.5	----	----	----	1,17
BQ.1.1.60	----	----	----	0,39
BQ.1.14	----	----	----	0,39
BQ.1.25	----	----	----	0,39
BQ.1.3.1	----	----	----	0,78
CL.1	----	----	----	0,39
DL.1	----	----	----	0,78
EF.1.1.1	----	----	----	1,17
EY.1	----	----	----	2,33
FE.1.1	----	----	----	0,78
FE.1.2	----	----	----	3,50
P.1	0,32	0,25	0,63	
P.1.10	----	0,13	----	----
P.1.8	----	0,13	----	----
P.1.9	----	----	----	----
P.2	0,32			
XBB	----	----	----	0,39
XBB.1	----	----	----	1,95
XBB.1.15	----	----	----	4,67
XBB.1.18	----	----	----	1,56
XBB.1.18.1	----	----	----	6,23
XBB.1.5	----	----	----	31,52
XBB.1.5.15	----	----	----	7,00
XBB.1.5.17	----	----	----	5,45
XBB.1.5.18	----	----	----	0,39
XBB.1.5.21	----	----	----	0,78
XBB.1.5.3	----	----	----	0,39
XBB.1.5.31	----	----	----	0,78
XBB.1.5.48	----	----	----	0,39
XBB.1.5.63	----	----	----	0,39
XBB.2.7	----	----	----	0,39
XZ			0,63	

Dentre os 19 genomas sequenciados do ano de 2020, 18 correspondem a linhagem B.1.1.33 (94,74% das amostras), e apenas 01 à linhagem B.1.1.28 (5,26%). Já no universo total de genomas analisados, ou seja, incluindo as sequências obtidas no GISAID 2020, foi observado, além das duas primeiras, a presença das linhagens P.2 (0,32%), P.1 (0,32%), B.1 (1,89%), B.1.1 (0,63%) e B.39 (0,95%), todas circulantes no Estado do Rio de Janeiro neste período do ano de 2020 (Figura 13).

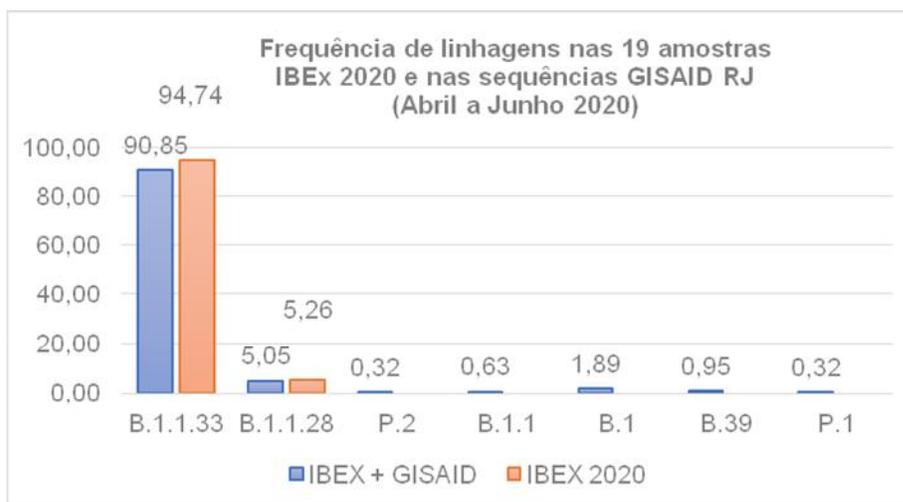


FIGURA 13 - Frequência de linhagens nas 19 amostras IBEx (laranja) e no universo total (azul) nos meses de Abril e Junho de 2020.

No ano de 2021 (Figura 14), a frequência de linhagens no período em estudo obtida por meio de genomas GISAID foi de 88,08% para AY.99.2 (denominação OMS: Delta), seguida de AY.99.1 (6,98%-Delta), ambas do clado 21J, e de outras linhagens do mesmo clado. A variante P.1 (denominação OMS Gamma) foi frequente apenas em 0,51% dos genomas de Outubro de 2021 (linhagens P.1-0,25%; P.1.8-0,13% e P.1.10-0,13%). A amostra 7842/2021 do IBEx, sequenciada pela tecnologia de nanoporos, pertence à linhagem AY.99.2 (variante Delta, clado 21J), que foi a mais frequente no Estado do Rio de Janeiro no período. Considerando-se que o universo de 2021 compreende apenas um mês, observa-se um proporcionalmente número maior de genomas depositados no GISAID (788) em relação ao ano de 2020 (298 genomas entre abril e junho).

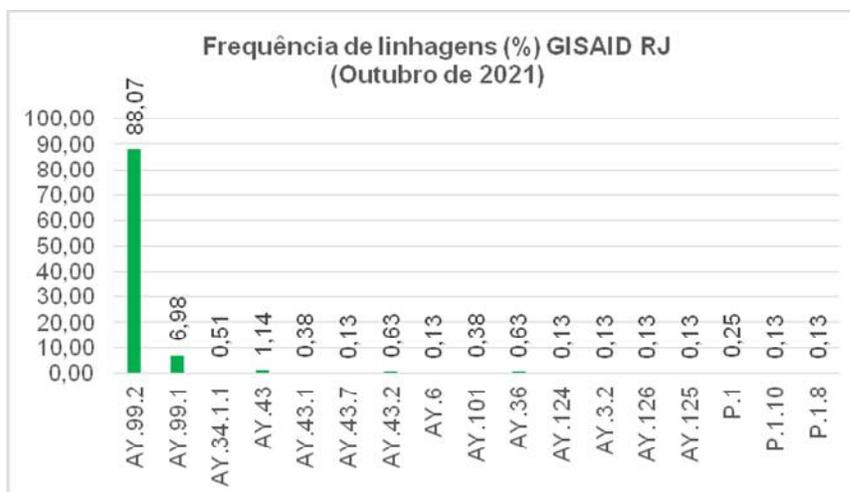


FIGURA 14 - Frequência de linhagens na amostra IBEx e nos 788 genomas GISAID do Estado do Rio de Janeiro. Amostras coletadas entre 1º e 31 de Outubro de 2021.

Em Novembro de 2021, com a introdução da variante BA.1 (Denominação OMS: Ômicron) 2021 no Brasil, observou-se a redução da frequência da VOC AY.99.2, que predominava no Estado (Figura 15). Observa-se, entre Dezembro de 2021 e Janeiro de 2022, ou seja, logo nos primeiros dois meses, uma rápida disseminação de BA.1 e BA.1.1 e, ainda, que 41,77% dos genomas sequenciados (Figura 16) pertencentes à linhagens da variante Ômicron A Figura 16 agrupa as linhagens das variantes Ômicron (BA.1, BA.1.1, BA.1.4.1, BA.1.4.2, BA1.14, BA.1.14.1, BA.1.15 E BA.1.17.2), Delta (AY.99.2, AY.99.1, AY.71, AY.59, AY.43.7, AY.43.2, AY.43.1, AY.43, AY.42.3, AY.36, AY.34.1.1, AY.126, AY.36, Y.34.1.1, AY126, AY.103 e AY.101) e Gamma (P.1 e P.1.9), evidenciando suas frequências.

A frequência da linhagem recombinante XZ, que exibe semelhanças significativas com BA.2” (NIVEDITHA et al., 2023) foi de 0,63% (apenas 1 genoma).

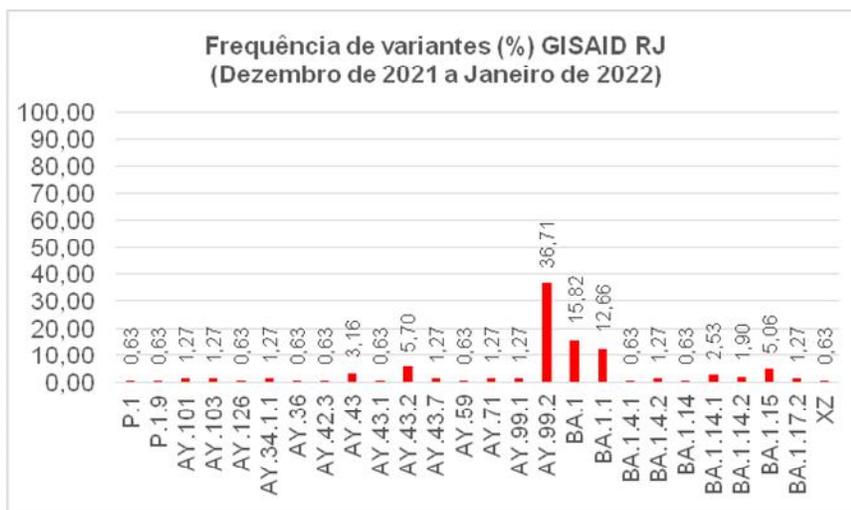


FIGURA 15 - Frequência de variantes nos 158 genomas GISAID do Estado do Rio de Janeiro. Amostras coletadas entre 1º de Dezembro e 31 de Janeiro de 2022, com linhagens das variantes Gamma, Delta e Ômicron.

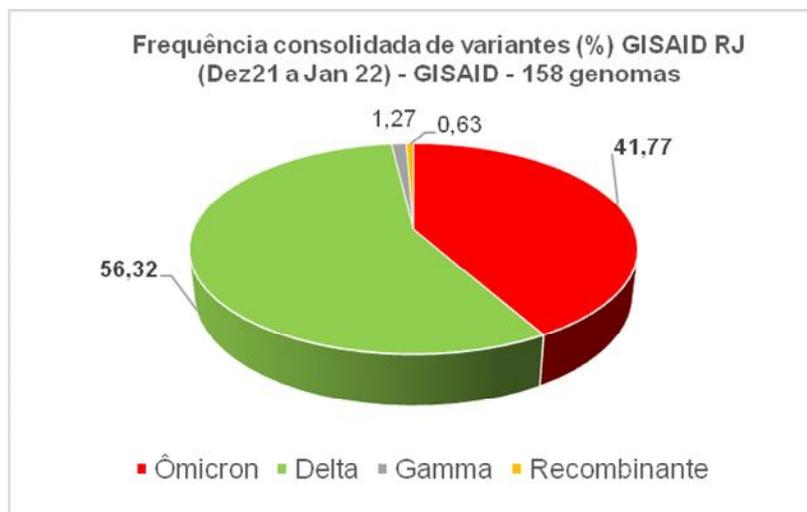


FIGURA 16 – Dados consolidados, mostrando as frequências das diversas linhagens dos genomas GISAID RJ agrupadas por tipo de variante. Em poucos meses, a variante Ômicron foi capaz de disseminar a reduzir a frequência de Delta, que predominou por grande parte do ano de 2021.

No ano de 2023 (Figura 17), as três linhagens de maior prevalência foram XBB.1.5 (31,52%, denominada de “kraken”, sendo uma recombinação de sublinhagens da Ômicron, BA.2.10.1 e BA.2.75.), BQ.1.1 (14,01%, descendente de BA.5) e XBB.1.5.15 (7,00%). Comparando-se as linhagens da variante Ômicron do final de 2021 até o ano de 2023 (Figura 18), observa-se o surgimento de novas linhagens e a existência de linhagens recombinantes de Ômicron, como as “XBB”.

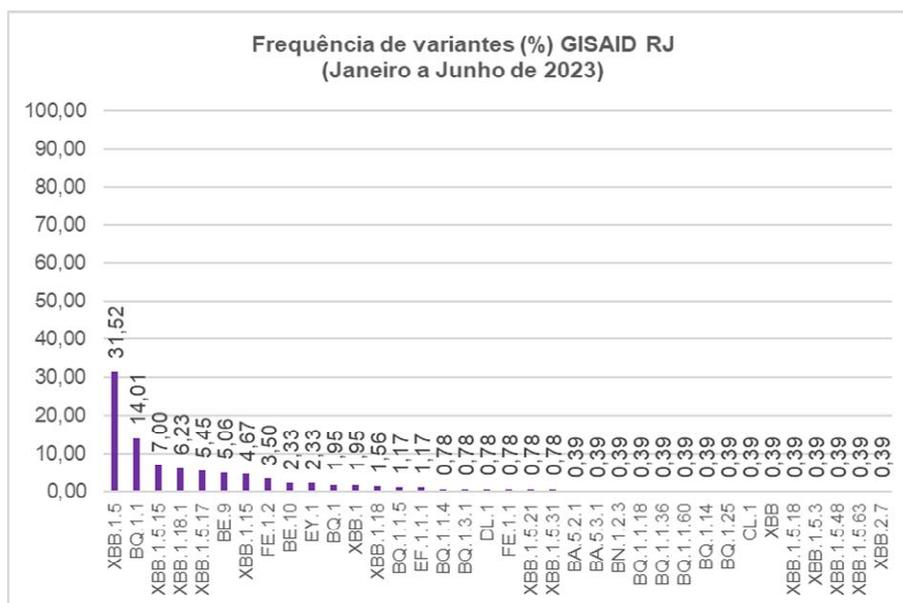


FIGURA 17 – Frequência de variantes nos 257 genomas GISAID do Estado do Rio de Janeiro. Amostras coletadas entre 1º de Janeiro e 30 de Junho de 2023.

4.4. SUBSTITUIÇÕES DE NUCLEOTÍDEOS E DE AMINOÁCIDOS

4.4.1. No universo de amostras sequenciadas

Quando comparadas ao genoma referência, foram observadas 177 substituições de aminoácidos nas 20 amostras sequenciadas, sendo 144 nas amostras de 2020 e 33 em apenas 1 amostra de 2021 (Nº 20), demonstrando o maior acúmulo de mutações na amostra de 2021, como seria esperado. As substituições de aminoácidos são mostradas na Quadro 6.

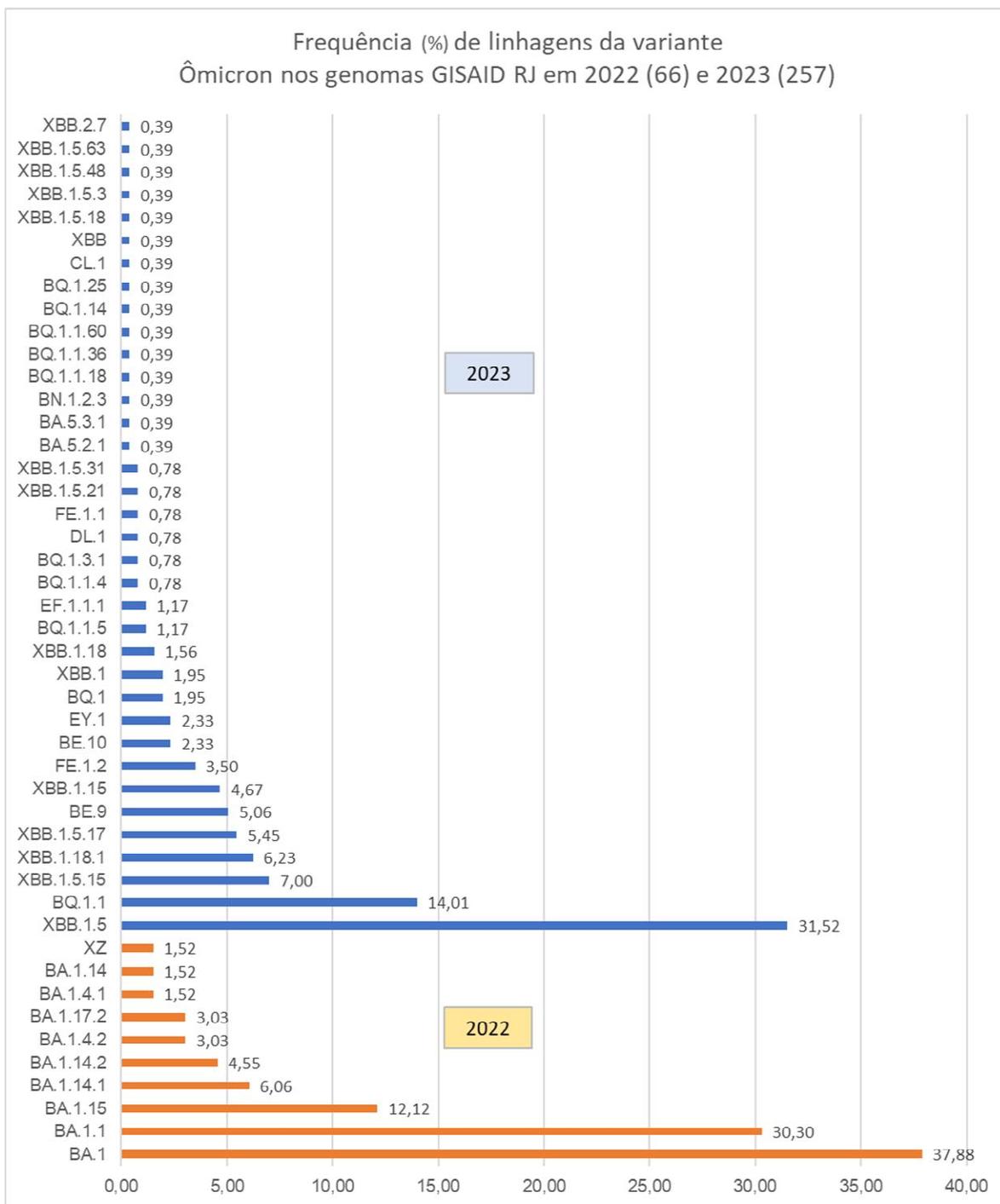


FIGURA 18 – Frequência de variantes nos 66 genomas Ômicron 2021/2022 (laranja) e nos 257 genomas da variante em 2023 (azul), GISAID do Estado do Rio de Janeiro, evidenciando a diferença entre as linhagens iniciais a as circulantes no 1º semestre de 2023.

QUADRO 6 - Substituições de aminoácidos nas amostras sequenciadas em 2020 (1 a 19) e 2021(20), por região do genoma.

	Índice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
	ID	58301/2020	59977/2020	57817/2020	58306/2020	59980/2020	32317/2020	58271/2020	11784/2020	57722/2020	1019/2020	11785/2020	11786/2020	1640/2020	2215/2020	2517/2020	32321/2020	57437/2020	57721/2020	9339/2020	7842/2021		
ORF1a	H417Y																						
	G728C																						
	T1055I																						
	I1091V																						
	G1155C																						
	T1250I																						
	H1160Y																						
	Y1285H																						
	V1295F																						
	A1306S																						
	A1462S																						
	D1600E																						
	P2046L																						
	P2287S																						
	V2930L																						
	T3255I																						
	R3542C																						
	T3646A																						
	V3751A																						
	I393OF																						
T4087I																							
ORF1b	T267M																						
	P314L																						
	I601G																						
	G662S																						
	L829I																						
	P1000L																						
	D1381Y																						
	P1570S																						
	T1621I																						
	A1918V																						
	A1959V																						
	A2431V																						
	T2453I																						

(continua)

(continuação)

	Índice	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
	ID	58301/2020	59977/2020	57817/2020	58306/2020	59980/2020	32317/2020	58271/2020	11784/2020	57722/2020	1019/2020	11785/2020	11786/2020	1640/2020	2215/2020	2517/2020	32321/2020	57437/2020	57721/2020	9339/2020	7842/2021	
S	T19R																					
	S98F																					
	G142D																					
	R158G																					
	A348S																					
	L452R																					
	T478K																					
	D614G																					
	P681R																					
	D950N																					
	K1191N																					
V1776F																						
ORF3a	T14I																					
	S26L																					
	S74F																					
	R134L																					
	T151I																					
	S171L																					
	G172V																					
M	I82T																					
ORF6	I33T																					
ORF7a	S44L																					
	V82A																					
	T120I																					
ORF7b	T40I																					
ORF9b	T60A																					
N	Q9L																					
	D63G																					
	R203M																					
	R203K																					
	G204R																					
	G215C																					
	G243C																					
	I292T																					
D377Y																						

No ano de 2020 não foram observadas substituições nos Genes E, M e nas ORFs 7b, 8 e 9b. A amostra de 2021 não apresentou substituições em E e ORF8. As principais substituições de nucleotídeos e aminoácidos desde período são mostradas na Quadro 7. Os SNP'S, inserções e deleções nas amostras sequenciadas são mostrados na Figura 19.

QUADRO 7 - Frequências das principais substituições de nucleotídeos e aminoácidos nas 19 amostras sequenciadas de 2020. A amostra de 2021 apresentou em comum as substituições C241T, C14408T(P314L) e A23403G(D614G).(em negrito)

Substituição	Frequência	Região/Gene	AA substituído
C241T	100%	5'UTR	N/A
C3037T	100%	ORF1a	N/A
C14408T	100%	ORF1b	P314L
A23403G	100%	S	D614G
T27299C	95%	ORF6	I33T
G28883C	100%	N	G204R
G28881A	100%	N	R203K
G28882A	100%	N	G204R
T29148C	90%	N	I292T

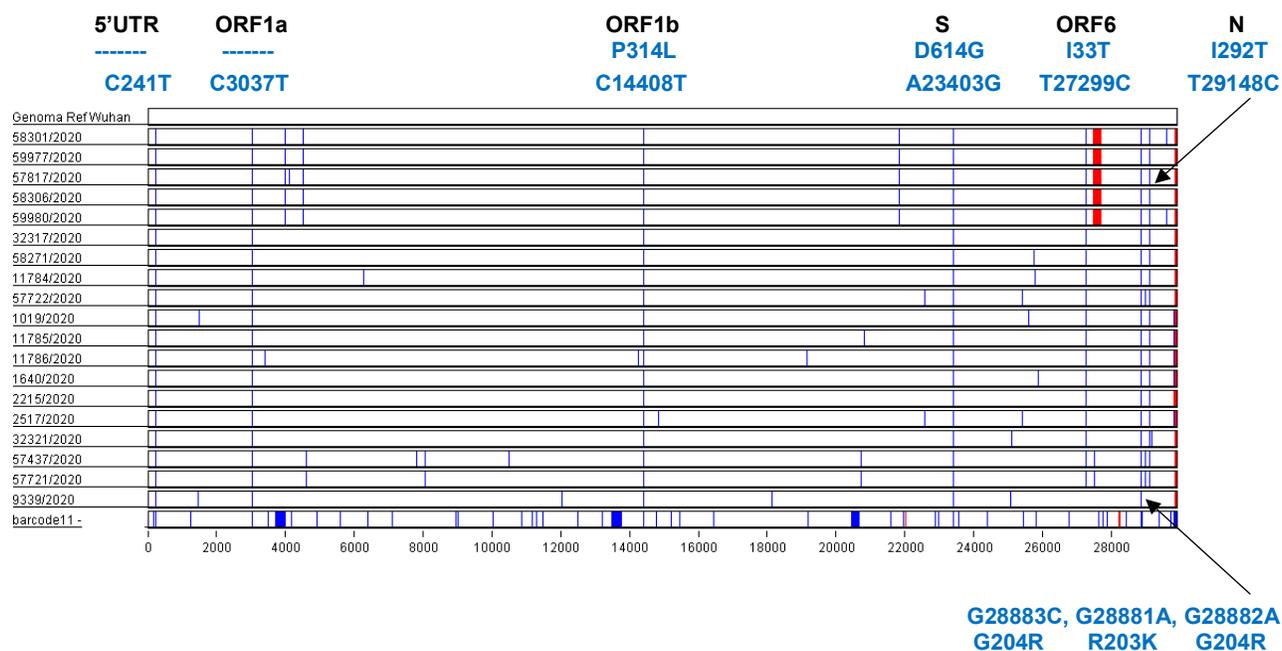


FIGURA 19 - Figura representativa dos SNP'S (azul) e deleções (vermelho) observados nas amostras sequenciadas 2020 e 2021. Fonte: Programa Base by Base disponível em Gatu Viral Bioinformatics Research Centre, com a descrição das principais substituições e regiões do genoma pelo autor.

Nessas amostras, pôde-se verificar que a grande maioria das substituições de aminoácidos em 2020 ocorreu em N (56), S (28), ORF1b (24) e ORF6(18), sendo essa grande frequência atribuída às substituições em N R203K(G28881A), G204R(G28883C) e I292T(T29148C); em ORF1b P314L(C14408T); em ORF6 I33T(T27299C) e em S D614G (A23403G). Todas as 19 sequências de 2020 apresentaram a mutação D614G em S e apenas uma não apresentou a mutação I33T em ORF6. As amostras de n^{os} 1 a 5 foram as únicas que apresentaram a mutação em S (S98F) e estão relacionadas entre si por terem sido coletadas de integrantes de uma mesma família.

Com relação ao genoma da amostra coletada em 2021, o total de substituições de aminoácidos foi bem maior que as de 2020 - 33 substituições (Figura 20), enquanto as amostras de 2020 tinham, em média, 7,7 substituições - com destaque para o maior número de mutações em S – 8 (dentre elas, L452R e D614G, descritas em literatura (MOTOZONO et al., 2021; KORBER et al., 2020) por aumentarem a transmissibilidade e infectividade) enquanto as amostras de 2020 apresentaram, em média, 1,47 substituições.

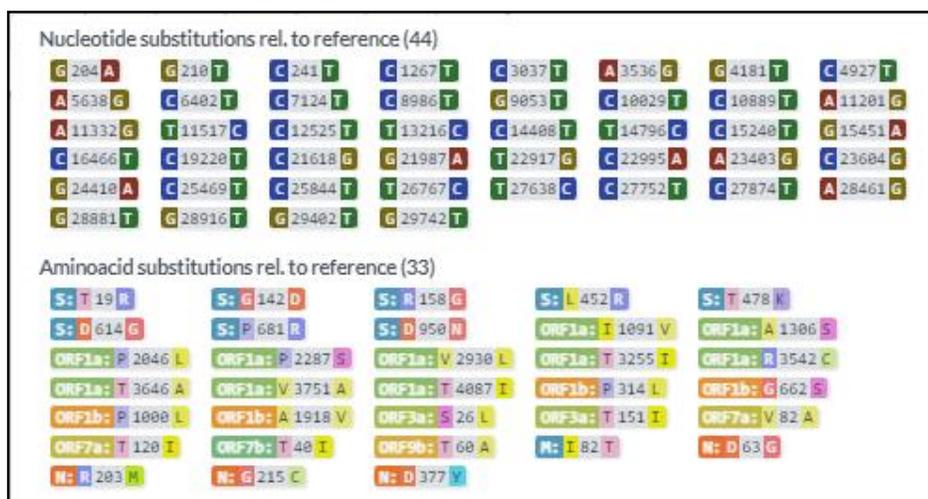


FIGURA 20 – Figura representativa das substituições de nucleotídeos (44) e de aminoácidos (33) apresentadas pela amostra 7842/2021 IBEX. Fonte: Nextclade (<https://clades.nextstrain.org>).

4.4.2. Nas sequências do Estado do Rio de Janeiro obtidas no GISAID entre 2020 e 2023

Nas sequências do GISAID 2020, foram observadas substituições de nucleotídeos/aminoácidos principalmente nos genes N, S, ORF1b e ORF6, o que também ocorreu também nas amostras IBEx sequenciadas. A partir de 2021, as regiões com maior número de mutações passaram a ser S, ORF1a, ORF1b e N evidenciando a mudança no perfil de mutações ao longo dos anos analisados (Quadro 8). O gene E demonstrou ser altamente conservado até o ano de 2022.

QUADRO 8- Comparativo de número de substituições de aminoácidos entre as amostras IBEx e sequências GISAID 2020 a 2023, por região do genoma.

	ORF1a	ORF1b	S	ORF3a	E	M	ORF6	ORF7a	ORF7b	ORF8	ORF9b	N	
Tamanho do gen (Nr de aminoácidos)	4401	2696	1274	276	76	223	62	122	44	122	98	420	4 regiões com > Nr substituições AAS
Amostras IBEX 2020 (19)	11	24	28	5	0	0	18	2	0	0	0	59	N, S, ORF1b, ORF6
Sequências GISAID 2020 (298)	159	362	383	37	0	1	271	7	0	17	12	868	N, S, ORF1b, ORF6
Amostra Delta 7842/IBEX 2021 (1)	10	4	8	2	0	1	0	2	1	0	1	4	ORF1a, S, N e ORF1b
Sequências GISAID 2021 (788)	7221	3601	5951	1006	0	805	26	1603	795	103	847	3284	ORF1a, S, ORF1b, N
Sequências GISAID 2021/2022 (158)	1270	573	2635	169	0	269	7	190	88	17	188	600	S, ORF1a, N, ORF1b
Sequências GISAID 2023 (257)	2624	1629	9102	326	438	600	177	21	1	175	266	1143	S, ORF1a, ORF1b, N

Ainda no Quadro 8, avaliando-se a proporção entre o número de mutações por região e o número de genomas analisados nos anos de 2020 (predominância da linhagem B.1.1.33), 2021 (predominância de Delta) e 2023 (predominância de Ômicron), evidencia-se ainda, que a ORF1b era uma das regiões com mais substituições de aminoácidos no início da pandemia e manteve este padrão até 2023, assim como os genes S e N. Sendo assim, a ORF3a passou a apresentar um número maior de mutações a partir de 2021, após a entrada da variante Delta (37 mutações em 298 genomas em 2020-0,124; 1006/788 em 2021-1,276 e 326/257 em 2023-1,268), assim como ocorreu com as mutações em M (1 mutação em 298 genomas em 2020- 0,003; 805/788 em 2021- 1,021 e 600/257 em 2023-2,335). Pelo mesmo raciocínio, evidencia-se que ORF6 em 2020 (linhagem B.1.1.33 predominante) era uma região proporcionalmente com número maior de mutações

do que nos anos seguintes (2020-0,909; 2021-0,033 e 2023-0,689), e que ORF7a acumulou mais mutações em genomas Delta (2021) do que em Ômicron (2020-0,024; 2021-2,035; 2023-0,082). As mutações em ORF7b foram mais marcantes em genomas da variante Delta (2020-sem mutações; 2021-1,009 e 2023-0,004). A região ORF8 demonstrou aumento progressivo do número de mutações (2020-0,006; 2021-0,131 e 2023-0,681). O gene N e a ORF9b também acumularam mais mutações a partir de 2021, sem aumentos significativos entre genomas Delta (2021) e Ômicron (2023) (Gene N: 2020-2,913; 2021-4,168; 2023-4,447 e ORF9b: 2020-0,040; 2021-1,075 e 2023-1,035). O gene S mostrou-se, desde 2020 ser um dos mais susceptíveis às mutações (Gene S: 2020-1,319; 2021-7,552; 2023-35,416).

Com relação às médias de número substituições de nucleotídeos e aminoácidos (Quadro 9), foi observado aumento de quase 4 vezes neste número entre os anos de 2020 e 2021, com a entrada das variantes P.1 (Gamma) e AY.99.2 (Delta). A média GISAID passou de 11 para 41 (variante Gamma) e 43 (variante Delta).

QUADRO 9- Comparativo de médias substituições de nucleotídeos, aminoácidos, inserções, deleções e mudanças de quadro de leitura no universo total de genomas.

MÉDIAS	Nr de genomas analisados	Substituições de nucleotídeos	Deleções	Inserções	Mudanças de quadro de leitura	Substituições de Aminoácidos	Deleções de Aminoácidos	Inserções de Aminoácidos	Cobertura (%)
Amostras IBEX 2020	19	13	5*	0	0	8	17	0	99,79
Sequências GISAID 2020	298	11	1	0	0	7	0	0	99,59
Sequências Delta GISAID 2021	784	43	13	0	0	32	4	0	99,51
Amostra Delta 7842/IBEX 2021	1	44	13	0	0	33	4	0	97,08
Sequências Delta GISAID 2022	89	44	13	0	0	33	4	0	99,51
Sequências Gamma GISAID 2021	4	44	10	4	0	29	3	0	99,68
Sequências Gamma GISAID 2022	2	41	9	4	0	27	3	0	99,71
Sequência Recombinante GISAID 2021/ 2022	1	52	0	0	0	43	0	0	99,67
Sequências Omicron GISAID 2021/2022	66	57	9	0	0	46	3	0	98,57
Sequências Omicron GISAID 2023	257	88	58	0	0	64	14	0	98,37
Total	1521								

*a média seria 53 se fossem consideradas as 5 amostras com 190 deleções em ORF7a

NOTA: Para o cálculo da média de deleções em amostras IBEX 2020 foram desconsiderados os dados atípicos (5 amostras de integrantes de uma mesma com 190 deleções em ORF7a).

Ao final de 2021, a média seguiu aumentando de forma mais discreta, em torno de 18%, após a entrada da variante BA.1 (Ômicron) no Rio de Janeiro e seguiu em tendência crescente no ano de 2023, com aumento de 54% entre as linhagens Ômicron observadas em 2021/2022 (57 substituições de nucleotídeos) e 2023 (88 substituições).

O Quadro 9 também evidencia que os resultados das amostras sequenciadas são próximos aos apresentados nas sequências do GISAID, nos períodos avaliados.

No universo de sequências 2020, as principais substituições de nucleotídeos/aminoácidos foram as mesmas encontradas nas amostras sequenciadas (descritas na Quadro 7).

As principais substituições de aminoácidos nos genes S e ORF1a (onde foram observadas a maioria das mutações) nos anos de 2020 a 2023 por variante e respectivos clados são mostrados nas Quadros 10 e 11. De acordo com a Quadro 10, observa-se, ao longo dos anos, o aumento do número de mutações em S entre as linhagens de SARS-CoV-2 e a substituição da variante Delta pela Ômicron a partir do final de 2021. Observa-se, também, que as variantes Delta e Gamma, em 2021, apresentam mutações em S bem características, compartilhando apenas a mutação D614G em comum. A mutação em S H655Y (Quadro 10), surgida inicialmente em variantes Gamma (2021), foi transmitida até as linhagens mais recentes, em 2023. O mesmo ocorreu com as mutações G142D, T478K e P681R surgidas em Delta (2021). A partir do final de 2021/início de 2022, com a circulação de Delta, Gamma e Ômicron, registra-se um número maior de mutações em S sendo compartilhadas entre as linhagens circulantes. Em 2023, todas as linhagens analisadas pertenciam à variante Ômicron. Observa-se, também, o aumento no número de linhagens (e conseqüentemente clados) da variante Ômicron entre os anos de 2021/2022 para 2023. A Quadro 10 mostra, ainda, que a substituição de aminoácidos D614G, presente nos genomas analisados desde 2020, permaneceu durante todo o período em estudo. Nas Quadros 10 e 11 é possível também observar a existência de mutações características da variante Delta que são comuns às encontradas na amostra 7842/2021 sequenciada no IBEx.

Assim como em S, novas mutações em ORF1a surgiram nas linhagens da variante Ômicron desde a entrada do vírus, no final de 2021, até o ano de 2023 (Quadro 11) As mutações K856R, deleção S2803_, L2084I, A2710T, T3255I,

P3395H, deleção L3674_ e I3758V foram mantidas desde as linhagens Ômicron iniciais.

QUADRO 10 - Principais substituições de aminoácidos no gene S observadas entre 2020 e 2023, de acordo com a variante (denominação OMS) e clado.

Substituição de AA em S	IBEX e GISAID 2020	IBEX e GISAID 2021	GISAID 2021/2022	GISAID 2023
T20N		20J		
P26S		20J		
D138Y		20J		
R190S		20J		
K417T		20J		
E484K		20J		
N501Y		20J		
D614G	20B	20J, 21J	20J, 21J e 21K	21K, 23A, 22E, 22F
H655Y		20J	20J, 21K	21K, 23A, 22E, 22F
T1027I		20J		
V1176F		20J		
T19R		21J	21J	
G142D		21J	21J, 21K	23A, 22E, 22F
E156_		21J	21J	
F157_		21J	21J	
R158G		21J	21J	
L452R		21J	21J	
T478K		21J	21J, 21K	21K
P681R		21J	21J, 21K	21K
D950N		21J	21J	
H69_			21K	21K, 22E
H70_				22E
V70_			21K	21K
T95I			21K	21K
Y145D			21K	21K
G339D			21K	21K
S375F			21K	21K
K417N			21K	21K, 23A, 22E, 22F
N440K			21K	21K, 22E
G446S			21K	
S477N			21K	21K
E484A			21K	
N501Y			21K	21K
Y505H			21K	21K, 23A, 22E, 22F
T547K			21K	21K
P681H			21K	21K, 23A, 22E, 22F

LEGENDA:

	B.1.1.33 (sem denominação OMS)
	DELTA
	GAMMA
	ÔMICRON

	GAMMA E ÔMICRON
	DELTA, GAMMA E ÔMICRON
	DELTA E ÔMICRON
	GAMMA E DELTA

(continua)

(continuação)

Substituição de AA em S	IBEX e GISAID 2020	IBEX e GISAID 2021	GISAID 2021/2022	GISAID 2023
N764K			21K	21K, 23A, 22E, 22F
D796Y			21K	21K, 23A, 22E, 22F
N856K			21K	21K
Q954H			21K	21K, 23A, 22E, 22F
N969K			21K	21K, 23A, 22E, 22F
L981F			21K	21K
G142				21K
V143				21K
Y144				21K, 22E
Y145D				21K
N211				21K
L212I				21K
INS_214EPE				21K
D843G				21K
I850L				21K
L24				23A, 22E, 22F
P25				23A, 22E, 22F
P26				23A, 22E, 22F
A27S				23A, 22E, 22F
V83A				23A,22F
H146Q				23A,22F
Q183E				23A,22F
V213E				23A,22F
G252V				23A,22F
G339H				23A,22F
R346T				23A, 22E, 22F
L368I				23A
S375F				23A, 22E, 22F
S376A				23A, 22E, 22F
R408S				23A, 22E, 22F
V445P				23A, 22F
G446S				23A, 22F
N460K				23A, 22E, 22F
S477N				23A, 22E, 22F
T478K				23A, 22E, 22F
F486P				23A, 22F
F490S				23A, 22F
Q954H				23A, 22E, 22F
V213G				22E
G339D				22E
N444T				22E
F486V				22E

LEGENDA:

 B.1.1.33 (sem denominação OMS)
 DELTA
 GAMMA
 ÔMICRON

 GAMMA E ÔMICRON
 DELTA, GAMMA E ÔMICRON
 DELTA E ÔMICRON
 GAMMA E DELTA

QUADRO 11 - Principais substituições de aminoácidos em ORF1a observadas entre 2020 e 2023, de acordo com a variante (denominação OMS) e clado.

Substituição de AA em ORF1a	IBEx e GISAID 2020	IBEx e GISAID 2021	GISAID 2021/2022	GISAID 2023
C241T	20B			
C3037T	20B			
I1091V		21J	21J	
A1306S		21J	21J	
P2046L		21J	21J	
P2287S		21J	21J	
T2910I		21J	21J	
V2930L		21J	21J	
T3255I		21J	21J	
T3646A		21J	21J	
T4087I		21J	21J	
S1188L		20J	20J	
K1795Q		20J	20J	
K47R				23A, 22F
S135R				23A, 22E, 22F
L450I				21K
Q556K				22E
T842I				23A, 22E, 22F
K856R			21K	21K
G1307S				23A, 22E, 22F
L2084I			21K	21K
A2344T				21K
D2579G				22F
A2710T			21K	21K
L3027F				23A, 22E, 22F
T3090I				23A, 22E, 22F
L3201P				21K, 22F
T3255I			21K	21K, 23A, 22E, 22F
P3395H			21K	21K, 23A, 22E, 22F
I3758V			21K	21K
L3829F				22E

LEGENDA:

 B.1.1.33 (sem denominação OMS)
 DELTA
 GAMMA
 ÔMICRON

 GAMMA E ÔMICRON
 DELTA, GAMMA E ÔMICRON
 DELTA E ÔMICRON
 GAMMA E DELTA

A Quadro 12 mostra o mapa de calor com as frequências de todas as mutações de aminoácidos em S encontradas nos genomas de 2020 a 2023, ressaltando em negrito as mutações únicas em S observadas nos genomas da variante Ômicron em 2023. É também importante ressaltar que mutações em S como T95I, Q493R, G496S, T547K, N856K e L981F, que eram frequentes em 100% dos genomas Ômicron iniciais, deixaram de existir em linhagens Ômicron de 2023.

QUADRO 12 – Mapa de calor com as frequências de todas as mutações em S encontradas nos genomas de 2020 a 2023. O número de genomas de cada variante encontra-se em parênteses. Em negrito estão destacadas as mutações únicas em S observadas nos genomas da variante Ômicron em 2023.

Substituição de AA	Gamma 2021 (%) (4)	Delta 2021 (%) (784)	Ômicron Dez21 a Jan22 (%) (66)	Ômicron 2023 (%) (257)
T19I				100,0
N679K			100,0	100,0
P681H			100,0	100,0
N764K			100,0	100,0
D796Y			100,0	100,0
N969K			100,0	100,0
H655Y	100,0		100,0	99,6
S477N			100,0	99,6
T478K		97,7	100,0	99,6
A27S				99,2
G142D		14,9		99,2
E484A			100,0	99,2
Q498R			97,0	99,2
N501Y	100,0		100,0	99,2
Y505H			87,9	99,2
Q954H			100,0	99,2
N460K				98,8
S373P			97,0	95,7
S375F			97,0	95,7
T376A				95,7
D405N				95,7
N440K			100,0	94,6
K417N			97,0	94,2
R408S				93,8
D614G	100,0	100,0	100,0	93,0
S371F				86,4
R346T				78,6
G339H				66,5

LEGENDA

- 99,1 A 100%
- 90 A 99%
- 80 A 89%
- 70 A 79%
- 60 A 69%
- 50 A 59%
- 40 A 49%
- 30 A 39%
- 20 A 29%
- 10 A 19%
- 1,1 A 9%
- ATÉ 1%
- 0%

(continua)

(continuação)

Substituição de AA	Gamma 2021 (%) (4)	Delta 2021 (%) (784)	Ômicron Dez21 a Jan22 (%) (66)	Ômicron 2023 (%) (257)
V213E				66,1
G446S			100,0	66,1
F490S				66,1
V445P				65,8
G252V				65,4
F486P				61,5
L368I				59,1
H146Q				58,0
V213G				33,9
G339D			100,0	33,5
F486V				33,5
L452R	0,0	97,7		33,1
K444T				30,7
F157L				4,7
F157S				4,7
F456L				4,3
F486S				4,3
L5F			1,5	2,3
Q613H				1,9
Q677H			1,5	0,8
A67V			100,0	0,4
K150E				0,4
H245Y				0,4
P251S				0,4
D253G				0,4
N282S				0,4
T299I				0,4
K304Q				0,4
A411S				0,4
N440R				0,4
A475V				0,4
T791A				0,4
Q804L				0,4
T883I				0,4
G1085R				0,4
N1125K				0,4
C1243F				0,4
Y45D				
T95I			100,0	
R158G				

(continua)

(continuação)

Substituição de AA	Gamma 2021 (%) (4)	Delta 2021 (%) (784)	Ômicron Dez21 a Jan22 (%) (66)	Ômicron 2023 (%) (257)
Q183E				
L212I			4,6	
R346K			30,3	
S371L			97,0	
Q493R			100,0	
G496S			100,0	
T547K			100,0	
V642G			1,5	
N856K			100,0	
L981F			100,0	
T19R		99,5		
T20N	100,0			
P26S	100,0			
E484K	100,0	0,1		
V1176F	100,0			
D138Y	100,0	0,2		
R190S	100,0			
T1027I	100,0			
P681R		99,6		
D950N		99,6		

Foi também observado, dentre os genomas do período de Dezembro de 2021 a Janeiro de 2022 o genoma recombinante XZ (Número 89 - RJ-COVID-19-UFRJ-62795/2022-01-17|EPI_ISL_16754208), coletado em 17 de janeiro de 2022, que até então não tinha sido observado no universo de genomas analisado. As principais mutações da recombinante XZ são mostrados na Quadro 13. A partir de 2023, genomas recombinantes passaram a ser mais frequentes.

QUADRO 13 – Informações de qualidade do genoma e principais mutações observadas na recombinante XZ (genoma Número 89 - RJ-COVID-19-UFRJ-62795/2022-01-17|EPI_ISL_16754208, de Janeiro de 2022). Fonte: dados obtidos no Nextclade.

Clado	Linhagem Pango	Denominação OMS	Status geral de qualidade	Total de Substituições	Total de deleções	Total de Inserções	Total de mudanças de quadros de leitura	Total de dados perdidos	Total de não ACGTNS	Total de subst. AAS	Total de deleções de AAS	Total de inserções de AAS	Total de AAS desconhecido	Cobertura	Deleções	Inserções	Mutações de quadro de leitura	Deleções de AAS	Inserções de AAS
recombinante	XZ	Recombinante	ruim	62	0	0	0	10	0	43	0	0	8	99,67	---	---	---	---	---
Substituições de nucleotídeos				Substituições de AAS						Mutações privadas de Nucleotídeos (reversões)			Mutações privadas de Nucleotídeos (mutações rotuladas)		Mutações privadas de Nucleotídeos (mutações não rotuladas)				
C:241T,C:3037T,G:4184A,C:9534T,C:10029T,C:10449A,G:11291A,C:14408T,C:17410T,A:18163G,C:19955T,G:21987A,T:2200G,G:22578A,C:22674T,T:22679C,C:22686T,A:22688G,G:22775A,A:22786C,C:22792T,G:22813T,T:22882G,G:22992A,C:22995A,A:23013C,A:23040G,A:23055G,A:23063T,T:23075C,A:23403G,C:23525T,T:23599G,C:23604A,C:23854A,G:23948T,A:24424T,T:24469A,C:25000T,C:25584T,C:26060T,C:26270T,C:26577G,G:26709A,C:26858T,A:27259C,C:27807T,A:28271T,C:28311T,G:28881A,G:28882A,G:28883C				E:T91M;Q19E,M:A63T,N:P13L,N:R203K,N:G204R,ORF1a:G1307S,ORF1a:T3090I,ORF1a:T3255I,ORF1a:P3395H,ORF1a:G3676S,ORF1b:P314L,ORF1b:R1315C,ORF1b:1566V,ORF1b:T2163I,ORF3a:T223I,ORF9b:P10S,S:G142D,S:V213G,S:G339D,S:S371F,S:S373P,S:S375F,S:T376A,S:D405N,S:R408S,S:K417N,S:N440K,S:S477N,S:T478K,S:E484A,S:Q493R,S:Q498R,S:N501Y,S:Y505H,S:D614G,S:H655Y,S:N679K,S:P681H,S:N764K,S:D796Y,S:Q954H,S:N969K						G670T,T2790C,T4321C,T9344C,G9424A,T9866C,T10198C,A10447G,T12880C,T15714C,G20055A,A20398G,T21618C,G26530A			C26858T 21L&22C&23A&22A&22D&22F&rec&23B		-11291A				

Em regiões de menor frequência de mutações em 2020, como no Gene M, foi observada apenas uma substituição (G26730T -V70F), no genoma N^o 232 - 456092|2020-04-10, representando a frequência de 0,347%. Já a partir de 2021, nota-se a mudança de perfil de variação, uma vez que foi observada a substituição em M T26767C (I82T) em 99,36% dos genomas, todos da variante Delta. Somente as variantes Gamma 2021 não apresentaram esta substituição em M. Com relação ao gene que codifica a proteína do Envelope (E), foi observado que a partir de 2023, 99,22% dos genomas passaram a incorporar a mutação T91, que possui efeito estabilizador nesta proteína (MOU, Kejie et al., 2021).

4.5. INSERÇÕES E DELEÇÕES

Inserções de aminoácidos foram pouco observadas no total de amostras e genomas avaliados. As inserções observadas nas linhagens mais frequentes de cada período são mostradas no Quadro 14.

QUADRO 14- Frequência de inserções nas principais linhagens em cada período de estudo.

Período	Linhagem (Pango) e Nr de indivíduos	Inserções (Nucleotídeos)	Inserções (Aminoácidos)	Frequência na linhagem
Abril a Junho de 2020	B.B.1.33 (288 genomas, 18 do IBEx)	29903:A	----	0,35
		0:CT	----	0,35
		0:TC	----	0,35
	B.1.1.28 (16 genomas, 1 do IBEx)	----	----	0,00
Outubro de 2021	AY.99.2 (785 genomas, 1 do lbex)	53:CC		0,13
		21886:GCTGTT	S:108:AV	0,13
	P.1 (4 genomas)	28262:AACA	----	50,00
Dez 21 a Jan22	AY.99.2 (58 genomas)	----	----	0,00
	BA.1 (25 genomas)	----	----	0,00
	BA1.1 (20 genomas)	----	----	0,00
	XBB1.5 (81 genomas)	----	----	0,00
Jan a Jun 23	BQ1.1 (36 genomas)	----	----	0,00
	XBB1.5.17 (14 genomas)	----	----	0,00

Deleções foram observadas tanto em amostras sequenciadas (Figura 21) quanto em genomas disponíveis no GISAID (Quadro 15). As amostras sequenciadas de N^{os} 1 a 5 (58301/2020, 59977/2020, 59980/2020, 58306/2020 e 57817/2020 do IBEx), todas IBEx/2020 da linhagem B.1.1.33 (Pango), apresentaram uma deleção de 190 a 191 nucleotídeos, que foi descrita e alvo de estudo pelo grupo da colaboração IBEx-UFRJ, citado anteriormente (SIMAS et al., 2023). O grupo de pesquisadores ampliou as análises, realizando o isolamento do vírus e estudos “in vitro”, para verificar o impacto da deleção no fitness viral, conforme publicação. (SIMAS, et al., 2023). Esses 5 genomas apresentaram um número de deleções de nucleotídeos muito superior aos demais sequenciados, ocasionando a deleção de 64 aminoácidos, todos em ORF7a.

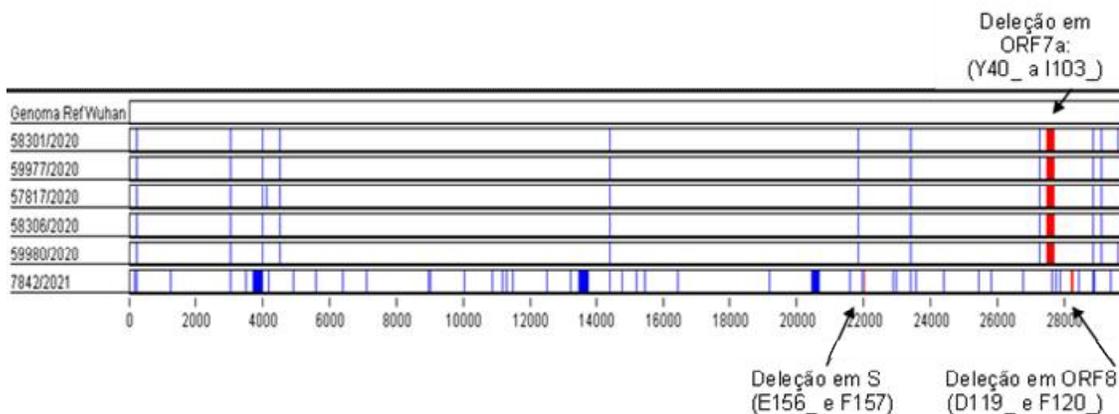


FIGURA 21 – Figura representativa dos SNP's e deleções (vermelho) observadas das amostras Nr 1 a 5 (2020) e Nr 20 (2021) . Fonte: Programa Base by Base disponível em Gatu Viral Bioinformatics Research Centre, com a sinalização das deleções observadas pelo autor.

No universo de genomas avaliados GISAID RJ (Quadro 15), a frequência de deleções em 2020 não passou de 6,25%, em genomas da linhagem B.1.1.28. Em 2021, os genomas da linhagem AY.99.2 (Delta) apresentaram as deleções de aminoácidos ORF8:D119-, ORF8:F120-, S:E156_, S:F157_ em 98,1% dos genomas, a qual foi mantida entre Dezembro de 2021/Janeiro de 2022, e a linhagem P.1 (Gamma) a deleção ORF1a:S3675-, ORF1a:G3676_, ORF1a:F3677_ em 75% de seus genomas. Entre Dezembro de 2021 e Janeiro de 2022, os genomas das linhagens BA.1 e BA.1.1 (Ômicron) apresentaram o conjunto de deleções ORF1a:S2083_, S:H69_, S:V70_ em 52,0% e 80,0% dos genomas, respectivamente. No 1º semestre de 2023, os genomas das linhagens XBB.1.5 e XBB.1.5.17 apresentaram frequências relativamente altas para o conjunto de deleções N:E31_, N:R32_, N:S33_, ORF1a:S3675_, ORF1a:G3676_, ORF1a:F3677_, ORF9b:E27_, ORF9b:N28_, ORF9b:A29_, S:L24_, S:P25_, S:P26_, S:Y144_ em 87,65% e 100,00% dos genomas, respectivamente. Já a linhagem BQ.1.1 apresentou o conjunto de deleções N:E31_, N:R32_, N:S33_, ORF1a:S3675_, ORF1a:G3676_, ORF1a:F3677_, ORF9b:E27_, ORF9b:N28_, ORF9b:A29_, S:L24_, S:P25_, S:P26_ em comum com as linhagens XBB.1.5 e XBB.1.5.17 e as deleções adicionais em S S:H69_ e S:V70_ em 61,11% dos genomas.

QUADRO 15- Frequências das deleções nas principais linhagens de SARS-COV-2 nos genomas GISAID RJ entre 2020 e 2023..

Período	Linhagem (Pango) e Nr de genomas	Deleções (Nucleotídeos)	Deleções (Aminoácidos)	Frequência na linhagem
Abril a Junho de 2020	B.B.1.33 (288 genomas, 18 do IBEx)	27511-27700	ORF7a:Y40-,ORF7a:E41-,ORF7a:G42-,ORF7a:N43-,ORF7a:S44-,ORF7a:P45-,ORF7a:F46-,ORF7a:H47-,ORF7a:P48-,ORF7a:L49-,ORF7a:A50-,ORF7a:D51-,ORF7a:N52-,ORF7a:K53-,ORF7a:F54-,ORF7a:A55-,ORF7a:L56-,ORF7a:T57-,ORF7a:C58-,ORF7a:F59-,ORF7a:S60-,ORF7a:T61-,ORF7a:Q62-,ORF7a:F63-,ORF7a:A64-,ORF7a:F65-,ORF7a:A66-,ORF7a:C67-,ORF7a:P68-,ORF7a:D69-,ORF7a:G70-,ORF7a:V71-,ORF7a:K72-,ORF7a:H73-,ORF7a:V74-,ORF7a:Y75-,ORF7a:Q76-,ORF7a:L77-,ORF7a:R78-,ORF7a:A79-,ORF7a:R80-,ORF7a:S81-,ORF7a:V82-,ORF7a:S83-,ORF7a:P84-,ORF7a:K85-,ORF7a:L86-,ORF7a:F87-,ORF7a:I88-,ORF7a:R89-,ORF7a:Q90-,ORF7a:E91-,ORF7a:E92-,ORF7a:V93-,ORF7a:Q94-,ORF7a:E95-,ORF7a:L96-,ORF7a:Y97-,ORF7a:S98-,ORF7a:P99-,ORF7a:I100-,ORF7a:F101-,ORF7a:L102-,ORF7a:I103-	1,74
		21992-21994	S:Y144-	0,35
		11288-11296,21992-21994	ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-,S:Y144-	0,35
		21992-21994	S:Y144-	0,35
		29779-29780	----	1,39
		29857-29873	----	0,35
		29857-29862	----	0,35
		8-9	----	0,35
		4-5	----	0,35

(continua)

(continuação)

Período	Linhagem (Pango) e Nr de genomas	Deleções (Nucleotídeos)	Deleções (Aminoácidos)	Frequência na linhagem
Abril a Junho de 2020	B.B.1.33 (288 genomas, 18 do IBEx)	12600-12602	ORF1a:S4112-	0,35
		935-937	ORF1a:T224-	0,35
		29756-29761	----	0,35
		21602-21613	S:Q14-,S:C15-,S:V16-,S:N17-	0,35
		22283-22291	S:L241-,S:L242-,S:A243-	0,35
		21984-21989	S:L141-,S:G142-	0,35
		20407-20412	ORF1b:F2314-,ORF1b:E2315-	0,35
Abril a Junho de 2020	B.1.1.28 (16 genomas, 1 do IBEx)	15-25	----	6,25
Outubro de 2021	AY.99.2 (785 genomas, 1 do lbex)	22029-22034,28248-28253,28271	ORF8:D119-,ORF8:F120-,S:E156-,S:F157-	98,09
	P.1 (4 genomas)	11288-11296	ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-	75,00
Dez 21 a Jan22	AY.99.2 (58 genomas)	22029-22034,28248-28253,28271	ORF8:D119-,ORF8:F120-,S:E156-,S:F157-	98,28
	BA.1 (25 genomas)	6513-6515,21765-21770	ORF1a:S2083-,S:H69-,S:V70-	52,00
	BA1.1 (20 genomas)			80,00
Jan a Jun 23	XBB1.5 (81 genomas)	11288-11296,21633-21641,21992-21994,28362-28370,29734-29759	N:E31-,N:R32-,N:S33-,ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-,ORF9b:E27-,ORF9b:N28-,ORF9b:A29-,S:L24-,S:P25-,S:P26-, S:Y144-	87,65
	BQ1.1 (36 genomas)	11288-11296,21633-21641,21765-21770,28362-28370,29734-29759	N:E31-,N:R32-,N:S33-,ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-,ORF9b:E27-,ORF9b:N28-,ORF9b:A29-,S:L24-,S:P25-,S:P26-, S:H69-,S:V70-	61,11
	XBB1.5.17 (14 genomas)	11288-11296,21633-21641,21992-21994,28362-28370,29734-29759	N:E31-,N:R32-,N:S33-,ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-,ORF9b:E27-,ORF9b:N28-,ORF9b:A29-,S:L24-,S:P25-,S:P26-, S:Y144-	100,00

4.6. MUDANÇAS DE QUADRO DE LEITURA (*FRAMESHIFTS*)

As inserções ou deleções de deslocamento de quadro geralmente resultam em uma tradução distorcida ou em uma parada prematura durante a tradução. No universo de amostras sequenciadas, as amostras numeradas de 1 a 5 (58301/2020, 59977/2020, 59980/2020, 58306/2020 e 57817/2020 do IBEx), que eram da linhagem mais frequente (B.1.1.33) no universo de amostras, apresentaram mudança de quadro em ORF7a:104-122, relacionada com a deleção de ORF7a (SARS-CoV-2 ORF7a_Δ190) previamente descrita (Figura 22).

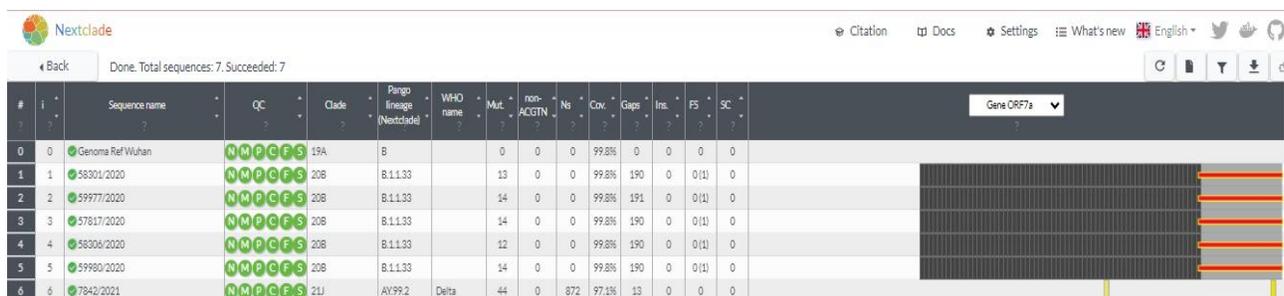


FIGURA 22- Figura representativa dos 5 genomas IBEx sequenciados que apresentaram deleção de nucleotídeos entre as posições 27701 e 27759 (sinalizada em preto) e mudança de quadro de leitura em ORF 7a (faixa de códons 104 a 122). Fonte: Nextclade (<https://clades.nextstrain.org>).

Nos genomas GISAID das linhagens mais frequentes, as mudanças de quadro de leitura não foram usuais (Quadro 16), sendo a maior frequência atribuída a um único genoma da linhagem XBB.1.5 (representando 1,23% dos 81 genomas desta linhagem).

QUADRO 16- Frequências das mudanças de quadro de leitura observadas nas principais linhagens de SARS-CoV-2 em cada período de estudo.

Período	Linhagem (Pango) e Nr de indivíduos	Mudanças de Quadro de leitura	Frequência na linhagem
Abril a Junho de 2020	B.B.1.33 (288 genomas, 18 do IBEx)	ORF7a:104-122	1,74
	B.1.1.28 (16 genomas, 1 do IBEx)	----	0,00

(continua)

(continuação)

Período	Linhagem (Pango) e Nr de indivíduos	Mudanças de Quadro de leitura	Frequência na linhagem
Outubro de 2021	AY.99.2 (785 genomas, 1 do Ibex)	ORF7a:47-122	0,13
	P.1 (4 genomas)	----	0,00
Dez 21 a Jan22	AY.99.2 (58 genomas)	----	0,00
	BA.1 (25 genomas)	----	0,00
	BA1.1 (20 genomas)	----	0,00
Jan a Jun 23	XBB1.5 (81 genomas)	ORF8:122	1,23
	BQ1.1 (36 genomas)	----	0,00
	XBB1.5.17 (14 genomas)	----	0,00

4.6 ANÁLISE FILOGENÉTICA

As árvores filogenéticas a seguir foram construídas no Software MEGA (Molecular Evolutionary Genetics Analysis) versão 11 pelo método estatístico de “Máxima verossimilhança”, com teste de filogenia com 1000 replicatas de “bootstrap”, e modelo de substituição Tamura-Nei. A Figura 23 mostra as 20 amostras sequenciadas, genomas GISAID de janeiro de 2020 representando o início da pandemia e genomas de outubro de 2021, além do genoma referência. Observa-se nesta Figura que os genomas de janeiro de 2020 (início da pandemia) e os de abril a junho de 2020 (1ª onda de casos no Rio de Janeiro) foram dispostos em clados diferentes, indicando que em poucos meses o vírus apresentou mutações que já divergiam dos genomas de janeiro de 2020. Como esperado, a amostra IBEx 2021 (barcode 11) apresentou proximidade com os genomas GISAID de outubro de 2021. Na Figura 23 observa-se ainda o distanciamento entre as amostras sequenciadas de 2020 e as sequências (amostra e genomas) de outubro de 2021. A árvore filogenética da Figura 24 representa os genomas sequenciados, o genoma referência Wuhan e genomas das linhagens mais frequentes nos anos de 2020 a 2023. Observa-se mais uma vez a distância entre as amostras sequenciadas

em 2020 e os genomas dos anos seguintes. A árvore evidencia também a variante P.1 como descendente de B.1.1.28 e que as linhagens BQ e XBB, de 2023, possuem um ancestral comum.

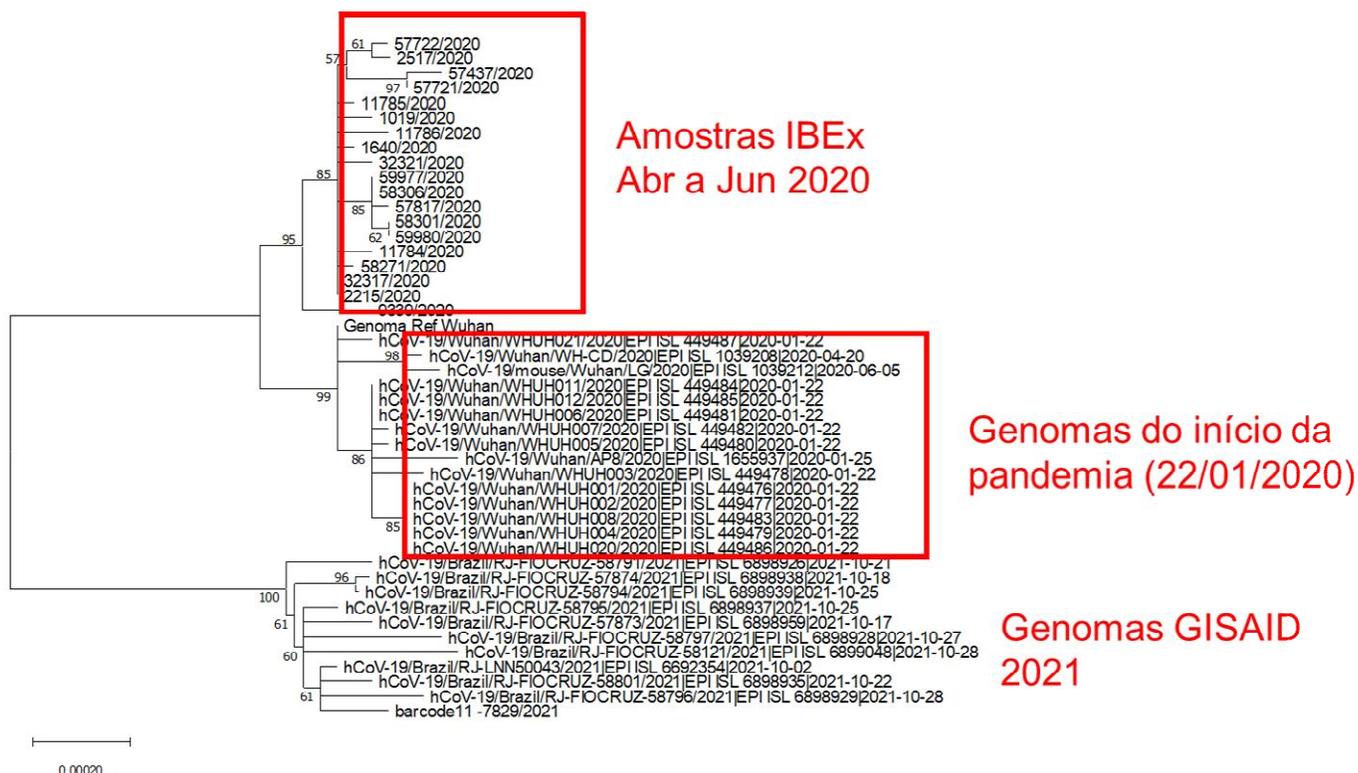


FIGURA 23 - Árvore filogenética construída pelo método estatístico de “Máxima verossimilhança”, com teste de filogenia com 1000 replicatas de “bootstrap” e modelo de substituição Tamura-Nei com genomas das 20 amostras sequenciadas, genomas GISAID do início de 2020 (representando o início da pandemia), genomas GISAID de outubro de 2021 e o genoma de referência Wuhan, evidenciando as distâncias evolutivas entre amostras e genomas de 2020 e 2021.

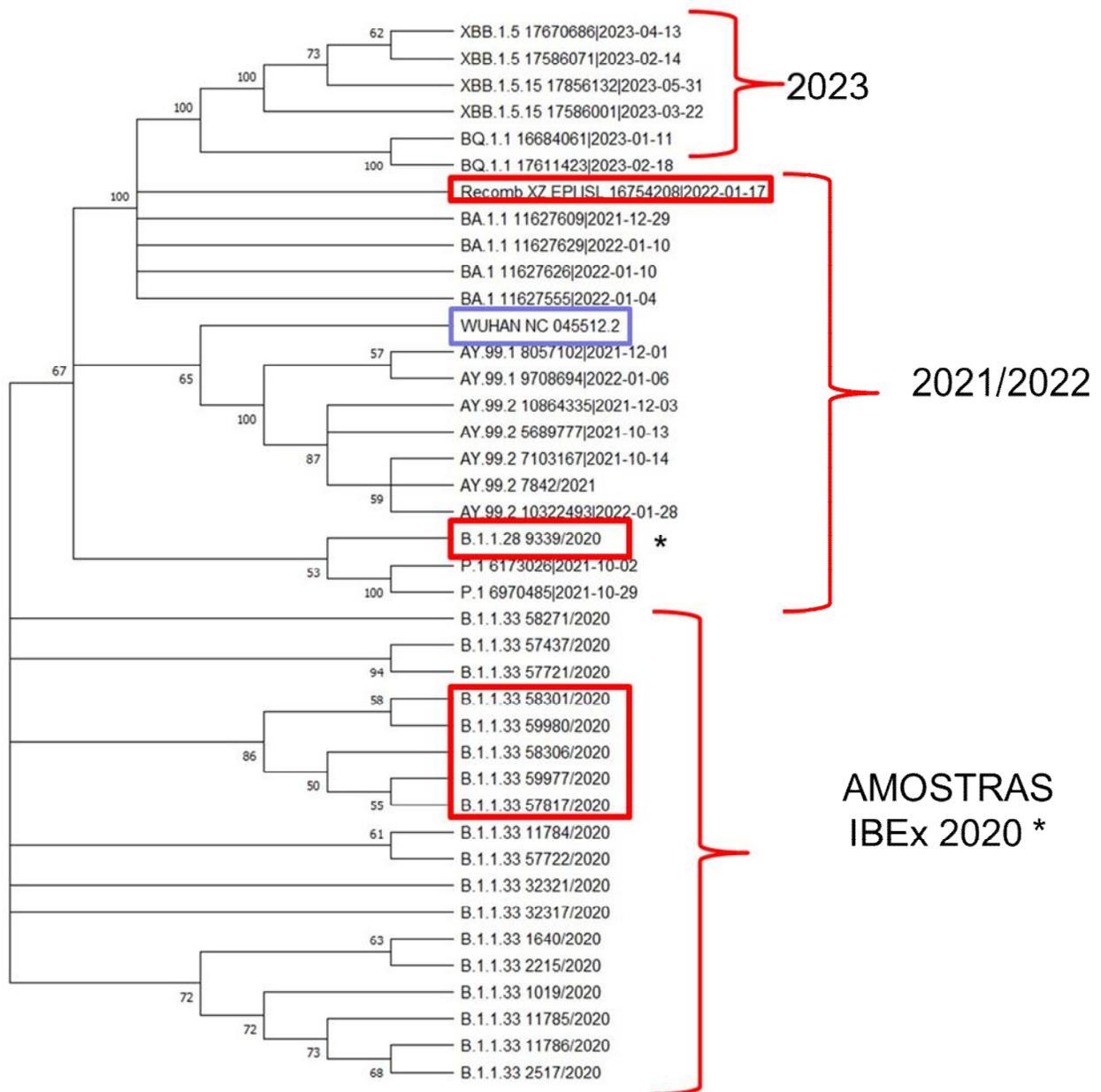


FIGURA 24 - Árvore filogenética construída pelo método estatístico de “Máxima verossimilhança”, com teste de filogenia com 1000 replicatas de “bootstrap” e modelo de substituição Tamura-Nei com genomas das amostras sequenciadas e genomas de linhagens de maior frequência entre os anos de 2020 e 2023, além do genoma de referência Wuhan. São evidenciados a variante P.1 com seu ancestral B.1.1.28, a existência de um ancestral comum entre BQ.1.1. e as linhagens XBB.1.5.15 e 1.5.17 e o distanciamento entre amostras de 2020 e genomas de 2023 .

5 DISCUSSÃO

O papel das forças armadas em acontecimentos como crises humanitárias, epidemias e guerras é bem conhecido não só no Brasil como no mundo. A epidemia de Ebola, ocorrida entre 2014 e 2016 na África Ocidental foi um dos exemplos deste processo, em que militares da França, Alemanha, Reino Unido, China, Canadá e Estados Unidos foram destacados para diversas tarefas de apoio no meio desta resposta específica. Na época, os destacamentos militares estrangeiros foram apresentados como a única resposta para uma lacuna de capacidade a nível internacional (GIBSON-FALL, FAWZIA, 2021). No Brasil, militares também foram enviados às ruas para auxiliar no combate ao vírus da Zika, transmitido pelo mosquito *Aedes aegypti* nos anos de 2015 a 2017, ao mesmo tempo que sinalizavam à população a gravidade da situação (WENHAM & FARIAS, 2019). A pandemia da COVID-19 consolidou ainda mais as forças armadas como intervenientes, ligando os militares às práticas de saúde (GIBSON-FALL, FAWZIA, 2021).

O Exército Brasileiro, por intermédio da operação COVID-19, coordenada pelo Ministério da Defesa, atuou em todo o país por meio dos esforços nas áreas operacional, ensino, humanitária e de saúde (PINHEIRO; PATUSCO; VALVERDE, 2022). Por intermédio de suas Organizações Militares de Saúde e pelo Sistema de Defesa Química, Biológica, Radiológica e Nuclear (SisDQBRN), do qual o IBEx faz parte, foram mobilizados pessoal e recursos no apoio ao diagnóstico e na pesquisa. Logo nos primeiros meses da pandemia, pesquisadores do IBEx atuaram no desenvolvimento e adaptação de protocolos baseados em RT-PCR para a rápida detecção de casos positivos para SARS-CoV-2 (CIPITELLI et al., 2020) e o Instituto tornou-se um importante centro para o Exército Brasileiro, na realização do diagnóstico laboratorial de casos de COVID-19.

Ainda no início da pandemia, a interação do IBEx com outras instituições de pesquisa brasileiras, como a Universidade Federal do Rio de Janeiro (UFRJ), Universidade Federal Fluminense (UFF) e o Instituto Oswaldo Cruz (IOC/Fiocruz) foi de fundamental importância e possibilitou o treinamento, a pesquisa e o sequenciamento de genomas completos de SARS-CoV-2. A aquisição do equipamento MinION, da Nanopore, pelo IBEx em meados de 2021, permite a

realização de estudos genômicos, como no presente trabalho.

O início da pandemia caracterizou-se pela ampla disseminação da linhagem B.1.1.33, que se distribuiu rapidamente por todas as regiões brasileiras e também por vários países ao redor do mundo (RESENDE et al., 2021). Sua provável origem pode ter sido uma linhagem semelhante à B.1.1.33 (“B.1.1.33 like”), introduzida no Brasil a partir da Europa, e que pode ter circulado no país antes da implementação das restrições de circulação e de viagens aéreas por algumas semanas, dando origem à linhagem B.1.1.33. O primeiro caso de COVID-19 foi detectado no Brasil em 25 de fevereiro de 2020 no Estado de São Paulo (RESENDE et al., 2021; CANDIDO et al., 2020), o mais populoso do país (IBGE, 2020) enquanto no Rio de Janeiro, o terceiro estado mais populoso (IBGE, 2020) o primeiro caso foi confirmado em 05 de março do mesmo ano, em uma paciente que retornava da Itália e Alemanha.

A primeira onda da COVID-19 no Rio de Janeiro, ocorrida entre os meses de abril e maio de 2020, pôde ser estudada em conjunto com pesquisadores da UFRJ, graças ao sequenciamento de 19 genomas completos, de amostras provenientes de indivíduos com diferentes faixas etárias, em atendimentos ambulatoriais ou hospitalares, em diferentes localidades no Estado do Rio de Janeiro. O sequenciamento foi realizado no Laboratório de Metabolismo Macromolecular Firmino Torres de Castro, chefiado pela Dra. Rosane Silva, como parte de uma colaboração para a vigilância da COVID-19 no Rio de Janeiro. As 19 amostras do ano de 2020 sequenciadas, referentes à primeira onda, mostraram a predominância da variante B.1.1.33, (94,74%) e a ocorrência de apenas uma amostra (9339/2020 IBEx) da linhagem B.1.1.28 (5,26%), ambas pertencentes ao mesmo clado, 20B, relatado dentre os clados circulantes no início da pandemia (VOLOCH et. al., 2021; RESENDE et al., 2021). A linhagem B.1.1.33 já se encontrava moderadamente dispersa em alguns países da América do Sul (5 a 18%) e com prevalência muito baixa na América do Norte, Europa e Oceania (<1%) (RESENDE et al., 2021). Dados epidemiológicos do período de coleta das amostras, evidenciavam até junho de 2020, 111.903 casos confirmados e 10.057 óbitos no Estado, com o maior número de ocorrências em municípios da região metropolitana, de onde a maioria das amostras foi obtida (<https://painel.saude.rj.gov.br/monitoramento/COVID19.html>, acesso em 06/09/23). O

município de Resende, de onde foi obtida uma das amostras, estava dentre os municípios com o menor número de casos confirmados (< 1000 casos). O Estado de São Paulo, mais populoso, apresentava na época 281.330 casos (<https://www.seade.gov.br/coronavirus/#>, acesso em 06/09/2023).

Dentre as amostras IBEx de 2020, foi possível verificar mutações características das duas linhagens circulantes no Rio de Janeiro (B.1.1.33 e B.1.1.28), em especial as mutações P314L em ORF1b, D614G em S, I33T em ORF6, e as mutações em N G204R, R203K, G204R e I292T. As amostras que despertaram maior interesse dentre as sequenciadas (números 1 a 5), da linhagem B.1.1.33, foram coletadas de integrantes de uma mesma família e apresentaram uma deleção importante de 190 nucleotídeos, afetando os códons 40 ao 103 da ORF7a (correspondendo a quase 50% da ORF). Este padrão de deleções não foi observado em nenhum dos genomas de alta qualidade obtidos no GISAID 2020 para a cidade do Rio de Janeiro, onde o maior número de deleções foi de 4 aminoácidos (amostra N° 04 - IC1112/202021). Os pacientes foram infectados no mesmo período e apresentaram sintomas leves de COVID-19, provavelmente associados à redução da supressão do Interferon I, consequentemente tornando o vírus mais susceptível à resposta imune. SIMAS e colaboradores pesquisaram o impacto desta deleção na replicação e demonstraram que os integrantes da família foram infectados com a mesma cepa viral, ou seja, mesmo com a deleção, o vírus foi capaz de se propagar. Em cultura celular, demonstraram que os vírus isolados foram capazes de replicar em cultura de células, por 24 horas, de forma semelhante aos vírus do tipo selvagem, mas com menos partículas infecciosas após 48 horas. Também foi possível observar que a deleção impactou severamente na estrutura e função da proteína. Sabe-se que a ORF7a desempenha um papel nas interações do vírus com o hospedeiro, por codificar uma das proteínas acessórias relacionadas com a replicação (proteína transmembrana do tipo 1) (SIMAS et.al, 2023), por suprimir a sinalização de Interferon tipo I in vitro (NEMUDRYI et al., 2021) e por interagir com alta eficiência na ligação aos monócitos CD14⁺ suprimindo a apresentação de antígenos feita pelo CD14⁺ (ZHOU et al., 2021). Também está envolvida na propagação da tempestade de citocinas (ZHOU et al., 2021). ZHOU e colaboradores evidenciaram que a ORF7a tem ação imunomoduladora, pois interage com alta eficiência com monócitos CD14⁺ no sangue periférico humano e desencadeia

respostas inflamatórias importantes relacionadas à tempestade de citocinas, como a produção de IL-6, IL-1 β , IL-8 e TNF- α , sendo um alvo terapêutico promissor para medicamentos para a pandemia de COVID-19. Estas deleções levaram às mudanças de quadro de leitura, que foram eventos pouco comuns nas amostras e genomas analisados de 2020 a 2023 (Quadro 16). Ainda sobre as amostras dos cinco familiares, foi observado que foram as únicas que apresentaram a mutação em S (S98F), que apesar de não estar localizada na região RBD, é descrita por aumentar a infectividade do SARS-CoV-2 (LI et al., 2023), provavelmente por ter diminuído a flexibilidade da proteína, favorecendo a sua estabilização (AL-ZYOUN, WALID & HADDAD; HAZEM, 2021). No caso dos cinco familiares, o efeito de maior infectividade associado à mutação S98F pode não ter sido evidenciado na infecção celular *in vitro*, possivelmente em virtude do impacto da deleção em ORF7a, no mesmo genoma.

Ampliando o estudo neste mesmo período, foram analisados mais 298 genomas disponíveis no GISAID-RJ, onde também foi verificada a alta prevalência de B.1.1.33 (90,85%). Estes dados estão de acordo com a tendência mostrada em estudos anteriores, onde foi encontrada a frequência de 90% de B.1.1.33 no início de abril de 2020 (RESENDE et al., 2021) e de 92% em maio do mesmo ano (DA SILVA et al., 2021) no Estado do Rio de Janeiro. As mutações nas proteínas S e N relatadas anteriormente foram as mais prevalentes tanto nas amostras sequenciadas quanto nos genomas GISAID, o que foi compatível com a literatura encontrada. É curioso a existência das variantes B.1.1.28.2 (P.2 ou Zeta) e B.1.1.28.1 (P.1 ou Gamma), ambas originadas da linhagem B.1.1.28, dentre os genomas analisados em amostras supostamente coletadas em abril de 2020, segundo informações fornecidas pela plataforma GISAID (P.2: Nr3-FIOCRUZ-2188-1P/2020|EPI_ISL_2157561|2020-04-13 e P.1: Nr298UFRJ-5943/2020|EPI_ISL_15977585|2020-04-29). Isto porque a variante P.1 somente foi descrita pela primeira vez em janeiro de 2021, isolada de uma família de pacientes que viajaram de Manaus para o Japão (FUJINO et al., 2021), indicando que possa ter havido algum equívoco no registro da data da coleta da amostra, uma vez que é de responsabilidade de quem deposita a informação no GISAID. O mesmo ocorre com relação à variante P.2, descrita por VOLOCH et. al, 2021 relacionando-a a amostras de outubro de 2020 e encontrada em nosso estudo

em Abril de 2020 dentre os genomas GISAID-RJ. Independente dos fatos, cabe ressaltar que a linhagem B.1.1.28, embora tenha se mostrado menos frequente que B.1.1.33 nos dois universos avaliados, foi capaz de evoluir geneticamente e prosseguir com dispersão por todo o país (P.1 e P.2), por meio de mutações principalmente na proteína S, com importância significativa na infectividade e transmissibilidade. Enquanto a variante P.1 apresentou as mutações em S características L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, T1027I, sendo K417T, E484K e N501Y localizadas no domínio de ligação ao receptor da proteína S (FUJINO et al., 2021), P.2 apresentou principalmente a mutação E484K. Tais mutações foram evidenciadas dentre os genomas GISAID avaliados em nosso estudo. Já a linhagem B.1.1.33, mais frequente à época, deu origem a variante N.9, que apresentava a mutação E484K (RESENDE et al., 2021) e a N.10, que apresentava 14 mutações definidoras de linhagem, incluindo mutações na região RBD (E484K e V445A) e três deleções na região NTD da proteína S (RESENDE et al., 2021). Estas linhagens originadas por B.1.1.33, entretanto, não tiveram o mesmo perfil de disseminação de P.1 e P.2 e não foram evidenciadas nem nas amostras nem nos genomas analisados.

Além das linhagens B.1.1.33 e B.1.1.28, foram observadas nos genomas GISAID-RJ outras variantes circulantes, como a B.39 (Clado 19A, mesmo clado da Cepa de Wuhan), B.1 (linhagem associada a surtos na Itália e outros países na Europa que foi disseminada mundialmente- Clado 20A,) (https://cov-lineages.org/lineage_list.html) e B.1.1 (Clado 20B, linhagem de origem europeia (https://cov-lineages.org/lineage_list.html)).

Dentre as diversas mutações apresentadas nos primeiros meses da pandemia, a mutação na proteína S D614G (uma substituição de ácido aspártico por glicina fora da região RBD), que provavelmente confere maior estabilidade à proteína (JACKSON et al., 2021), frequente em 100% dos genomas IBEx sequenciados, foi a primeira descrita a apresentar vantagem sob seleção positiva, estando associada à maior capacidade do vírus em infectar células, aumentando significativamente a infecciosidade e a transmissibilidade do vírus. (KORBER et al., 2020); (HARVEY et al., 2021; JACKSON et al., 2021). Esta mutação permanece frequente inclusive nos dias atuais, em linhagens da variante Ômicron (GISAID, 2023). A associação entre a mutação D614G e a P314L, encontrada em 100% das

amostras sequenciadas neste trabalho, está localizada no gene da RNA polimerase RNA-dependente (RdRp - ORF1b). A primeira mutação é descrita por tornar o local de clivagem da Furina mais acessível, enquanto a segunda, por aumentar a estabilidade da proteína. Juntas, essas mutações foram relatadas como responsáveis por aumentar a entrada e a replicação viral, respectivamente (HADDAD et al., 2021).

Com relação às mutações na proteína do nucleocapsídeo (N) observadas em 2020, verificou-se que as mutações R203K e G204R, presentes simultaneamente no genoma, também foram frequentes em 100% das amostras de 2020, e amplamente relacionadas em estudos científicos, nos anos de 2020 e 2021, ao aumento da replicação e patogênese e também de casos da variante B.1.1.7 (Alpha) mundialmente (JOHNSON et al., 2022). As variantes P.1 e P.2 também foram carreadoras destas mutações, mostrando que a associação dessas mutações conferiu vantagem ao “fitness viral”, facilitando provavelmente o processo de montagem do nucleocapsídeo (WU H et al., 2021). Pesquisadores associaram o aumento da gravidade da COVID-19 mundialmente e a frequência dessas duas mutações, atribuindo-se a elas uma vantagem durante a montagem do nucleocapsídeo. JOHNSON e colaboradores também atribuíram a essa associação de mutações o aumento da replicação, aptidão e patogênese do SARS-CoV-2, destacando, portanto, a importância de mutações na proteína do nucleocapsídeo, não somente as mutações da glicoproteína S, para a disseminação viral durante a pandemia. Outra mutação em N frequente em 90% dos genomas sequenciados em 2020 foi a I292T, que juntamente com a I33T em ORF6 é descrita como marcadoras da variante B.1.1.33 (PANZERA Y et al., 2021). Estas mutações em N não foram verificadas nas variantes Gamma, Delta e Ômicron que vieram a surgir meses depois.

O ano de 2021 foi marcado pelo início da vacinação no Estado do Rio de Janeiro, iniciada em 18 de janeiro, progredindo-se a vacinação por faixa etária ao longo do ano. Ainda assim, foram observados o aumento do número de casos, de acordo com o painel de dados sobre a COVID-19 no Estado do Rio de Janeiro (<https://epirio.svs.rio.br/painel/painel-rio-COVID-19/>) e o avanço das variantes B.1.1.28.2 (P.2 ou Zeta) e B.1.1.28.1 (P.1 ou Gamma), além do surgimento de B.1.617.2 (AY.99.2, Delta), provavelmente devido à dificuldade de manutenção de

regras mais restritivas de circulação de pessoas e distanciamento social e às altas taxas de mutação do vírus. Uma das amostras IBEx sequenciada (Nº 20 - 7842/2021) foi coletada em outubro daquele ano, proveniente de um paciente idoso (85 anos), que faleceu durante a internação (não há informações sobre vacinação) e correspondeu à variante AY.99.2 (Delta), a mais frequente também nos 788 genomas adicionais GISAID-RJ avaliados. Nesta amostra, observou-se a presença de 8 mutações em S, dentre elas a D614G e a L452R, ambas descritas por proporcionarem o aumento da infectividade e transmissibilidade, respectivamente, mas não associadas ao aumento da virulência (HARVEY et al., 2021).

Em virtude do sequenciamento de amostras de 2021 2022 não ter gerado resultados satisfatórios, ampliamos as análises, obtendo mais 788 genomas GISAID-RJ de alta qualidade e cobertura, do período de outubro de 2021, concomitante ao da amostra sequenciada, nos quais verificamos a prevalência de AY.99.2 (88,07%), seguida de AY.99.1 (com apenas 6,98%). Em geral, houve em 2021 um número de mutações bem maior que as amostras e genomas de 2020, o que é esperado em virtude do rápido acúmulo de mutações ao longo do tempo, em vírus de genoma RNA. Desta forma, foi possível caracterizar mutações bem marcantes de Delta e observar a clara diferenciação para as mutações em Gamma.

Naquele ano, P.1 e P.2 reduziram drasticamente suas frequências, sendo substituídas pela linhagem mais prevalente AY.99.2 (88,07% em outubro) como observado na Figura 14 e na Quadro 5, nos resultados. À medida que a população foi sendo vacinada, os governos estaduais foram gradualmente afrouxando as medidas de enfrentamento da propagação do coronavírus. Em 28 de outubro de 2021 a Secretaria de Estado da Saúde do Rio de Janeiro (SES-RJ) publicou a Resolução SES Nº 2499 em que regulamentou a flexibilização do uso de máscaras no estado, orientando que os municípios poderiam desobrigar o uso de máscaras em locais abertos e sem aglomeração quando 65% de toda a sua população ou 75% do público-alvo (indivíduos a partir dos 12 anos de idade) tivessem completado o esquema primário de imunização contra a COVID-19. Logo em seguida, não só o Rio de Janeiro como o país apresentaram um aumento significativo de casos, com a chegada da variante B.1.1.529 (BA, Ômicron) conforme mostram as Figuras 15 e 16 aliada ao relaxamento das medidas do governo e as comemorações de fim de ano e férias escolares. Desde a primeira introdução da variante Ômicron no Brasil, em 30

de novembro de 2021 no estado de São Paulo, observamos nas frequências apresentadas nos genomas GISAID-RJ de novembro de 2021 a janeiro de 2022 (Figuras 15 e 16) que a Ômicron rapidamente substituiu a anteriormente dominante variante Delta. Em outubro de 2021, AY.99.2 representava 88,07% dos genomas. Entre dezembro de 2021 a janeiro de 2022, a frequência caiu para 36,71% (Figura 15). Estes dados puderam ser relacionados ao aumento acentuado dos casos de COVID-19 em janeiro de 2022. Embora tenha sido a maior “onda” no estado em número de casos (459.099), ela foi marcada por um número menor de internações, 4.607, e 1.475 óbitos ([Boletim epidemiológico dos casos de COVID-19 no Estado do Rio de Janeiro](#)), refletindo a efetividade da vacinação contra COVID-19. LAMARCA et al., 2023 relataram a mesma ocorrência avaliando as frequências de novembro de 2021 (94,0% Delta e 6% Ômicron) a Janeiro (mais de 99% Ômicron). O total de 26 diferentes tipos de linhagens (Figura 15) de SARS-CoV-2 foram observadas dentre os genomas GISAID-RJ, dentre eles, genomas Gamma, Delta e Ômicron. A variedade de linhagens circulante, aliada mais uma vez ao período de férias escolares, relaxamento de medidas de prevenção e do uso de máscaras e o alto número de casos no estado pode ter facilitado o surgimento de linhagens recombinantes como a XZ (Figura e Quadro 13), que exibe semelhanças significativas com BA.2 (NIVEDITHA et al., 2023), linhagem que provavelmente circulava no estado, mas não estava presente no universo de genomas GISAID de alta qualidade sequenciados. As linhagens recombinantes podem combinar as mutações vantajosas das linhagens parentais, representando um risco de geração de variantes altamente infecciosas (OU et al., 2022). Linhagens recombinantes são mais comuns quando o número de casos é alto, como ocorreu com a chegada da Ômicron, facilitando o evento de coinfeção. Torna-se, portanto, extremamente importante o controle não só do número de casos, mas da diversidade de linhagens circulando ao mesmo tempo. (LAMARCA et al., 2023). As linhagens da variante Ômicron seguiram a tendência de maior número de substituições de nucleotídeos e aminoácidos que as variantes até então existentes. Os genes S e ORF1a passaram a ser os de maior número de mutações. Somente em S, foram observadas 29 novas mutações, como pôde ser observado na Quadro 12. Foram “herdadas” de Gamma as mutações N501Y e H655Y, e de Delta a T478K. D614G, existente desde a linhagem B.1.1.33, manteve-se em Ômicron em 2021/2022.

Inserções de aminoácidos foram pouco observadas no total de amostras e genomas avaliados (Quadro 14). Deleções importantes foram relatadas nas amostras sequenciadas (Figura 21 e Quadro 15). Alguns grupos/conjuntos de deleções (Quadro 15), característicos de cada linhagem, fizeram parte de genomas de AY.99.2 (Delta) e P.1 (Gamma) em 2021, das linhagens BA.1 e BA.1.1 (Ômicron) analisadas do período compreendido entre Dezembro de 2021 e Janeiro de 2022, bem como de XBB.1.5 e XBB.1.5.17 no 1º semestre de 2023. Algumas deleções mostraram-se favoráveis às linhagens, como as deleções em S E156_ e F157_ em AY.99.2, que foram descritas por MISHRA e colaboradores como específicas do domínio N-terminal (NTD) e que contribuíram para aumentar a infectividade e reduzir a sensibilidade aos anticorpos induzidos por vacinas. Outra deleção em S, a Y144_, presente em XBB.1.5 e XBB.1.5.17 (dentre as três mais frequentes de 2023), contribuiu para reduzir a afinidade da ligação dos anticorpos ao vírus, facilitando a evasão da resposta imune (SAHA et al., 2021).

Os genomas GISAID 2023 foram também avaliados, buscando-se obter um panorama mais atual das linhagens circulantes e mutações. No período de janeiro a junho, foi observado que todas as linhagens circulantes pertenciam à variante Ômicron (Figura 17), ressaltando-se que as 10 linhagens que surgiram entre dezembro de 2021 e janeiro de 2022 não permaneceram em 2023 (Figura 18), dando lugar a 36 novas linhagens, muitas delas recombinantes, mesmo com a queda do número de casos e mortes no estado e no país. As mutações em S continuaram a ocorrer com bastante frequência. Neste universo, foram observadas 40 novas mutações em comparação às linhagens de SARS-CoV-2 anteriormente existentes (Quadro 12). Nossos dados e o curso da pandemia mostraram que a proteína S demonstrou alta capacidade de sofrer mutações ao longo dos anos, sem perda de sua função, apresentando particular plasticidade estrutural e funcional (TELENTI et al., 2021).

Em maio de 2023, a OMS anunciou mudanças no status da COVID-19, que deixou de ser classificada como emergência de saúde pública de interesse internacional. A pandemia, entretanto, não terminou, uma vez que a disseminação do vírus se mantém de forma global e com transmissão sustentada. A tendência atual é que a pandemia continue através de pequenos aumentos no número de casos, causados por uma redução da imunidade coletiva (TELENTI, 2021). Autores

ressaltam que crescimento cíclico e a redução do tamanho da população de SARS-CoV-2 são especialmente preocupantes, pois podem facilmente fixar novas mutações e gerar novas variantes tanto por pressão seletiva quanto por deriva genética (TELENTI et al., 2021).

As mutações observadas nas amostras sequenciadas e nos genomas de 2020 a 2023 mostram claramente o processo de evolução natural do vírus. A cada nova variante introduzida e disseminada, foram observadas mutações que conferiram vantagens na infectividade ou no escape à resposta imune. Nossos resultados mostraram que a pandemia no estado caracterizada por fases que foram impulsionadas pelo aumento do número de mutações ao longo dos anos de 2020 a 2023, muitas dessas estudadas pela comunidade científica, que se mostraram importantes no aumento da infectividade, transmissibilidade ou no escape do vírus à resposta imune. As regiões S, ORF1a, ORF1b e N mantiveram-se desde o início da pandemia como as quatro mais frequentes em número de mutações (Quadro 8). O gene E, ao contrário, manteve-se conservado durante os primeiros anos (Quadro 8), passando a apresentar a mutação T9I em 99,2% dos genomas da variante Ômicron coletados em 2023. (MOU et al., 2021). Estas variadas mutações levaram a uma substituição sucessiva de linhagens, que foram se adaptando e sustentando sua transmissão, sendo a primeira fase iniciada com a chegada da variante B.1.1.33 (mutação D614G) em 2020, observada nas amostras sequenciadas. Infelizmente, as frequências de P.2 e P.1 não puderam ser observadas nos períodos avaliados (P.2 predominou entre novembro de 2020 e janeiro de 2021, seguida de P.1 a partir de março de 2021 (DA SILVA et al., 2021). Em outubro de 2021, com o Estado em sua quarta “onda” e vários países do mundo enfrentando novas “ondas” da COVID-19, observamos 14 sublinhagens Delta nos genomas analisados (Figura 14). Dentre elas, AY.99.2 (encontrada em nossa amostra e a mais frequente nos genomas GISAID) e AY.101 (com frequência de apenas 0,38% nos genomas de outubro de 2021) encontravam-se disseminadas de forma global (ROMANO et al., 2022). Nossos dados estão de acordo com os de ROMANO et al., 2022, onde a prevalência de AY.99.2 também foi a maior dentre as sublinhagens Delta.

A vigilância genômica e a identificação e o estudo de novas variantes são

cruciais no monitoramento da propagação dos vírus circulantes, sendo também útil na melhoria de técnicas de diagnóstico, produção de novas vacinas e tratamentos contra o vírus da COVID-19. Embora a vacinação tenha resultado no declínio da infecção por COVID-19 e de hospitalizações em todo o mundo, numerosos estudos demonstraram que a eficácia das vacinas contra a infecção diminuiu com o tempo (ABBASIAN et al., 2023). Dado o vasto território brasileiro e a constatação da importância econômica do Rio de Janeiro e da região sudeste, com três dos estados mais populosos do Brasil (IBGE, 2023) e também devido a concentração de comércio, indústria e serviços, ressalta-se a necessidade de vigilância genômica, uma vez que a tendência natural é que estas regiões, mais populosas e com conexões econômicas nacionais e internacionais sejam as principais disseminadoras de futuras linhagens.

Mesmo com a diminuição de casos, mortes e severidade da COVID-19, novas variantes continuarão a surgir no Brasil e no mundo e muitas mutações continuarão a ser descritas. Dados recentes de filodinâmica disponíveis na plataforma GISAID em 24 de setembro de 2023 (<https://gisaid.org/phylogenomics/brazil/>) mostram que, atualmente, muito embora tenhamos prevalência da variante Ômicron no Brasil, podemos observar que ela está representada por uma altíssima diversidade de linhagens circulantes, muitas delas recombinantes, reforçando a necessidade do contínuo sequenciamento genômico seguido por estudos, como o realizado neste trabalho, para o constante rastreamento de novas variantes de preocupação, bem como, para o maior entendimento da dinâmica de transmissão do SARS-CoV-2, a resposta imunológica, permitindo o desenvolvimento de novas vacinas e antivirais.

REFERÊNCIAS

ABBASIAN, Mohammad Hadi et al. Global landscape of SARS-CoV-2 mutations and conserved regions. **Journal of Translational Medicine**, v. 21, n. 1, p. 152, 2023.

AKSAMENOV, Ivan et al. Nextclade: clade assignment, mutation calling and quality control for viral genomes. **Journal of open source software**, v. 6, n. 67, p. 3773, 2021.

AL-ZYOUD, Walid; HADDAD, Hazem. Dynamics prediction of emerging notable spike protein mutations in SARS-CoV-2 implies a need for updated vaccines. **Biochimie**, v. 191, p. 91-103, 2021.

BARNES, Christopher O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. **Nature**, v. 588, n. 7839, p. 682-687, 2020.

BORGES, Alessandra A. et al. SARS-CoV-2: origem, estrutura, morfogênese e transmissão. **BARRAL-NETTO, M.; BARRETO, ML; PINTO JUNIOR, EP**, 2020.

CANDIDO, Darlan S. et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. **Science**, v. 369, n. 6508, p. 1255-1260, 2020.

CEVIK, Muge et al. Virology, transmission, and pathogenesis of SARS-CoV-2. **bmj**, v. 371, 2020.

CHIARA, Matteo et al. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. **Briefings in Bioinformatics**, v. 22, n. 2, p. 616-630, 2021.

CIPITELLI, Marcio da Costa et al. SARS-CoV-2 diagnostic diary: from rumors to the first case. Early reports of molecular tests from the military research and diagnostic institute of Rio de Janeiro. **Memórias do Instituto Oswaldo Cruz**, v. 115, 2020.

CORONA-ÔMICABR, R. 3º Relatório técnico sobre a circulação de variantes do SARS-CoV-2 no Brasil Projeto Corona-Ômica-BR. , [s.d.].

COUTARD, Bruno et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. **Antiviral research**, v. 176, p. 104742, 2020.

DA SILVA FRANCISCO JUNIOR, Ronaldo et al. Turnover of SARS-CoV-2 lineages shaped the pandemic and enabled the emergence of new variants in the state of Rio de Janeiro, Brazil. **Viruses**, v. 13, n. 10, p. 2013, 2021.

GORBALENYA et al., 2020 The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **Nat Microbiol.** 2020 Apr;5(4):536-544. doi: 10.1038/s41564-020-0695-z. Epub 2020 Mar 2. PMID: 32123347; PMCID: PMC7095448.

GORDON, David E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. **Nature**, v. 583, n. 7816, p. 459-468, 2020.

HARVEY, William T. et al. SARS-CoV-2 variants, spike mutations and immune escape. **Nature Reviews Microbiology**, v. 19, n. 7, p. 409-424, 2021.

HOFFMANN, Markus et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. **cell**, v. 181, n. 2, p. 271-280. e8, 2020.

HUI, Pei. Next generation sequencing: chemistry, technology and applications. **Chemical Diagnostics: From Bench to Bedside**, p. 1-18, 2014.

ILMJÄRV, Sten et al. Concurrent mutations in RNA-dependent RNA polymerase and spike protein emerged as the epidemiologically most successful SARS-CoV-2 variant. **Scientific reports**, v. 11, n. 1, p. 13705, 2021.

FLOWER, Thomas G. et al. Structure of SARS-CoV-2 ORF8, a rapidly evolving coronavirus protein implicated in immune evasion. **Biorxiv**, 2020.

FONANGER J. et al. **Working paper on SARS-CoV-2 spike mutations arising in Danish mink, their spread to humans and neutralization data.** https://files.ssi.dk/Mink-cluster-5-short-report_AFO2 (2020)

FUJINO, Takahisa et al. Novel SARS-CoV-2 variant in travelers from Brazil to Japan. **Emerging infectious diseases**, v. 27, n. 4, p. 1243, 2021.

GE, Xing-Yi et al. Detection of alpha-and betacoronaviruses in rodents from Yunnan, China. **Virology journal**, v. 14, p. 1-11, 2017.

GIBSON-FALL, Fawzia. Military responses to COVID-19, emerging trends in global civil-military engagements. **Review of International Studies**, v. 47, n. 2, p. 155-170, 2021.

GIOVANETTI, Marta et al. Evolution patterns of SARS-CoV-2: Snapshot on its genome variants. **Biochemical and biophysical research communications**, v. 538, p. 88-91, 2021.

GLOBAL INITIATIVE ON SHARING INFLUENZA DATA – phylodynamics
<https://gisaid.org/phylogenetics/brazil/> (acesso em 11/03/2023, com informações de 3331 genomas coletados entre fevereiro de 2020 e fevereiro de 2023)

GOEL, Rishi R. et al. mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. **Science**, v. 374, n. 6572, p. abm0829, 2021.

GREANEY, Allison J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. **Cell host & microbe**, v. 29, n. 3, p. 463-476. e6, 2021.

GROSCHE, Victória Riquena et al. Insights on the SARS-CoV-2 genome variability: the lesson learned in Brazil and its impacts on the future of pandemics. **Microbial Genomics**, v. 7, n. 11, 2021.

GRUDLEWSKA-BUDA, Katarzyna et al. SARS-CoV-2—morphology, transmission and diagnosis during pandemic, review with element of meta-analysis. **Journal of Clinical Medicine**, v. 10, n. 9, p. 1962, 2021.

GUSEV, Evgenii et al. SARS-CoV-2-specific immune response and the pathogenesis of COVID-19. **International journal of molecular sciences**, v. 23, n. 3, p. 1716, 2022.

HADDAD, Dania et al. SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. **PloS one**, v. 16, n. 5, p. e0251368, 2021.

HASSAN, Sk Sarif; CHOUDHURY, Pabitra Pal; ROY, Bidyut. SARS-CoV2 envelope protein: non-synonymous mutations and its consequences. **Genomics**, v. 112, n. 6, p. 3890-3892, 2020.

HELMY, Yosra A. et al. The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. **Journal of clinical medicine**, v. 9, n. 4, p. 1225, 2020.

HOFFMANN, Markus et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. **cell**, v. 181, n. 2, p. 271-280. e8, 2020.

HU, Taishan et al. Next-generation sequencing technologies: An overview. **Human Immunology**, v. 82, n. 11, p. 801-811, 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Portal das cidades. Disponível em: <https://cidades.ibge.gov.br/brasil/rj/pesquisa/48/48986?tipo=ranking&indicador=48985&ano=2020>, acesso em: 05 set 2023.

ITO, Jumpei et al. Convergent evolution of SARS-CoV-2 Ômicron subvariants leading to the emergence of BQ. 1.1 variant. **Nature communications**, v. 14, n. 1, p. 2671, 2023.

JACKSON, Cody B. et al. Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. **Biochemical and Biophysical Research Communications**, v. 538, p. 108-115, 2021.

JAIN, Miten et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. **Nature biotechnology**, v. 36, n. 4, p. 338-345, 2018.

JOHNSON, Bryan A. et al. Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis. **PLoS pathogens**, v. 18, n. 6, p. e1010627, 2022.

KE, Rongqin et al. Fourth generation of next-generation sequencing technologies: Promise and consequences. **Human mutation**, v. 37, n. 12, p. 1363-1367, 2016.

KEMP, Steven A. et al. SARS-CoV-2 evolution during treatment of chronic infection. **Nature**, v. 592, n. 7853, p. 277-282, 2021.

KIM, Dongwan et al. The architecture of SARS-CoV-2 transcriptome. **Cell**, v. 181, n. 4, p. 914-921. e10, 2020.

KLINAKIS, Apostolos; COURNIA, Zoe; RAMPIAS, Theodoros. N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains. **Computational and structural biotechnology journal**, v. 19, p. 5556-5567, 2021.

KONNO, Yoriyuki et al. SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. **Cell reports**, v. 32, n. 12, 2020.

KORBER, Bette et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. **Cell**, v. 182, n. 4, p. 812-827. e19, 2020.

KUMAR, Swatantra et al. Morphology, genome organization, replication, and pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In: **Coronavirus Disease 2019 (COVID-19)**. Springer, Singapore, 2020. p. 23-31.

LAMARCA, Alessandra Pavan et al. Phylodynamic analysis of SARS-CoV-2 spread in Rio de Janeiro, Brazil, highlights how metropolitan areas act as dispersal hubs for new variants. **Microbial Genomics**, v. 8, n. 9, 2022.

LEI, Xiaobo et al. Activation and evasion of type I interferon responses by SARS-CoV-2. **Nature communications**, v. 11, n. 1, p. 3810, 2020.

LI, Meng et al. Comparative susceptibility of SARS-CoV-2, SARS-CoV, and MERS-CoV across mammals. **The ISME Journal**, v. 17, n. 4, p. 549-560, 2023.

LI, Xiaowei et al. Molecular immune pathogenesis and diagnosis of COVID-19. **Journal of pharmaceutical analysis**, v. 10, n. 2, p. 102-108, 2020.

LIU, GuanQun; GACK, Michaela U. Distinct and orchestrated functions of RNA sensors in innate immunity. **Immunity**, v. 53, n. 1, p. 26-42, 2020

LIU, Lihong et al. Striking antibody evasion manifested by the Ômicron variant of SARS-CoV-2. **Nature**, v. 602, n. 7898, p. 676-681, 2022.

LIU, Lin et al. Comparison of next-generation sequencing systems. **Journal of Biomedicine and Biotechnology**, v. 2012, 2012.

LIU, Yen-Chin; KUO, Rei-Lin; SHIH, Shin-Ru. COVID-19: The first documented coronavirus pandemic in history. **Biomedical journal**, v. 43, n. 4, p. 328-333, 2020.

LIU, Zhuoming et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. **Cell host & microbe**, v. 29, n. 3, p. 477-488. e4, 2021.

LU, Hengyun; Giordano, Francesca; Ning, Zemin. Oxford Nanopore MinION sequencing and genome assembly. **Genomics, proteomics & bioinformatics**, v. 14, n. 5, p. 265-279, 2016.

LU, Roujian et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. **The lancet**, v. 395, n. 10224, p. 565-574, 2020.

MALIK, Yasmin A. Properties of coronavirus and SARS-CoV-2. **The Malaysian journal of pathology**, v. 42, n. 1, p. 3-11, 2020.

MÁRQUEZ, Sully et al. SARS-CoV-2 genome sequencing from COVID-19 in Ecuadorian patients: a whole country analysis. **MedRxiv**, 2021.

MARTY, Francisco M.; CHEN, Kaiwen; VERRILL, Kelly A. How to obtain a nasopharyngeal swab specimen. **The New England journal of medicine**, v. 382, n. 22, p. e76-e76, 2020.

MATSUYAMA, Shutoku; TAGUCHI, Fumihiro. Two-step conformational changes in a coronavirus envelope glycoprotein mediated by receptor binding and proteolysis. **Journal of virology**, v. 83, n. 21, p. 11133-11141, 2009.

MCCOMBIE, W. Richard; MCPHERSON, John D.; MARDIS, Elaine R. Next-generation sequencing technologies. **Cold Spring Harbor perspectives in medicine**, v. 9, n. 11, 2019.

MCMAHAN, Katherine et al. Correlates of protection against SARS-CoV-2 in rhesus macaques. **Nature**, v. 590, n. 7847, p. 630-634, 2021.

MCNAB, Finlay et al. Type I interferons in infectious disease. **Nature Reviews Immunology**, v. 15, n. 2, p. 87-103, 2015.

MENEZES, Diego et al. SARS-CoV-2 Genomic Surveillance in Brazil: A Systematic Review with Scientometric Analysis. **Viruses**, v. 14, n. 12, p. 2715, 2022.

METZKER, Michael L. Sequencing technologies—the next generation. **Nature reviews genetics**, v. 11, n. 1, p. 31-46, 2010.

MICHELON, Cleonice Maria. Main SARS-CoV-2 variants notified in Brazil. **Rev. Bras. Anal. Clin.**, v. 53, p. 109-116, 2021.

MINISTÉRIO DA SAÚDE DO BRASIL. Site oficial do Ministério da Saúde. Disponível em: <<https://COVID.saude.gov.br>>, acesso em 06/02/23.

MIORIN, Lisa et al. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. **Proceedings of the National Academy of Sciences**, v. 117, n. 45, p. 28344-28354, 2020.

MISHRA, Tarun et al. SARS-CoV-2 spike E156G/Δ157-158 mutations contribute to increased infectivity and immune escape. **Life Science Alliance**, v. 5, n. 7, 2022.

MORADI, Jale et al. Variation analysis of SARS-CoV-2 complete sequences from Iran. **Future virology**, v. 17, n. 12, p. 863-872, 2022.

MOTOZONO, Chihiro et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. **Cell host & microbe**, v. 29, n. 7, p. 1124-1136. e11, 2021.

MOU, Kejie et al. Emerging mutations in envelope protein of SARS-CoV-2 and their effect on thermodynamic properties. **Informatics in medicine unlocked**, v. 25, p. 100675, 2021.

NEMUDRYI, Artem et al. SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. **Cell Reports**, v. 35, n. 9, 2021.

OU, Junxian et al. Tracking SARS-CoV-2 Ómicron diverse spike gene mutations identifies multiple inter-variant recombination events. **Signal Transduction and Targeted Therapy**, v. 7, n. 1, p. 138, 2022.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). Site oficial da OMS. Disponível em: <<http://COVID19.who.int/table>>, acesso em 13/09/2023.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE., Folha informativa sobre a COVID-19, disponível em <<https://www.paho.org/pt/COVID19>>, acesso em 20/06/23)

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **Sequenciamento genômico do SARS-CoV-2. Guia de implementação para máximo impacto na saúde pública. 8 de janeiro de 2021.** [s.l.] Pan American Health Organization, 2021.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **Atualização epidemiológica: Variantes do SARS-CoV-2 nas Américas.**[s.l: s.n.], OPAS, 24 DE MARÇO DE 2021. Disponível em: <http://www.paho.org/>, acesso em 20 de março de 2023

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **Update on the emergence of SARS-CoV-2 Ómicron sublineages**[s.l: s.n.], OPAS, 22 DE AGOSTO DE 2023. Disponível em: <https://www.paho.org/en/documents/update-emergence-sars-cov-2-Ómicron-sublineages-22-august-2023>, acesso em 17 de setembro de 2023

Painel Saúde RJ. Monitoramento da COVID-19 no Estado do Rio de Janeiro. Disponível em: <<https://painel.saude.rj.gov.br/monitoramento/COVID19.html#>>, acesso em: 04/06/23.

PAN AMERICAN HEALTH ORGANIZATION - Weekly COVID-19 Epidemiological Update - Region of the Americas Issue 57, published 7 March 2023 Disponível em: <<https://www.paho.org/en/documents/paho-weekly-COVID-19-epidemiological-update-ew9-7-march-2023>>, acesso em 15/03/23.

PANZERA, Yanina et al. A deletion in SARS-CoV-2 ORF7 identified in COVID-19 outbreak in Uruguay. **Transboundary and emerging diseases**, v. 68, n. 6, p. 3075-3082, 2021.

PANZERA, Yanina et al. Transmission cluster of COVID-19 cases from Uruguay: emergence and spreading of a novel SARS-CoV-2 ORF6 deletion. **Memórias do Instituto Oswaldo Cruz**, v. 116, p. e210275, 2022.

PATER, Adrian A. et al. High throughput nanopore sequencing of SARS-CoV-2 viral genomes from patient samples. **Journal of biological methods**, v. 8, n. COVID 19 Spec Iss, 2021.

Plataforma ArcGIS. Site oficial da Plataforma ArcGIS. Disponível em: <<https://experience.arcgis.com/experience/38efc69787a346959c931568bd9e2cc4>>, acesso em 24/09/2023.

PICCOLI, Luca et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. **Cell**, v. 183, n. 4, p. 1024-1042. e21, 2020.

PINHEIRO; PATUSCO; VALVERDE. A atuação do Exército Brasileiro durante a pandemia do COVID-19. **Observatório Militar da Praia Vermelha**. ECEME: Rio de Janeiro. 2022.

PLANTE, Jessica A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. **Nature**, v. 592, n. 7852, p. 116-121, 2021.

QUICK, Joshua et al. Real-time, portable genome sequencing for Ebola surveillance. **Nature**, v. 530, n. 7589, p. 228-232, 2016.

QUICK, Josh. Amplicon clean-up using SPRI “beads”. 2020.

QUICK, Josh. DNA quantification using the Quantus fluorometer. 2020.

QUICK, Josh. Ebola virus sequencing protocol. 2019.

QUICK, Josh. nCoV-2019 sequencing protocol v3 (LoCost). 2020.

QUICK, Josh. One-pot native barcoding of amplicons. 2019.

QUICK, Josh et al. Starting a MinION sequencing run using MinKNOW. 2019.

RASTOGI, Meghana et al. SARS coronavirus 2: from genome to infectome. **Respiratory research**, v. 21, n. 1, p. 1-15, 2020.

RESENDE, Paola Cristina, et al. Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B. 1.1. 33 during the early pandemic phase in Brazil. **Frontiers in microbiology**, 2021, 11: 3565.

RESENDE, Paola Cristina, et al. A potential SARS-CoV-2 variant of interest (VOI) harboring mutation E484K in the Spike protein was identified within lineage B. 1.1. 33 circulating in Brazil. **Viruses**, 2021, 13.5: 724.

RESENDE, Paola Cristina et al. Identification of a new B. 1.1. 33 SARS-CoV-2 Variant of Interest (VOI) circulating in Brazil with mutation E484K and multiple deletions in the amino (N)-terminal domain of the Spike protein. **Preprint from Virological**. < <https://virological.org/t/identification-of-a-new-b-1-1-33-sars-cov-2-variant-of-interest-circulating-in-brazil-with-mutation-e484k-and-multiple-deletions-in-the-amino-n-terminal-domain-of-the-spike-protein/675> > accessed, v. 14, 2021.

Rio de Janeiro (Estado). Secretaria de Estado de Saúde. Boletim Epidemiológico dos Casos de COVID-19 no Estado do Rio de Janeiro 2020 – 2021. Rio de Janeiro: SES-RJ, 2022. Disponível em : < https://painel.saude.rj.gov.br/arquivos/Boletim_2020_2021_2611.pdf >, acesso em: 26/052/2023.

ROMANO, Camila Malta et al. Early Emergence and Dispersal of Delta SARS-CoV-2 Lineage AY. 99.2 in Brazil. **Frontiers in Medicine**, v. 9, p. 930380, 2022.

RONAGHI, Mostafa; UHLÉN, Mathias; NYRÉN, Pål. A sequencing method based on real-time pyrophosphate. **Science**, v. 281, n. 5375, p. 363-365, 1998.

RUSK, Nicole. Torrents of sequence. **Nature Methods**, v. 8, n. 1, p. 44-44, 2011.

SAHA, Indrajit et al. Hotspot mutations in SARS-CoV-2. **Frontiers in Genetics**, v. 12, p. 753440, 2021.

SALLAS, Janaína et al. A vigilância genômica do SARS-CoV-2 no Brasil na resposta à pandemia da COVID-19. **Revista Panamericana de Salud Pública**, v. 45, p. e75, 2021.

SANTOS, Ana Alice Reis et al. Diferentes aspectos da resposta imunológica na COVID-19. **Research, Society and Development**, v. 10, n. 14, p. e423101422531-e423101422531, 2021.

SCHUSTER, Stephan C. Next-generation sequencing transforms today's biology. **Nature methods**, v. 5, n. 1, p. 16-18, 2008.

SHEMESH, Maya et al. SARS-CoV-2 suppresses IFN β production mediated by NSP1, 5, 6, 15, ORF6 and ORF7b but does not suppress the effects of added interferon. **PLoS pathogens**, v. 17, n. 8, p. e1009800, 2021.

SHU, Yuelong; MCCAULEY, John. GISAID: Global initiative on sharing all influenza data—from vision to reality. **Eurosurveillance**, v. 22, n. 13, p. 30494, 2017.

SILVA, Rowersan Cabral; LIMA, Airton; DA SILVA SOUZA, Liliane Cristina. Principais métodos de sequenciamento de DNA. **Scientific Electronic Archives**, v. 15, n. 10, 2022.

SIMAS, Maria Clara da Costa et al. Evaluation of SARS-CoV-2 ORF7a Deletions from COVID-19-Positive Individuals and Its Impact on Virus Spread in Cell Culture. **Viruses**, v. 15, n. 3, p. 801, 2023.

SLATKO, Barton E.; GARDNER, Andrew F.; AUSUBEL, Frederick M. Overview of next-generation sequencing technologies. **Current protocols in molecular biology**, v. 122, n. 1, p. e59, 2018.

STARR, Tyler N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. **cell**, v. 182, n. 5, p. 1295-1310. e20, 2020.

STOKEL-WALKER, Chris et al. What do we know about the adaptive immune response to COVID-19?. **bmj**, v. 380, 2023.

SUBRAMANIAN, Anuradhaa et al. Symptoms and risk factors for long COVID in non-hospitalized adults. **Nature medicine**, v. 28, n. 8, p. 1706-1714, 2022.

The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **Nature microbiology**, v. 5, n. 4, p. 536-544, 2020.

TIMMERS, Luis Fernando Saraiva Macedo et al. SARS-CoV-2 mutations in Brazil: from genomics to putative clinical conditions. **Scientific reports**, v. 11, n. 1, p. 11998, 2021.

THOMS, Matthias et al. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. **Science**, v. 369, n. 6508, p. 1249-1255, 2020.

THOMSON, Emma C. et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. **Cell**, v. 184, n. 5, p. 1171-1187. e20, 2021.

TURNER, Jackson S. et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. **Nature**, v. 596, n. 7870, p. 109-113, 2021.

TYSON, John R. et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. **BioRxiv**, 2020.

V'KOVSKI, Philip et al. Coronavirus biology and replication: implications for SARS-CoV-2. **Nature Reviews Microbiology**, v. 19, n. 3, p. 155-170, 2021.

WORLD HEALTH ORGANIZATION. Tracking SARS-CoV-2 variants. Disponível em: <<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>>, acesso em 05 de maio de 2023

WORLD HEALTH ORGANIZATION. XBB.1.5 Updated Rapid Risk Assessment, 24 February 2023. Disponível em:
<<https://www.who.int/docs/defaultsource/coronaviruse/22022024xbb.1.5ra.pdf>,
acesso em 20/04/23.

WROBEL, Antoni G. et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. **Nature structural & molecular biology**, v. 27, n. 8, p. 763-767, 2020.

V'KOVSKI, Philip et al. Coronavirus biology and replication: implications for SARS-CoV-2. **Nature Reviews Microbiology**, v. 19, n. 3, p. 155-170, 2021.

VOLOCH, Carolina M. et al. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. **Journal of virology**, v. 95, n. 10, p. 10.1128/jvi.00119-21, 2021.

WICK, Ryan R. et al. Completing bacterial genome assemblies with multiplex MinION sequencing. **Microbial genomics**, v. 3, n. 10, 2017.

WORLD HEALTH ORGANIZATION et al. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health, 8 January 2021. 2021.

WORLD HEALTH ORGANIZATION. (2020). Coronavirus disease 2019 (COVID-19): situation report, 73.

WU, Haibo et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. **Cell host & microbe**, v. 29, n. 12, p. 1788-1801. e6, 2021.

XIA, Hongjie et al. Evasion of type I interferon by SARS-CoV-2. **Cell reports**, v. 33, n. 1, 2020.

YUEN, Chun-Kit et al. SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. **Emerging microbes & infections**, v. 9, n. 1, p. 1418-1428, 2020.

ZANDI, Milad et al. The role of SARS-CoV-2 accessory proteins in immune evasion. **Biomedicine & Pharmacotherapy**, p. 113889, 2022.

ZHOU, Ziliang et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14⁺ monocytes. **IScience**, v. 24, n. 3, 2021.

ANEXOS

ANEXO A - Termo de Consentimento Livre e Esclarecido



**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
CML - 1ª R M
INSTITUTO DE BIOLOGIA DO EXÉRCITO
IBEx (Lab Microscopia Clin e Bacteriol./1894)**

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (Elaborado em 07/10/2022)

Título do projeto de pesquisa: “Análise genômica de SARS-COV-2 no Estado do Rio de Janeiro em amostras diagnosticadas no Instituto de Biologia do Exército e em genomas depositados no GISAID, durante a pandemia da COVID-19”

Você está sendo convidado a participar de um estudo que irá realizar o sequenciamento do genoma do vírus SARS-CoV-2, causador da COVID-19, em amostras positivas de militares do Exército Brasileiro e seus dependentes que foram diagnosticadas no Instituto de Biologia do Exército (IBEx), para rastrear possíveis mutações genéticas de interesse à saúde pública e contribuir para a melhor compreensão do processo evolutivo do vírus. A pesquisadora responsável Dra. Nádia Vaez Gonçalves da Cruz, e a pesquisadora Aline Lima de Moura Natarelli, ambas oficiais farmacêuticas servindo no Instituto de Biologia do Exército, pretendem realizar um estudo com as seguintes características:

Objetivos do estudo:

- Realizar a vigilância genômica de SARS-CoV-2, por meio do sequenciamento rápido do genoma completo do vírus, utilizando a plataforma de sequenciamento MinION (um equipamento produzido pela empresa Oxford Nanopore Technologies, que é conectado a um computador para fornecer dados das sequências genéticas dos vírus) em amostras positivas para SARS-Cov-2

diagnosticadas no IBEx.

- Rastrear possíveis mutações genéticas nos genomas sequenciados, verificar a classificação das variantes nas amostras selecionadas, bem como verificar informações epidemiológicas junto aos pacientes que concordarem em participar da pesquisa, de forma a confrontar o perfil genético do vírus com o quadro clínico.

- Realizar análises filogenéticas (análises para determinar as relações de origem entre os vírus nas amostras), para que seja possível a avaliar a evolução dos vírus nas amostras sequenciadas.

Benefícios aos participantes e à sociedade:

Este estudo permitirá que o participante tenha conhecimento sobre as características do vírus SARS-Cov-2 que o infectou. Proporcionará, ainda, à sociedade, ia possam ampliar as informações sobre a diversidade genética e circulação de variantes virais na população-alvo do estudo, contribuindo para a vigilância da COVID-19 em nosso país.

Descrição dos procedimentos para coleta da amostra:

Será realizada a coleta para a realização do exame de PCR. Nesta coleta, haverá a introdução de um swab (uma espécie de cotonete estéril que serve para coleta de exames microbiológicos com a finalidade de estudos clínicos ou pesquisa) nos dois orifícios nasais do paciente. Haverá somente o desconforto da introdução do swab no nariz. Não há riscos para o paciente, nem na coleta da amostra nem no sequenciamento da amostra. O material coletado não será armazenado. Imediatamente após a análise e realização do estudo, o material será descartado de acordo com as normas vigentes.

Garantia de liberdade:

Sua participação é totalmente **voluntária**. Você pode escolher não participar ou desistir a qualquer momento. Além disso, ninguém o constrangerá ou o obrigará a participar do estudo, sendo a participação, subordinada apenas a sua escolha individual.

Direito de confidencialidade e acessibilidade:

Cada participante poderá ter acesso, a qualquer momento, ao seu próprio resultado. As informações relacionadas ao estudo são tratadas como confidenciais e servirão para a elaboração de literatura científica. No entanto, se qualquer informação for divulgada em relatório ou publicação, isto será feito sob a forma codificada, sem aparecer nenhum nome, para que o sigilo seja mantido.

Despesas e compensações:

A sua participação no estudo não terá custo algum. As despesas porventura existentes serão de responsabilidade das pesquisadoras. Pela sua participação no estudo, você também não receberá nenhum valor ou compensação financeira.

Garantia de acesso aos pesquisadores:

Em qualquer fase do estudo você terá pleno acesso as pesquisadoras:

Dra. Nádia Vaez Gonçalves da Cruz (pesquisadora responsável): pelo telefone (21) 99179-5859; Divisão de Ensino e Pesquisa, Instituto de Biologia do Exército - IBEx. Rua Francisco Manoel, 102. Benfica; e-mail: nadiavaez@hotmail.com; pesquisadora Aline Lima de Moura Natarelli: pelo telefone (21) 98161-0031 ou e-mail: alinelimamn@gmail.com

Havendo necessidade de fazer alguma reclamação ou denunciar qualquer abuso ou improbidade desta pesquisa, será possível, ainda, entrar em contato com o Comitê de Ética do Centro de Capacitação Física do Exército (CEP-CCFEx), Av João Luiz Alves s/nº, Sala do CEP-CCFEx, Prédio da EsEFEx - Complexo Desportivo Sylvio Magalhães Padilha - Rio de Janeiro, RJ, ou pelo telefone (21)2586-2297, às segundas e quartas-feiras, de 9h às 12h, ou através do e-mail: l: cep@ccfex.eb.mil.br

**Este documento tem 2 vias e o Sr tem direito a receber uma das vias.
O Sr tem direito a tempo para tomar a decisão de participar ou não desta
pesquisa, assim como tem direito à indenização de danos causados pela
participação na pesquisa**

CONSENTIMENTO PÓS INFORMAÇÃO

_____, ____ / ____ / ____.

Eu, _____,
portador da Carteira de Identidade nº _____ li o texto acima e compreendi a natureza e o objetivo do estudo do qual fui convidado a participar. A explicação que recebi menciona os riscos e benefícios do estudo. Eu entendi que sou livre para interromper minha participação no estudo a qualquer momento sem justificar minha decisão, e sei que qualquer problema relacionado será livre de custos para mim. Eu concordo voluntariamente em participar deste estudo e receberei uma cópia desse Termo de Consentimento Livre e Esclarecido (TCLE) e a outra ficará com o pesquisador responsável por essa pesquisa.

Assinatura do voluntário

_____ Idade _____

Assinatura do pesquisador

Contato do Pesquisador responsável:

Dra. Nádia Vaez Gonçalves da Cruz: (21) 99179-5859; e-mail:
nadiavaez@hotmail.com

ANEXO B – Esquema de *primers* v4 Artic Network

Nr sequencial	Ref	início	fim	Left /right	Pool	Sentido
1	MN908947.3	25	50	SARS-CoV-2_1_LEFT	1	+
2	MN908947.3	408	431	SARS-CoV-2_1_RIGHT	1	-
3	MN908947.3	324	344	SARS-CoV-2_2_LEFT	2	+
4	MN908947.3	705	727	SARS-CoV-2_2_RIGHT	2	-
5	MN908947.3	644	666	SARS-CoV-2_3_LEFT	1	+
6	MN908947.3	1017	1044	SARS-CoV-2_3_RIGHT	1	-
7	MN908947.3	944	966	SARS-CoV-2_4_LEFT	2	+
8	MN908947.3	1337	1362	SARS-CoV-2_4_RIGHT	2	-
9	MN908947.3	1245	1266	SARS-CoV-2_5_LEFT	1	+
10	MN908947.3	1623	1650	SARS-CoV-2_5_RIGHT	1	-
11	MN908947.3	1540	1562	SARS-CoV-2_6_LEFT	2	+
12	MN908947.3	1925	1948	SARS-CoV-2_6_RIGHT	2	-
13	MN908947.3	1851	1875	SARS-CoV-2_7_LEFT	1	+
14	MN908947.3	2228	2250	SARS-CoV-2_7_RIGHT	1	-
15	MN908947.3	2154	2180	SARS-CoV-2_8_LEFT	2	+
16	MN908947.3	2544	2571	SARS-CoV-2_8_RIGHT	2	-
17	MN908947.3	2483	2508	SARS-CoV-2_9_LEFT	1	+
18	MN908947.3	2861	2885	SARS-CoV-2_9_RIGHT	1	-
19	MN908947.3	2826	2850	SARS-CoV-2_10_LEFT	2	+
20	MN908947.3	3183	3210	SARS-CoV-2_10_RIGHT	2	-
21	MN908947.3	3078	3102	SARS-CoV-2_11_LEFT	1	+
22	MN908947.3	3470	3492	SARS-CoV-2_11_RIGHT	1	-
23	MN908947.3	3390	3412	SARS-CoV-2_12_LEFT	2	+
24	MN908947.3	3769	3794	SARS-CoV-2_12_RIGHT	2	-
25	MN908947.3	3683	3705	SARS-CoV-2_13_LEFT	1	+
26	MN908947.3	4067	4093	SARS-CoV-2_13_RIGHT	1	-
27	MN908947.3	3992	4018	SARS-CoV-2_14_LEFT	2	+

28	MN908947.3	4387	4409	SARS-CoV-2_14_RIGHT	2	-
29	MN908947.3	4312	4339	SARS-CoV-2_15_LEFT	1	+
30	MN908947.3	4685	4710	SARS-CoV-2_15_RIGHT	1	-
31	MN908947.3	4620	4648	SARS-CoV-2_16_LEFT	2	+
32	MN908947.3	4995	5017	SARS-CoV-2_16_RIGHT	2	-
33	MN908947.3	4923	4953	SARS-CoV-2_17_LEFT	1	+
34	MN908947.3	5302	5331	SARS-CoV-2_17_RIGHT	1	-
35	MN908947.3	5230	5259	SARS-CoV-2_18_LEFT	2	+
36	MN908947.3	5620	5643	SARS-CoV-2_18_RIGHT	2	-
37	MN908947.3	5561	5584	SARS-CoV-2_19_LEFT	1	+
38	MN908947.3	5932	5957	SARS-CoV-2_19_RIGHT	1	-
39	MN908947.3	5867	5894	SARS-CoV-2_20_LEFT	2	+
40	MN908947.3	6247	6272	SARS-CoV-2_20_RIGHT	2	-
41	MN908947.3	6184	6210	SARS-CoV-2_21_LEFT	1	+
42	MN908947.3	6553	6582	SARS-CoV-2_21_RIGHT	1	-
43	MN908947.3	6478	6507	SARS-CoV-2_22_LEFT	2	+
44	MN908947.3	6859	6885	SARS-CoV-2_22_RIGHT	2	-
45	MN908947.3	6747	6776	SARS-CoV-2_23_LEFT	1	+
46	MN908947.3	7122	7148	SARS-CoV-2_23_RIGHT	1	-
47	MN908947.3	7057	7084	SARS-CoV-2_24_LEFT	2	+
48	MN908947.3	7440	7467	SARS-CoV-2_24_RIGHT	2	-
49	MN908947.3	7381	7403	SARS-CoV-2_25_LEFT	1	+
50	MN908947.3	7747	7770	SARS-CoV-2_25_RIGHT	1	-
51	MN908947.3	7672	7695	SARS-CoV-2_26_LEFT	2	+
52	MN908947.3	8063	8092	SARS-CoV-2_26_RIGHT	2	-
53	MN908947.3	7997	8019	SARS-CoV-2_27_LEFT	1	+
54	MN908947.3	8370	8395	SARS-CoV-2_27_RIGHT	1	-
55	MN908947.3	8304	8326	SARS-CoV-2_28_LEFT	2	+
56	MN908947.3	8691	8714	SARS-CoV-2_28_RIGHT	2	-
57	MN908947.3	8596	8619	SARS-CoV-	1	+

				2_29_LEFT		
58	MN908947.3	8990	9013	SARS-CoV-2_29_RIGHT	1	-
59	MN908947.3	8919	8944	SARS-CoV-2_30_LEFT	2	+
60	MN908947.3	9306	9329	SARS-CoV-2_30_RIGHT	2	-
61	MN908947.3	9168	9192	SARS-CoV-2_31_LEFT	1	+
62	MN908947.3	9535	9564	SARS-CoV-2_31_RIGHT	1	-
63	MN908947.3	9470	9497	SARS-CoV-2_32_LEFT	2	+
64	MN908947.3	9842	9866	SARS-CoV-2_32_RIGHT	2	-
65	MN908947.3	9782	9805	SARS-CoV-2_33_LEFT	1	+
66	MN908947.3	10150	10176	SARS-CoV-2_33_RIGHT	1	-
67	MN908947.3	10076	10099	SARS-CoV-2_34_LEFT	2	+
68	MN908947.3	10465	10491	SARS-CoV-2_34_RIGHT	2	-
69	MN908947.3	10393	10419	SARS-CoV-2_35_LEFT	1	+
70	MN908947.3	10785	10810	SARS-CoV-2_35_RIGHT	1	-
71	MN908947.3	10713	10742	SARS-CoV-2_36_LEFT	2	+
72	MN908947.3	11092	11116	SARS-CoV-2_36_RIGHT	2	-
73	MN908947.3	11000	11023	SARS-CoV-2_37_LEFT	1	+
74	MN908947.3	11388	11414	SARS-CoV-2_37_RIGHT	1	-
75	MN908947.3	11305	11330	SARS-CoV-2_38_LEFT	2	+
76	MN908947.3	11689	11720	SARS-CoV-2_38_RIGHT	2	-
77	MN908947.3	11624	11651	SARS-CoV-2_39_LEFT	1	+
78	MN908947.3	12011	12033	SARS-CoV-2_39_RIGHT	1	-
79	MN908947.3	11937	11963	SARS-CoV-2_40_LEFT	2	+
80	MN908947.3	12317	12339	SARS-CoV-2_40_RIGHT	2	-
81	MN908947.3	12234	12255	SARS-CoV-2_41_LEFT	1	+
82	MN908947.3	12618	12643	SARS-CoV-2_41_RIGHT	1	-
83	MN908947.3	12519	12546	SARS-CoV-2_42_LEFT	2	+
84	MN908947.3	12895	12920	SARS-CoV-2_42_RIGHT	2	-
85	MN908947.3	12831	12856	SARS-CoV-2_43_LEFT	1	+
86	MN908947.3	13218	13240	SARS-CoV-2_43_RIGHT	1	-

87	MN908947.3	13124	13148	SARS-CoV-2_44_LEFT	2	+
88	MN908947.3	13506	13528	SARS-CoV-2_44_RIGHT	2	-
89	MN908947.3	13463	13485	SARS-CoV-2_45_LEFT	1	+
90	MN908947.3	13833	13859	SARS-CoV-2_45_RIGHT	1	-
91	MN908947.3	13752	13775	SARS-CoV-2_46_LEFT	2	+
92	MN908947.3	14120	14144	SARS-CoV-2_46_RIGHT	2	-
93	MN908947.3	14045	14075	SARS-CoV-2_47_LEFT	1	+
94	MN908947.3	14428	14457	SARS-CoV-2_47_RIGHT	1	-
95	MN908947.3	14338	14362	SARS-CoV-2_48_LEFT	2	+
96	MN908947.3	14717	14743	SARS-CoV-2_48_RIGHT	2	-
97	MN908947.3	14647	14674	SARS-CoV-2_49_LEFT	1	+
98	MN908947.3	15023	15050	SARS-CoV-2_49_RIGHT	1	-
99	MN908947.3	14953	14983	SARS-CoV-2_50_LEFT	2	+
100	MN908947.3	15336	15358	SARS-CoV-2_50_RIGHT	2	-
101	MN908947.3	15214	15237	SARS-CoV-2_51_LEFT	1	+
102	MN908947.3	15596	15619	SARS-CoV-2_51_RIGHT	1	-
103	MN908947.3	15535	15557	SARS-CoV-2_52_LEFT	2	+
104	MN908947.3	15917	15941	SARS-CoV-2_52_RIGHT	2	-
105	MN908947.3	15855	15881	SARS-CoV-2_53_LEFT	1	+
106	MN908947.3	16239	16260	SARS-CoV-2_53_RIGHT	1	-
107	MN908947.3	16112	16137	SARS-CoV-2_54_LEFT	2	+
108	MN908947.3	16483	16508	SARS-CoV-2_54_RIGHT	2	-
109	MN908947.3	16386	16408	SARS-CoV-2_55_LEFT	1	+
110	MN908947.3	16767	16796	SARS-CoV-2_55_RIGHT	1	-
111	MN908947.3	16692	16714	SARS-CoV-2_56_LEFT	2	+
112	MN908947.3	17082	17105	SARS-CoV-2_56_RIGHT	2	-
113	MN908947.3	16986	17013	SARS-CoV-2_57_LEFT	1	+
114	MN908947.3	17381	17405	SARS-CoV-2_57_RIGHT	1	-
115	MN908947.3	17323	17345	SARS-CoV-2_58_LEFT	2	+
116	MN908947.3	17688	17711	SARS-CoV-	2	-

				2_58_RIGHT		
117	MN908947.3	17615	17642	SARS-CoV-2_59_LEFT	1	+
118	MN908947.3	17997	18022	SARS-CoV-2_59_RIGHT	1	-
119	MN908947.3	17911	17939	SARS-CoV-2_60_LEFT	2	+
120	MN908947.3	18307	18328	SARS-CoV-2_60_RIGHT	2	-
121	MN908947.3	18244	18267	SARS-CoV-2_61_LEFT	1	+
122	MN908947.3	18624	18652	SARS-CoV-2_61_RIGHT	1	-
123	MN908947.3	18550	18578	SARS-CoV-2_62_LEFT	2	+
124	MN908947.3	18936	18961	SARS-CoV-2_62_RIGHT	2	-
125	MN908947.3	18869	18891	SARS-CoV-2_63_LEFT	1	+
126	MN908947.3	19252	19277	SARS-CoV-2_63_RIGHT	1	-
127	MN908947.3	19183	19208	SARS-CoV-2_64_LEFT	2	+
128	MN908947.3	19558	19586	SARS-CoV-2_64_RIGHT	2	-
129	MN908947.3	19485	19513	SARS-CoV-2_65_LEFT	1	+
130	MN908947.3	19877	19901	SARS-CoV-2_65_RIGHT	1	-
131	MN908947.3	19810	19836	SARS-CoV-2_66_LEFT	2	+
132	MN908947.3	20186	20216	SARS-CoV-2_66_RIGHT	2	-
133	MN908947.3	20090	20117	SARS-CoV-2_67_LEFT	1	+
134	MN908947.3	20472	20497	SARS-CoV-2_67_RIGHT	1	-
135	MN908947.3	20377	20405	SARS-CoV-2_68_LEFT	2	+
136	MN908947.3	20766	20792	SARS-CoV-2_68_RIGHT	2	-
137	MN908947.3	20677	20699	SARS-CoV-2_69_LEFT	1	+
138	MN908947.3	21050	21080	SARS-CoV-2_69_RIGHT	1	-
139	MN908947.3	20988	21013	SARS-CoV-2_70_LEFT	2	+
140	MN908947.3	21358	21387	SARS-CoV-2_70_RIGHT	2	-
141	MN908947.3	21294	21316	SARS-CoV-2_71_LEFT	1	+
142	MN908947.3	21675	21700	SARS-CoV-2_71_RIGHT	1	-
143	MN908947.3	21532	21561	SARS-CoV-2_72_LEFT	2	+
144	MN908947.3	21904	21933	SARS-CoV-2_72_RIGHT	2	-
145	MN908947.3	21865	21889	SARS-CoV-2_73_LEFT	1	+

146	MN908947.3	22247	22274	SARS-CoV-2_73_RIGHT	1	-
147	MN908947.3	22091	22113	SARS-CoV-2_74_LEFT	2	+
148	MN908947.3	22474	22503	SARS-CoV-2_74_RIGHT	2	-
149	MN908947.3	22402	22428	SARS-CoV-2_75_LEFT	1	+
150	MN908947.3	22785	22805	SARS-CoV-2_75_RIGHT	1	-
151	MN908947.3	22648	22677	SARS-CoV-2_76_LEFT	2	+
152	MN908947.3	23028	23057	SARS-CoV-2_76_RIGHT	2	-
153	MN908947.3	22944	22974	SARS-CoV-2_77_LEFT	1	+
154	MN908947.3	23327	23351	SARS-CoV-2_77_RIGHT	1	-
155	MN908947.3	23219	23246	SARS-CoV-2_78_LEFT	2	+
156	MN908947.3	23611	23635	SARS-CoV-2_78_RIGHT	2	-
157	MN908947.3	23553	23575	SARS-CoV-2_79_LEFT	1	+
158	MN908947.3	23927	23955	SARS-CoV-2_79_RIGHT	1	-
159	MN908947.3	23853	23876	SARS-CoV-2_80_LEFT	2	+
160	MN908947.3	24233	24258	SARS-CoV-2_80_RIGHT	2	-
161	MN908947.3	24171	24194	SARS-CoV-2_81_LEFT	1	+
162	MN908947.3	24545	24567	SARS-CoV-2_81_RIGHT	1	-
163	MN908947.3	24426	24448	SARS-CoV-2_82_LEFT	2	+
164	MN908947.3	24814	24836	SARS-CoV-2_82_RIGHT	2	-
165	MN908947.3	24750	24772	SARS-CoV-2_83_LEFT	1	+
166	MN908947.3	25122	25150	SARS-CoV-2_83_RIGHT	1	-
167	MN908947.3	25051	25076	SARS-CoV-2_84_LEFT	2	+
168	MN908947.3	25438	25461	SARS-CoV-2_84_RIGHT	2	-
169	MN908947.3	25331	25353	SARS-CoV-2_85_LEFT	1	+
170	MN908947.3	25711	25740	SARS-CoV-2_85_RIGHT	1	-
171	MN908947.3	25645	25672	SARS-CoV-2_86_LEFT	2	+
172	MN908947.3	26026	26050	SARS-CoV-2_86_RIGHT	2	-
173	MN908947.3	25951	25979	SARS-CoV-2_87_LEFT	1	+
174	MN908947.3	26338	26360	SARS-CoV-2_87_RIGHT	1	-
175	MN908947.3	26255	26277	SARS-CoV-	2	+

				2_88_LEFT		
176	MN908947.3	26635	26661	SARS-CoV-2_88_RIGHT	2	-
177	MN908947.3	26564	26587	SARS-CoV-2_89_LEFT	1	+
178	MN908947.3	26956	26979	SARS-CoV-2_89_RIGHT	1	-
179	MN908947.3	26873	26895	SARS-CoV-2_90_LEFT	2	+
180	MN908947.3	27256	27283	SARS-CoV-2_90_RIGHT	2	-
181	MN908947.3	27152	27177	SARS-CoV-2_91_LEFT	1	+
182	MN908947.3	27534	27560	SARS-CoV-2_91_RIGHT	1	-
183	MN908947.3	27447	27473	SARS-CoV-2_92_LEFT	2	+
184	MN908947.3	27826	27855	SARS-CoV-2_92_RIGHT	2	-
185	MN908947.3	27700	27726	SARS-CoV-2_93_LEFT	1	+
186	MN908947.3	28082	28104	SARS-CoV-2_93_RIGHT	1	-
187	MN908947.3	27996	28021	SARS-CoV-2_94_LEFT	2	+
188	MN908947.3	28394	28416	SARS-CoV-2_94_RIGHT	2	-
189	MN908947.3	28190	28214	SARS-CoV-2_95_LEFT	1	+
190	MN908947.3	28572	28598	SARS-CoV-2_95_RIGHT	1	-
191	MN908947.3	28512	28536	SARS-CoV-2_96_LEFT	2	+
192	MN908947.3	28893	28914	SARS-CoV-2_96_RIGHT	2	-
193	MN908947.3	28827	28849	SARS-CoV-2_97_LEFT	1	+
194	MN908947.3	29206	29227	SARS-CoV-2_97_RIGHT	1	-
195	MN908947.3	29136	29161	SARS-CoV-2_98_LEFT	2	+
196	MN908947.3	29512	29534	SARS-CoV-2_98_RIGHT	2	-
197	MN908947.3	29452	29475	SARS-CoV-2_99_LEFT	1	+
198	MN908947.3	29827	29854	SARS-CoV-2_99_RIGHT	1	-