



CENTRO UNIVERSITÁRIO DE BRASÍLIA - CEUB
PROGRAMA DE INICIAÇÃO CIENTÍFICA

KELTON DE OLIVEIRA CONCEIÇÃO
ROBERTA GOMES GONTIJO

**ANÁLISE *IN SILICO* DO GENOMA DE ESTIRPES DE Sars-CoV-2 PARA
CORRELAÇÃO ENTRE VARIABILIDADE GENÉTICA E DADOS EPIDEMIOLÓGICOS**

BRASÍLIA

2021

KELTON DE OLIVEIRA CONCEIÇÃO

ROBERTA GOMES GONTIJO

**ANÁLISE *IN SILICO* DO GENOMA DE ESTIRPES DE Sars-CoV-2 PARA
CORRELAÇÃO ENTRE VARIABILIDADE GENÉTICA E DADOS EPIDEMIOLÓGICOS**

Relatório final de pesquisa de Iniciação Científica apresentado à Assessoria de Pós-Graduação e Pesquisa.

Orientação: Paulo Roberto Martins Queiroz

BRASÍLIA

2021

DEDICATÓRIA

Este trabalho é dedicado a todas as pessoas que perderam suas vidas em decorrência da pandemia do COVID-19. Deixamos aqui nossas sinceras condolências.

AGRADECIMENTOS

Primeiramente gostaríamos de agradecer a Deus, pois sem ele nada disso seria possível. Também deixamos nossos agradecimentos a nossos pais Elisangela Gomes Chaves; Ivanildo da Conceição; Renato Valadares Gontijo; Selma Pedroso de Oliveira; que tanto fazem por nós; aos nossos irmãos Giovanna Gomes Chaves; Kellidy de Oliveira Conceição; Manuela Gomes Chaves; Pedro Camilo Valadares Gontijo que nos inspiram à excelência; a nossa amiga Camila Beatriz de Lima Ferreira por sempre nos incentivar; ao meu namorado Dalton Pedro Morelo Ferraz por sempre estar ao meu lado fomentando pelo meu sucesso e compreendendo minha ausência enquanto me dedicava à realização deste trabalho; e ao nosso orientador, Paulo Roberto Martins Queiroz, que teve tanta paciência e resiliência ao nos auxiliar na produção desse trabalho.

“Que consigamos lutar contra o obscurantismo que anda tão forte e vejamos (como diria Carl Sagan) a ciência como uma vela no escuro.”

Stephane Werner

RESUMO

O novo coronavírus ou SARS-CoV-2, que é um vírus pertencente à família Coronaviridae, que é um grupo de vírus com ácido ribonucleico de cadeia simples (RNA), envelopado de sentido simples. Esse nome tem origem do latim corona, que tem como significado "coroa" ou também "auréola" por apresentar espinhos que cobrem toda a sua membrana superficial. Este vírus tem sua transmissão acelerada que se sucede através de secreções ou por meio de superfícies contaminadas. Os danos que esse vírus pode causar em paralelo com a presteza que se dissemina é motivo de alerta e preocupação pelas autoridades sanitárias. Dessa forma, esse trabalho teve como objetivo analisar a variabilidade genética das várias estirpes de SARS-CoV-2 correlacionando com dados epidemiológicos para achar uma maneira de relacionar o perfil molecular dos tipos de coronavírus com possíveis padrões epidemiológicos. Assim, inicialmente, encontrou-se um genoma médio de 29.789 bp, com distribuições das bases nitrogenadas em 29,89% de A; 32,16% de T; 18,36% de C; e 19,62% de G, fato esse divergente ao analisar cada nacionalidade. Outro ponto importante é a baixa proporção de CG de 38%, em relação à Sars-CoV e Mers-CoV. Ao analisar as mutações no material genético, a taxa de transformação de citosina e guanina em timina é maior que as demais substituições, além de existir processos que transformam outra base nitrogenada em timina, assim, futuramente, aumentando a proporção de timina no genoma. Também cabe destacar que a maior parte das mutações ocorrem em regiões ORF1ab e na proteína estrutural spike (S), o qual tais mutações podem influenciar, respectivamente, na transmissão e infecção; e no aumento da carga viral. Além do exposto, utilizando enzimas de restrição nesse genoma foi possível identificar o aparecimento de 5 enzimas que não cortam a grande parte dos genomas, sendo coincidentemente essas enzimas encontradas na Venezuela, Índia, Itália. Acerca do exposto, foi possível identificar divergências no Sars-CoV-2 de acordo com a localidade, a prevalência das mutações, assim como, o local mais acometido do genoma e correlacionando essas mutações a possíveis dados epidemiológicos.

Palavras-chave: Bioinformática; Spike; Coronaviridae.

LISTAS DE FIGURAS, TABELAS, QUADROS, GRÁFICOS, SÍMBOLOS E ABREVIações

Tabela e Figura	Página
Quadro 1 – Principais gêneros de coronavírus que causam doenças em seres humano e algumas características principais.	11
Gráfico 1 - Relação de casos de COVID (gráfico A) e óbitos (gráfico B) no mundo.	16
Gráfico 2 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) do Reino Unido após o surgimento da variante Alpha.	19
Gráfico 3 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) da África do Sul após o surgimento da variante Beta.	20
Gráfico 4 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) do Brasil após o surgimento da variante Gama.	21
Gráfico 5 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) da Índia após o surgimento da variante Delta.	23
Quadro 2 – Média e moda do tamanho dos genomas de Sars-Cov-2 sequenciados em cada continente.	27
Figura 6 - Mapa comparativo entre as sequências consenso dos países da Oceania e China. Os traços verticais em cinza indicam as respectivas mutações.	28
Figura 7 - Mapa comparativo entre as sequências consenso dos países da África e China. Os traços verticais em cinza indicam as respectivas mutações.	28
Figura 8 - Mapa comparativo entre as sequências consenso dos países da Ásia e China. Os traços verticais em cinza indicam as respectivas mutações.	30
Figura 9 – Análise filogenética das estirpes de Sars-Cov-2 provenientes de 71 países que foram usadas nas simulações in silico com enzimas de restrição. Coeficiente de correlação de 0,97.	34

SUMÁRIO

1	INTRODUÇÃO	09
2	FUNDAMENTAÇÃO TEÓRICA	10
2.1	HISTÓRIA DO CORONAVÍRUS	10
2.2	NOVO CORONAVÍRUS (SARS-COV-2)	12
2.3	ASPECTOS SOCIAIS	14
2.4	EPIDEMIOLOGIA	15
2.5	VARIANTES	17
2.5.1	VARIANTE ALPHA – B.1.1.7 – REINO UNIDO	18
2.5.2	VARIANTE BETA – B.1.3.5.1 - ÁFRICA DO SUL	19
2.5.3	VARIANTE GAMA – P.1 – BRASIL	20
2.5.4	VARIANTE DELTA – B. 1.6.1.7.2 – ÍNDIA	22
3	MÉTODO	24
4	RESULTADOS E DISCUSSÃO	26
5	CONSIDERAÇÕES FINAIS (OU CONCLUSÕES)	33
6	REFERÊNCIAS	35

1 INTRODUÇÃO

O vírus SARS-CoV-2, identificado pela primeira vez em Wuhan na China, vem causando a pandemia do COVID-19, declarada assim pela Organização Mundial da Saúde (OMS) em março de 2020, sendo ela associada à síndrome respiratória aguda grave, por isso da sua denominação (VALENCIA, 2020).

Esse novo padecimento é causado por um novo tipo de Coronavírus, ou seja, faz parte da família Coronaviridae, essa sendo formada por quatro gêneros: Alphacoronavírus; Betacoronavírus; Gammacoronavírus e Deltacoronavírus (MALIK, 2020). O SARS-CoV-2 é um vírus de formato esférico, diâmetros de cerca de 125 nm², envelopado e com seu nucleocapsídeo helicoidalmente simétrico. O genoma consiste de uma molécula de ácido ribonucleico (RNA), não segmentado, sentido positivo (RNA+), ou seja, um RNA de extremidade 5' junto com uma cauda de extremidade 3' poli(A) que permite se comportar como um RNAm (MALIK, 2020; CHEN, 2020).

A virulência e patogenicidade desse vírus, pode estar concatenada a sua atração pelos receptores da enzima conversora de angiotensina 2 (ACE2), que possui grande expressão principalmente nos pulmões e coração, explicando de certa forma o mecanismo de infecção primária do trato respiratório (UZUNIAN, 2020).

A transmissão, podendo ser direta ou indireta, ocorre pelo contato da membrana mucosa com o material contaminado, e pode ocorrer através dos sintomáticos e dos assintomáticos, esses que são cerca de 2/3 do total dos infectados e podem produzir uma carga viral suficiente para que ocorra esse contágio. Ocorrendo após um período de incubação que ocorre no intervalo de 1 a 14 dias, embora existam relatos de casos em que esse período seja de até 24 dias. Essa disseminação ocorre, principalmente, através de gotículas respiratórias, mas também pode ocorrer através do contato com superfícies contaminadas, fezes, sangue ou fômites (VALENCIA, 2020; RIO, 2020) Outro fator importante é que as condições climáticas e do tempo podem favorecer a sobrevivência e transmissão do vírus, como ambientes de baixa umidade e de baixa temperatura (PAL et al, 2020).

Assim, torna-se de extrema relevância sua análise molecular, pois a partir dela pode-se diferenciar e descrever todas as estirpes existentes; suas particularidades; padrões existentes; padrões de mutação; sobrevivência; alterações nas suas proteínas estruturais e não estruturais; sua fisiopatologia; etiopatogenia e sintomatologia.

Acerca do exposto, devido às suas peculiaridades, existe a necessidade de buscar maneiras de combater e entender o funcionamento da dinâmica do SARS-COV-2 nas populações. Buscou-se analisar e comparar o material genético das principais estirpes da família coronavírus e das diferentes estirpes do novo coronavírus espalhadas pelo mundo, para entender seu funcionamento, evolução e buscar fatores que ajudem no seu controle. A realização da pesquisa visa estabelecer e entender se a variabilidade das estirpes do novo coronavírus estão dentro de padrões epidemiológicos predizíveis. Essa estratégia permitiria estabelecer, então, procedimentos de correlação dos perfis genéticos das estirpes do coronavírus com quadros clínicos específicos. Isso ajudaria no estabelecimento de estratégias de manejo da população em surtos futuros.

OBJETIVOS

Objetivo Geral

Analisar a variabilidade genética das várias estirpes de Sars-CoV-2 e correlacionar com dados epidemiológicos utilizando a análise *in silico* e levantamento de dados epidemiológicos.

Objetivo Específico

Obter a sequência genômica das várias estirpes de Sars-CoV-2 em bancos de sequências de nucleotídeos;

Analisar a variabilidade genética das várias estirpes de Sars-CoV-2 por ferramentas de bioinformática;

Realizar o levantamento epidemiológico das várias estirpes por pesquisa bibliográfica e em bancos de dados relacionados ao Sars-CoV-2;

Relacionar o perfil molecular dos tipos de coronavírus com possíveis distribuições de padrões epidemiológicos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1. História do Coronavírus

Os coronavírus são membros da ordem Nidovirales e família Coronaviridae, que são caracterizados, respectivamente, por apresentar um genoma muito grande, cerca de 30 kb - fato incomum para vírus de RNA -, e pela “coroa” ou “halo”, identificado na microscopia eletrônica, o qual seriam os peplômeros - uma proteína estrutural chamada spike (S) - cobrindo sua superfície (VALENCIA, 2020; ARTIKA; DEWANTARI; WIYATNO, 2020). Ademais, são vírus de ácido ribonucléico (RNA); não segmentado; de fita simples; de sentido positivo (ssRNA +), ou seja, um RNA com extremidade 5' junto com uma cauda de extremidade 3' poli(A), que permite se comportar como um RNAm; e envelopado (PAL et al, 2020; DECARO; LORUSSO, 2020).

Atualmente, a família Coronaviridae, da subfamília Orthocoronavirinae são divididos em quatro gêneros: os *Alphacoronavirus* e *Betacoronavirus* - responsáveis exclusivamente pelas infecções em mamíferos - e os *Gammacoronavirus* e *Deltacoronavirus* – causadores de doença principalmente em aves (DECARO; LORUSSO, 2020; ARTIKA; DEWANTARI; WIYATNO, 2020; DHAMA et al, 2020; LIU; LIANG; FUNG, 2021). Desse modo, os vírus dos gêneros *Alphacoronavirus* e *Betacoronavirus* sofreram mutações e passaram a infectar humanos dando origem a sete coronavírus humanos (Quadro 1).

Quadro 1 – Principais gêneros de coronavírus que causam doenças em seres humano e algumas características principais.

Nome	Gênero	Possível Ancestral	Doença Associada	Ano que foi identificado
HCoV-229E	Alphacoronavírus	Originado em morcegos e transmitido a humanos por camelídeos	Resfriado Comum	1966
HCoV-OC43	Betacoronavírus	Originado em roedores e transmitido a humanos por bovídeos	Resfriado Comum	1967
SARS-CoV	Betacoronavírus	Originado em morcegos e transmitido a	Síndrome respiratória	2002

		humanos por carnívoros selvagens	aguda grave	
HCoV-NL63	Alphacoronavírus	Originado e transmitido a humanos por morcegos	Resfriado Comum	2004
HCoV-HKU1	Betacoronavírus	Originado e transmitido a humanos por morcegos	Resfriado Comum	2005
MERS-CoV	Betacoronavírus	Originado em morcegos e transmitido a humanos por camelídeos	Síndrome respiratória aguda grave	2012
SARS-CoV-2	Betacoronavírus	Originado em morcegos e transmitido a humanos por pangolins	Síndrome respiratória aguda grave	2019

Fontes: (ARTIKA; DEWANTARI; WIYATNO, 2020; DECARO; LORUSSO, 2020; VALENCIA, 2020; RIO; MALANI, 2020; LIU; LIANG; FUNG, 2021).

Nesse contexto, a recente história dos coronavírus humano iniciou-se em 1966 com a descoberta do primeiro vírus, o HCoV-229E, causador de doença respiratória leve. Após essa descoberta, foram identificados outros seis vírus, incluindo o recém-descoberto Sars-CoV-2, capazes de infectar humanos (DECARO; LORUSSO, 2020; DHAMA et al, 2020). Outro ponto crucial dessa evolução é a piora das manifestações clínicas que passou de um resfriado comum para uma Síndrome Respiratória Aguda Grave, e levou a duas epidemias e uma pandemia em menos de 20 anos, sendo a Sars-CoV, em 2002, iniciada na China; o MERS-CoV, em 2012, iniciada na Arábia Saudita; e o Sars-CoV-2, em 2019, iniciada na China, e tornou-se a pandemia da COVID-19 (LIU; LIANG; FUNG, 2021; DECARO; LORUSSO, 2020).

2.2. Novo Coronavírus (Sars-CoV-2)

Esse novo vírus passou a ser conhecido em 31 de dezembro de 2019 quando a Organização Mundial da Saúde (OMS) é alertada sobre vários casos de pneumonia em

Wuhan (Hubei, China), que, inicialmente, tinha sua origem desconhecida (SILVA, 2020; OPAS/OMS, 2020).

Assim, após o sequenciamento e análise do material genético do patógeno foi identificado o novo coronavírus, o SARS-CoV-2 (síndrome respiratória aguda coronavírus 2), pertence à família coronaviridae, da ordem Nidovirales e do gênero Betacoronavírus (PAL et al, 2020; SILVA, 2020; OMS, 2020). Além disso, com base na sua proximidade genética, o SARS-CoV-2 provavelmente se originou do CoV derivado do morcego e foi transmitido aos humanos por meio de um hospedeiro intermediário, provavelmente o pangolim malaio (KADAM et al, 2021).

Um mês depois, em 30 de janeiro de 2020, a OMS declara que o surto constitui uma Emergência de Saúde Pública de Importância Internacional (ESPII) – o maior nível de alerta da organização -, mas em 11 de março de 2020, menos de três meses após o alerta da China, a Organização Mundial da Saúde (OMS) declara pandemia do COVID-19 (VALENCIA, 2020; SILVA, 2020; OMS, 2020). Deste modo, deixando claro que a demora de para iniciar as medidas de controle pelos órgãos de saúde influenciaram a disseminação do vírus e o cenário perfeito para a criação de novas estirpes.

O Sars-CoV-2 contém um genoma (>29 kb) de RNA de fita simples, sentido positivo empacotado em um envelope esférico com projeções em sua superfície (KADAM et al, 2021). Esse genoma contém diversos quadros de leitura aberto (ORF's), ou seja, uma parte do genoma que tem capacidade de ser traduzida e região não traduzida (UTR), sendo assim, organizados em: calda 5'cap-UTR- ORF1ab – ORF's - 3'UTR-poli(A) (PAL et al, 2020; CHEN; LIU; GUO, 2020). Além disso, no início de cada gene existe uma sequência reguladora da transcrição (TRSs) que pode ser necessária para a expressão de cada gene estrutural ou de acentuação (PAL et al, 2020; KADAM et al, 2021).

Dois terços do genoma correspondem ao ORF1ab produzindo duas poliproteínas: PP1a e PP1ab que, a partir do seu processamento, surgem dezesseis proteínas não estruturais putativas (NSPs), sendo responsáveis pelas diferentes funções vitais, incluindo a formação do complexo replicase-transcriptase (RTC), responsável pela replicação e transcrição do RNA viral (CHEN; LIU; GUO, 2020; VALENCIA, 2020; DHAMA et al, 2020; CHAN et al., 2020). Outrossim, vale salientar a existência de NSPs que modulam a imunidade do hospedeiro, o qual interfere na produção de proteínas e nas respostas imunes inatas (KADAM et al, 2021).

O outro terço do genoma, representado pelas ORFs, encontram-se as proteínas acessórias e as estruturais, sendo que os genes que codificam as proteínas acessórias estão presentes entre os genes estruturais (CHAN et al., 2020; KADAM et al, 2021). Existem pelo menos seis ORFs que codificam proteínas acessórias encontradas no SARS-CoV-2 incluindo 3a, 6, 7a, 7b, 8 (8b) e 9b, sendo que, o 8a e hemaglutinina esterase (HE) estão ausentes, e o 8b apresenta mais nucleotídeos se comparado aos demais CoVs. Suas funções não são totalmente conhecidas, mas estudos anteriores indicam que apresentam papel contra a resposta imune do hospedeiro, regulação da atividade de canal iônico e podem estar relacionados com a infecção e a patogenicidade do vírus (KADAM et al, 2021).

Ademais, também podem-se encontrar as proteínas estruturais spike (S); pequena membrana (E); membrana (M) e nucleocapsídeo (N), que são essenciais para a montagem do vírion e a infecção do CoVs (PAL et al, 2020; CHEN; LIU; GUO, 2020; KADAM et al, 2021). Assim, as proteínas spike (S) constituem os picos na superfície viral e são responsáveis pela ligação, através do RBD (domínio receptor-obrigatório), aos receptores do hospedeiro, no caso a enzima conversora de angiotensina-2 (ACE2) e a enzima transmembrana serina protease 2 (TMPRSS2), proteína encontrada no endotélio respiratório e do trato digestivo (CHEN; LIU; GUO, 2020; RIO; MALANI, 2020; YUKI; FUJIOGI; KOUTSOGIANNAKI, 2020; KADAM et al, 2021; DHAMA et al, 2020).

A proteína pequena membrana (E) desempenha papel na montagem e liberação do vírus e está envolvida na patogênese viral. A proteína membrana (M) possui três domínios transmembranares e é a mais abundante na partícula viral, sendo responsável, devido à interação com as proteínas S e E, por definir a forma do envelope viral. Por fim, a proteína nucleocapsídeo (N) contém dois domínios, os quais podem formar um complexo com o genoma viral facilitando a interação da proteína M durante sua montagem e transcrição. Ademais, essa proteína N é antagonista do interferon (IFN) e repressor viral da interferência do RNA, que parece ser benéfico para a replicação viral (CHEN; LIU; GUO, 2020; KADAM et al, 2021; DHAMA et al, 2020).

2.3. Aspectos Sociais

O material genético do SARS-CoV-2 pode ser detectado na parte superior e inferior do trato respiratório, tanto em pacientes sintomáticos quanto em assintomáticos, mostrando o importante papel do epitélio nasal como depósito viral e a importância da

mucosa respiratória na dispersão do vírus (SALIAN et al, 2021). A transmissão do SARS-CoV-2 ocorre principalmente por meio de aerossóis, as micropartículas de secreção respiratória e/ou gotículas, mas a transmissão fecal-oral e o contato direto com superfícies também foram relatados durante a pandemia (HARRISON et al, 2020). Estudos demonstraram que esses aerossóis podem permanecer de forma viável no ar por pelo menos 3 h tendo como tempo de meia-vida de 1 h, sendo ainda suficientemente contagioso para o ser humano (SALIAN et al, 2021).

O SARS-CoV-2 tem alta infectabilidade e não há predileção por faixas etárias, sendo todas as idades passíveis de serem infectadas, mas, verificando-se os dados epidemiológicos pode-se denotar que o vírus tem menor propensão pelas crianças e apesar da gravidade da COVID-19, quando observada nesta mesma população, essas são em parte assintomáticos ou de gravidade exígua. Isto posto, infere-se também que as crianças não fazem parte dos principais transmissores desse vírus (BAGGIO et al., 2020).

Alguns sinais e sintomas mais genéricos podem ajudar no direcionamento da suspeita de infecção causada pelo SARS-CoV-2, segundo o Ministério da Saúde, a síndrome gripal (SG) é a manifestação mais usual, esta podendo ser estabelecida como um quadro respiratório agudo descrito por, mal-estar, febre, fadiga, tosse, dispneia leve, anorexia, dor de garganta, dor no corpo, dor de cabeça ou congestão nasal, podendo ocorrer em alguns casos também diarreia, náusea e vômito. Os sinais/sintomas podem variar de pessoa para pessoa, manifestando-se algumas vezes equitativamente a uma pneumonia, pneumonia grave ou síndrome respiratória aguda grave (SRAG) (ISER et al, 2020).

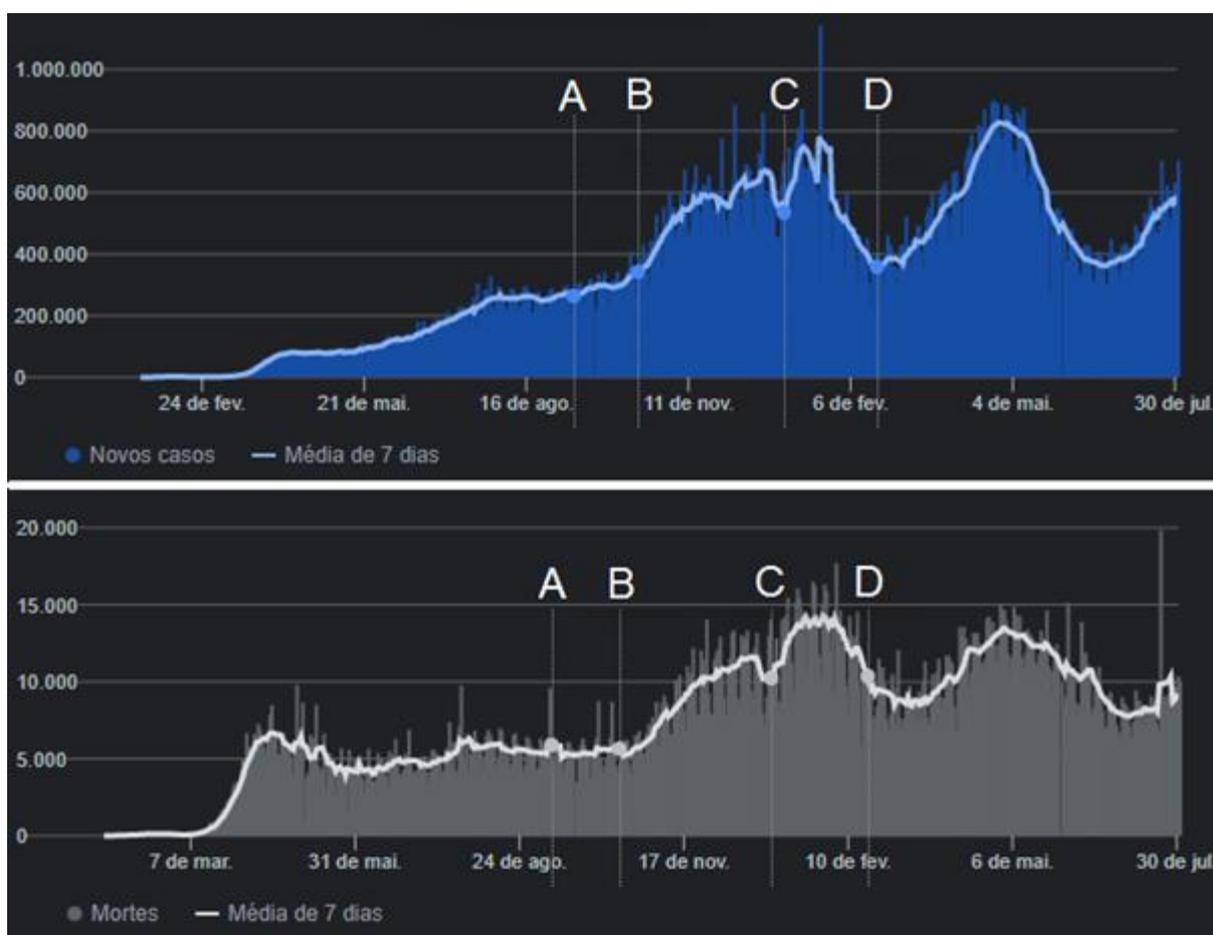
2.4. Epidemiologia

O marco central para a disseminação do Sars-CoV-2 foi coincidir o seu surgimento com o Ano Novo Lunar Chinês, o qual milhões de pessoas acabam viajando para outras cidades. Um pouco depois desse período houve o surgimento de diversos casos em todo mundo, assim o epicentro da pandemia migrou da Ásia, para a Europa e, depois, as Américas, voltando para a Europa com a variante Alfa e voltando para a Ásia com a variante Delta (MALLAH et al, 2021).

Devido à facilidade de transmitir o vírus, diversos países observaram, em pouco tempo, o número de casos e óbitos aumentarem de forma exponencial (Gráfico 1), os quais impactaram o sistema de saúde e diversas estruturas da sociedade. Em respostas a esses

danos, diversos países tomaram decisões históricas e adotaram intervenções não farmacêuticas (INP), como regras de restrições sociais, suspensão de aulas; fechamento de estabelecimentos públicos/privados e quaisquer locais que gerassem aglomerações, além do uso de máscaras como forma de parar a disseminação da doença. Destarte, foi observado um declínio dessas taxas (casos e óbitos) em respostas a essas medidas (SALVATORE et al, 2021; MALLAH et al, 2021; TUFAN; KAYAASLAN, 2020).

Gráfico 1 - Relação de casos de COVID (gráfico A) e óbitos (gráfico B) no mundo.



Legenda: A = Aparecimento da variante Alpha; B = Aparecimento da variante Beta; C = Aparecimento da variante Gama; D = Aparecimento da variante Delta.

Fonte: JHU CSSE COVID-19 Adaptado.

Outro ponto que a pandemia do COVID-19 influenciou foi o colapso do sistema de saúde, assim, devido à baixa adesão das medidas não farmacêuticas, ocorreu aumento significativo nos casos, principalmente naqueles que necessitavam de atendimento especializado. Destarte, diante ao grande fluxo de pacientes, a falta de leitos de enfermaria,

de leitos de UTI (Unidade de Terapia Intensiva), medicamentos, oxigênio e recursos humanos levaram ao colapso do sistema e, conseqüentemente, interferindo na mortalidade, sendo que muitos pacientes morreram sem conseguir um atendimento (BARRETO et al, 2021).

Após a implementação das medidas não farmacológicas diversos países experimentaram a estabilização da curva de casos e óbitos e, posteriormente, sua redução (Gráfico 2; 3; 4 e 5), assim relaxaram as medidas de controles sanitários, liberaram o retorno de atividades não essenciais, como bares e restaurantes, retorno de eventos com público e, em alguns casos, a liberação do uso obrigatório de máscaras. Além do início da vacinação que auxiliou todo esse processo (SALVATORE et al, 2021).

Entretanto, com o afrouxamento das medidas e o surgimento de novas variantes, os números de casos voltaram a crescer consideravelmente (Gráficos 2; 3; 4 e 5) produzindo a segunda onda de infecção com a variante Alpha e, posteriormente a terceira onda com a Delta (SALVATORE et al, 2021; FRAMPTON et al, 2021; MORA et al, 2021). Ao observar a figura 01 pode-se notar o impacto dessas variantes em escala mundial devido o aumento da capacidade de transmissão do vírus e a mudança do perfil dos pacientes infectados o qual, antes eram mais idosos e pessoas com alguma doença prévia e, atualmente, passou a afetar a população em geral, desde jovens a pessoas hígdas (ORTIZ-PRADO et al, 2020; MALLAH et al, 2021).

2.5. Variantes

Do ponto de vista genético, a família coronavírus apresenta uma excepcional plasticidade genética que facilita a produção de estirpes. Um dos pontos que favorecem o processo é o seu genoma de RNA de fita simples (ssRNA), o qual manifesta, em geral, uma taxa de erro de 10^{-4} durante a replicação do RNA, pois a RNA polimerase, ao contrário da DNA polimerase, não tem as capacidades de revisão nem os mecanismos de reparo de incompatibilidade pós-replicação (DECARO; LORUSSO, 2020; ARTIKA; DEWANTARI; WIYATNO, 2020). Ademais, outro processo que favorece essa variabilidade é a capacidade de recombinação dos CoV, assim, devido maquinaria de replicação, o vírus pode produzir um RNA híbrido contendo sequências de outros coronavírus ou de diferentes vírus do organismo, deste modo podem aumentar sua patogenicidade na mesma espécie ou adaptar-se a espécies diferentes (DECARO; LORUSSO, 2020).

Entre as mutações observadas no Sars-CoV-2 a transição de pirimidina, de C para T (a literatura usa o nucleotídeo T condigno os bancos de dados representarem o genoma em DNA) domina completamente outras mutações. Também é observada a transversão de G para T devido à ação de espécies reativas de oxigênio. Outro ponto fundamental, sendo reflexo dessas mutações é o baixo conteúdo, cerca de 38%, de G + C do RNA genômico comparado com outros CoVs (KADAM et al, 2021; DHAMA et al, 2020).

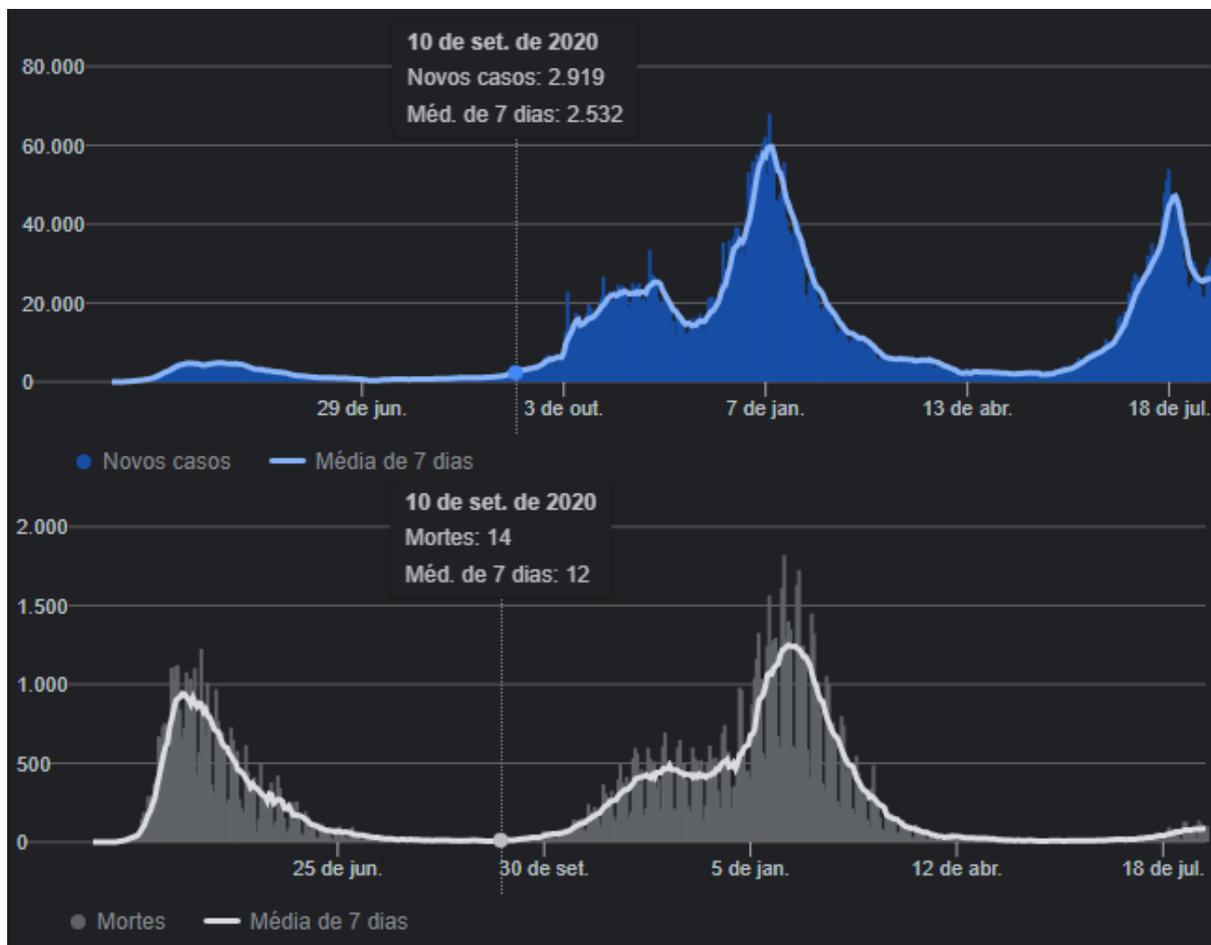
Nesse contexto, a maioria das mutações não conduzem à formação de novas proteínas, sendo que 80% não desencadeiam alterações. Ademais, foi observado que existiam locais altamente recorrentes a mutações, sendo nas proteínas acessórias nsp6, nsp11, nsp13 e na proteína estrutural S, as quais são responsáveis, respectivamente, pelo processo de replicação e invasão do hospedeiro (KADAM et al, 2021; DHAMA et al, 2020). Diante do processo de substituição e recombinação do Sars-CoV-2 houve o surgimento de diversas estirpes, algumas com maior patogenicidade, virulência e transmissibilidade, sendo denominadas de variantes preocupantes e que merecem ser destacadas, sendo a variante Alpha, originada no Reino Unido; variante Beta, originada na África do Sul; variante Gama, originada no Brasil; e variante Delta, originada na Índia (OMS, 2021).

2.5.1 – Variante Alpha – B.1.1.7 – Reino Unido

Em meados de setembro de 2020, o governo do Reino Unido foi notificado sobre o surgimento dessa variante, denominada B.1.1.7 ou 501Y.V1 (FRAMPTON et al, 2021). Foram identificados nessa estirpe 9 mutantes que incluem deleções e substituição dentro da chave que codifica a proteína estrutural S (Spike), sendo que se destacavam, principalmente, a N501Y, pois apresenta domínio dentro da RBD que interage com ACE2, e as demais mutações: del69-70HV, del144Y, A570D, D614G, P681H, T761I, S982A, D1118H (FRAMPTON et al, 2021; SOCHER et al, 2021; OSTROV, 2021; SUPASA et al., 2021).

As seguintes alterações aumentaram a patogenicidade, carga viral e transmissibilidade do vírus, o qual proporcionou a dominância da estirpe na região e impactaram, inicialmente, no Reino Unido, resultando no surgimento da segunda onda (Gráfico 2) e, posteriormente, o mundo com o aumento no número de casos, óbitos e colapso no sistema de saúde em diversas localidades, além da mudança do perfil epidemiológico da população afetada, o qual passou da população idoso para a população mais jovem (FRAMPTON et al, 2021; SOCHER et al, 2021).

Gráfico 2 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) do Reino Unido após o surgimento da variante Alpha.



Fonte: JHU CSSE COVID-19.

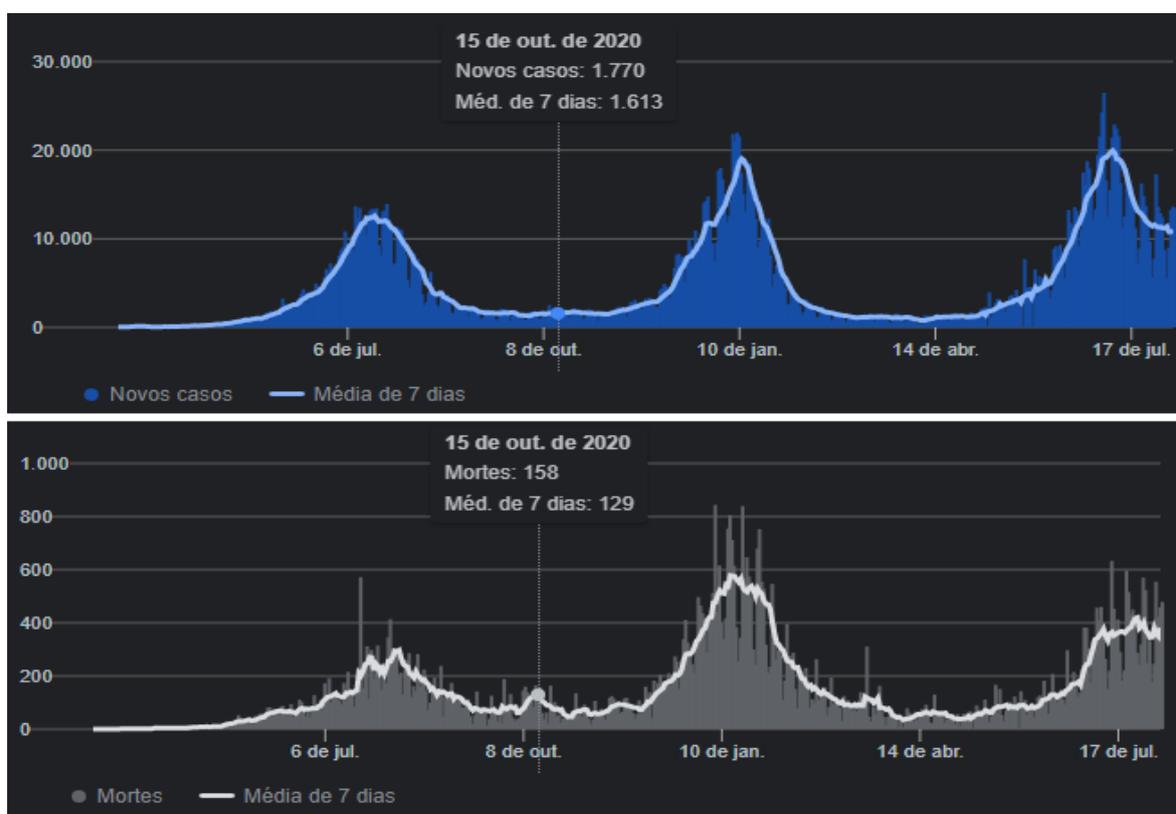
2.5.2 – Variante Beta – B.1.3.5.1 - África do Sul

Em meados de outubro de 2020 surgiu uma nova variante na África do Sul, denominada B.1.3.5.1 e, também, conhecida como 501Y.V2. Diferente da variante Alpha, na estirpe dessa região foram encontradas 10 mutações no domínio da proteína S, sendo as principais a K417N, E484K e N501Y, pois sofrem alterações na RBD e, as demais mutações: del- L242-244, L18F, D80A, D215G, R246I, D614G e A701V (LI et al, 2021; ZHOU et al, 2021). Coincidentemente, comparando com a variante B.1.1.7, essas duas estirpes compartilham de mutações semelhantes, sendo a N501Y e D614G.

Outro ponto importante é o número cerca de 19 vezes maior do número de mutações para o RBD da estirpe de Wuhan e 2,7 vezes maior do que para B.1.1.7,

aumentando a infectividade viral. Ademais, foi observada resistência parcial ao plasma convalescente e a alguns anticorpos monoclonais (mAbs), favorecendo a reinfeção; aumento da afinidade do RBD pela ACE2; aumento da transmissibilidade e virulência, o qual acarretou a dominância e desencadeou a segunda onda de casos e óbitos na África do Sul (Gráfico 3) (LI et al, 2021; ZHOU et al, 2021; MOYO-GWETE et al, 2021).

Gráfico 3 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) da África do Sul após o surgimento da variante Beta.



Fonte: JHU CSSE COVID-19.

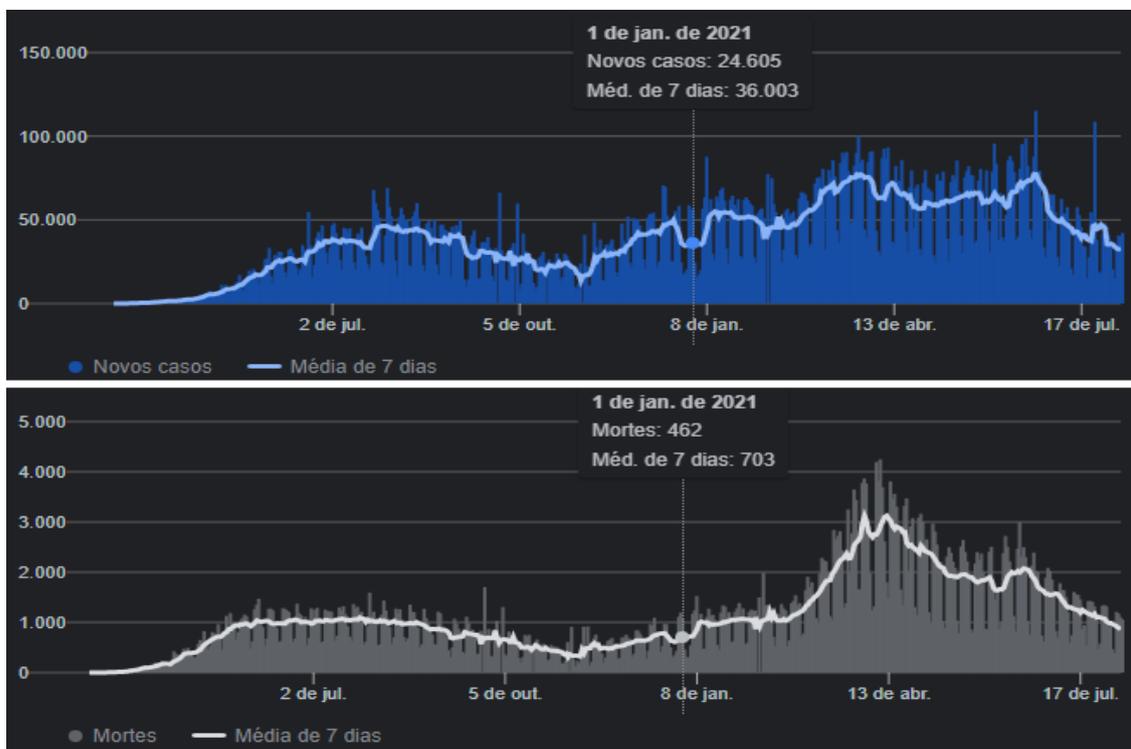
2.5.3 – Variante Gama – P.1 – Brasil

No início de 2021, em janeiro, em decorrência do descontrole da pandemia e a baixa adesão da população às medidas sanitárias, surgiu uma nova linhagem no Brasil denominada P.1 ou de 501Y.V3. Essas estirpes apresentaram 11 mutações, sendo três no domínio RBD (K417T, E484K e N501Y) e nove ao longo da proteína estrutural S (L18F, T20N, P26S, D138Y, R190S, H655Y, T1027I e V1176F) (ZHOU et al, 2021; MOYO-GWETE et al, 2021; HOFFMANN et al., 2021). Ao comparar a variante P.1 com a variante B.1.1.7 pode-se encontrar a

repetição da mutação N501Y e com a B.1.351 encontram-se as mutações N501Y, E484K, L18F. Outro ponto importante e que relaciona para as futuras variantes, é a relação do N501Y com o aumento da transmissão e o E484K como fator que influencia na resistência a mAbs (FARIA et al, 2021; HOFFMANN et al., 2021).

Ademais, semelhantemente a variante Beta, a Gama também apresenta resistência à neutralização pelos anticorpos monoclonais, associada e risco aumentado de reinfecção (KAKU et al, 2021; SOUZA et al, 2021). Ao analisar os dados epidemiológicos, houve mudança da população adoecida, passando de idosos ou pacientes com comorbidade para população jovem e hígida, também ocorreu aumento da transmissibilidade (cerca de 1,7 a 2,4 vezes em relação a não P.1), nota-se o aumento dos casos na Figura 04, e aumento do tempo de hospitalização, o qual refletiu no aumento da carga viral. O prolongamento do tempo de internação favoreceu o colapso do sistema de saúde e aumento do número de óbitos (Gráfico 4) não podendo afirmar com veemência que aumento da virulência (FARIA et al, 2021).

Gráfico 4 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) do Brasil após o surgimento da variante Gama.



Fonte: JHU CSSE COVID-19.

2.5.4 – Variante Delta – B. 1.6.1.7.2 – Índia

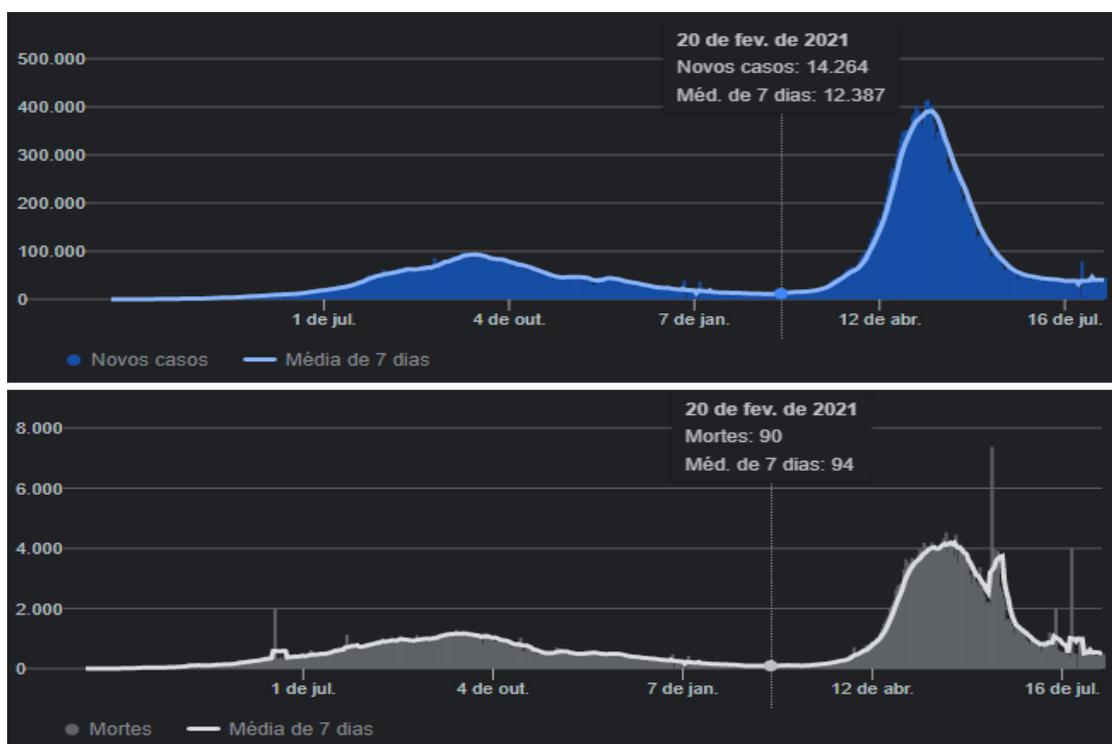
Atualmente, a variante que mais preocupa o mundo é a variante Delta ou B. 1.6.1.7.2. Identificada em outubro de 2020 na Índia, mas tornando uma variante preocupante em meados de fevereiro de 2021. Essa linhagem apresenta alta capacidade de transmissibilidade e, provavelmente, tornará a variante dominante no mundo em poucos meses (MORA et al, 2021; SALVATORE et al, 2021).

Essa estirpe apresenta 10 mutações, sendo as mais relevantes a T478K e L452R no domínio RBD, detalhe que essa variante não apresenta mutação em E484Q, e as del157-158, D614G, T19R, T478K, P681R, D950N, G142D, R158G que estão ao longo da proteína Spike (MORA et al, 2021; BERNAL et al, 2021; DAVIS et al, 2021; PLANAS et al, 2021; LI et al, 2021). Ao comparar a linhagem 1.6.1.7.2 com a B.1.1.7 pode-se identificar as mutações semelhantes, como D614G, P681H; ao se comparar com a linhagem B.1.3.5.1 encontrou-se D614G; e ao se comparar com a linhagem P.1 não se encontrou nenhuma alteração similar.

Mutações que favorecem a redução da neutralidade de anticorpos monoclonais e, possibilitando, escape imunológico, assim como, ocorre nas variantes Beta e Gama; também acarretou o aumento da transmissibilidade de 1,6 em comparação a linhagem de Wuhan, fato esse observado no aumento de casos na Índia (Gráfico 5) e, conseqüentemente, o colapso do sistema de saúde e crescimento nos números de óbitos (SALVATORE et al, 2021; YANG; SHAMAN, 2021; LI et al, 2021).

Além disso, foi observada taxa de replicação viral potencialmente mais alta da variante Delta, sugerindo que a taxa de infecção durante o estágio inicial era provavelmente maior. A frequência da triagem da população deve ser otimizada e abre um alerta devido a possibilidade de transmissão antes do início dos sintomas (LI et al, 2021).

Gráfico 5 – Relação de casos (gráfico azul) e óbitos (gráfico cinza) da Índia após o surgimento da variante Delta.



Fonte: JHU CSSE COVID-19.

3 MÉTODO

- **Obtenção dos Genomas das estirpes de Sars-CoV-2**

As sequências de nucleotídeos dos genomas das estirpes do vírus Sars-CoV-2 foram obtidas do GenBank® (<http://www.ncbi.nlm.nih.gov>) do NCBI (*National Center for Biotechnology Information*). As sequências de nucleotídeos dos respectivos genomas foram posteriormente convertidas em formato .txt ou .FASTA para melhor acesso e análise dos dados obtidos nos devidos programas de bioinformática.

Nessa etapa foram selecionadas 2.786 sequências, publicadas de dezembro a junho de 2020, as quais tiveram como critério de inclusão o tamanho do genoma superior a 29.000 pb, e de exclusão genomas que já tinham sido selecionados previamente e que apresentavam sequência inferior a 29.000 pb. Assim, foram selecionadas sequências do coronavírus provenientes de 71 países. Na busca das sequências do coronavírus no GenBank foram coletadas todas as sequências dos genomas disponíveis em publicações de grupos de pesquisa em 67 países. Já em 4 países, Austrália, Índia, China, Estados Unidos das Américas, em que as amostras passavam de 200, foram selecionadas o máximo proporcional às

amostras disponíveis. Ademais, devido à expansão da pandemia, houve a necessidade de uma nova seleção de dados, estes de julho 2020 até janeiro de 2021 obedecendo aos mesmos critérios na primeira análise.

- **Identificação de Bases Nitrogenadas**

Com a seleção dos dados foi criada uma tabela no Excel e, a partir do algoritmo The Sequence Manipulation Suite (<https://bioinformatics.org/sms/>), foram identificadas a quantidade de bases nitrogenadas e suas proporções no genoma de cada uma das estirpes obtidas. A partir da análise foi possível identificar as proporções de adenina (A), citosina (C), guanina (G) e timina (T) do genoma de cada amostra, dividi-las em seus respectivos países de origem, calcular suas médias, modas e mediana. Também foi possível calcular a média e modas por continente e a busca de uma média e moda global.

- **Identificação das Mutações no Genoma**

Usando a plataforma do programa de bioinformática Geneious (<https://www.geneious.com/>) foi possível criar uma sequência consenso para cada país analisado. Pela comparação dessas sequências com a sequência consenso da China foi possível observar alterações das bases nitrogenadas, assim como, sua localização dentro do genoma. As sequências do coronavírus publicadas pelos grupos de pesquisa da China foram usadas como padrão de comparação devido ao fato das primeiras descrições das ocorrências desse vírus naquele país.

- **Análise de Variância Molecular (AMOVA) e filogenética**

Com a formação da sequência consenso para cada nacionalidade, foi criada uma tabela, de natureza binária e, a partir do algoritmo The Sequence Manipulation Suite (<https://bioinformatics.org/sms/>) foram identificadas enzimas de restrição capazes de clivar o genoma em pontos específicos. O código 0 foi definido para a ausência do sítio de restrição e 1 para a presença do sítio de restrição para uma determinada enzima.

A partir dos dados binários foi construída uma matriz 289L x 13C que foi utilizada para o cálculo da análise da variância molecular pelo programa Arlequin (EXCOFFIER et al,

2010) e a obtenção de uma árvore filogenética pelo programa PAST (HAMMER et al, 2001) pelo método de Cluster analysis e usando o coeficiente de Jaccard.

- **Levantamento epidemiológico da dinâmica do Sars-CoV-2**

Em paralelo com a atividade *in silico* foi feito o levantamento epidemiológico da dinâmica das estirpes do Sars-CoV-2 na população. Essa estratégia foi feita por meio de uma pesquisa bibliográfica em sítios de internet que disponibilizavam literatura científica fidedigna, tais como, PubMed – NCBI (<https://www.ncbi.nlm.nih.gov/pubmed/>), Science Direct (<https://www.sciencedirect.com/>) e Mendeley (https://www.mendeley.com/?interaction_required=true).

O levantamento de dados para a composição da dinâmica epidemiológica foi complementado por sítios de internet, tais como, Coronavirus Source Data (<https://ourworldindata.org/coronavirus-source-data>) e Coronavirus Pandemic (COVID-19) (<https://ourworldindata.org/coronavirus>).

Foi realizado um levantamento de dados a respeito das mutações e dados epidemiológicos, os quais encontraram divergências entre as literaturas, pois as estirpes não experimentavam o mesmo ambiente de contágio e multifatores interferiam nas análises. Entretanto, foram observadas mutações em determinadas partes do genoma que favorecem os processos epidemiológicos.

4 RESULTADOS E DISCUSSÃO

Inicialmente foram selecionadas 2786 sequências, publicadas entre dezembro e junho de 2020, de diversos países. Entretanto, 203 sequências (7,29%) precisaram ser descartadas, pois algumas apresentavam apenas parte do material genético do vírus ou foram selecionadas mais de uma vez. Continuando a análise, devido à expansão da pandemia, houve a necessidade de uma nova seleção de dados, estes de julho 2020 até janeiro de 2021. Então, foram adicionadas ao estudo mais 753 sequências, totalizando 3.539 sequências de genomas correspondentes ao Sars-Cov-2. Outro ponto fundamental da seleção de dados foi o consenso de selecionar sequências cujo genoma fosse maior que 29.000 pb.

Após a primeira análise foi identificado que em diversas sequências encontrava-se com uma base indefinida (indicada pela letra N). Sendo assim, foi feita uma nova análise, na

qual foram removidas 153 amostras (4,32%) que não se enquadraram, e selecionadas 2.633 sequências contendo apenas as bases nitrogenadas adenina, guanina, citosina e timina.

Analisando o material genético do Sars-CoV-2 encontrou-se um tamanho médio de genoma de 29.789 pb, com as distribuições médias das bases nitrogenadas seguindo a seguinte ordem: adenina com 8.904 pb (29,89%); timina com 9.582 pb (32,16%); citosina com 5.472 pb (18,36%) e guanina com 5.847 pb (19,62%). É importante citar a que a base nitrogenada guanina apresentou moda com maior frequência (66,19%) em relação à citosina (63,38%), timina (60,56%) e adenina (59,15%).

Ao comparar a média global das amostras do genoma Sars-CoV-2 encontrou-se a soma citosina e guanina correspondendo a aproximadamente 38% do genoma total fato esse, comparado aos genomas Sars-CoV e Mers-CoV que são, respectivamente, de 41,10% e 41,22%, mostra-se inferior ao esperado. Outro fato que chamou atenção foi a quantidade de timina que chegou a 32,16%, enquanto no genoma da estirpe HCoV-HKU1 encontrou-se 40,28% e no HCoV-OC43 com 35,74%.

Em seguida, foram analisados os genomas de coronavírus obtidos de publicações de grupos de pesquisa dos seguintes países:

- 2 países da Oceania (Austrália e Nova Zelândia);
- 26 países da Ásia (Vietnã, Turquia, Timor-Leste, Tailândia, Sri Lanka, Rússia, Paquistão, Nepal, Malásia, Líbano, Jordânia, Japão, Israel, Iraque, Irã, Índia, Geórgia, Filipinas, Coreia do Sul, China, Cisjordânia, Cazaquistão, Camboja, Bangladesh, Bahrein, Arábia Saudita);
- 16 países da Europa (Suécia, Sérvia, Romênia, Reino Unido, Tcheca, Polônia, Holanda, Malta, Itália, Grécia, França, Finlândia, Espanha, Dinamarca, Bélgica, Alemanha);
- 10 países da África (Zâmbia, Tunísia, Serra Leoa, Quênia, Nigéria, Marrocos, Mali, Gana, Egito, África do Sul) e;
- 17 países da América (México, Canadá, Estados Unidos, República Dominicana, Porto Rico, Jamaica, Cuba, Guatemala, Belize, Venezuela, Colômbia, Equador, Peru, Chile, Argentina, Uruguai, Brasil);

e foram encontradas divergências entre as médias do genoma de cada continente (Quadro 2), assim como, nas nacionalidades. Um fato importante observado nessa tabela foi que apesar das variações de média, a América e Europa apresentam a mesma moda.

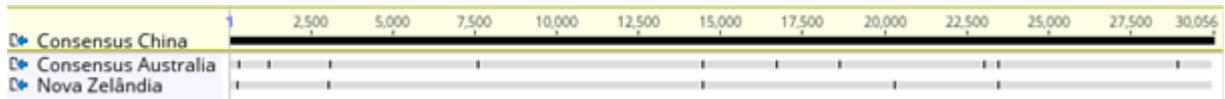
Quadro 2 – Média e moda do tamanho dos genomas de Sars-Cov-2 sequenciados em cada continente.

Localidade	Média (pb)	Moda (pb)
Mundo	29.789	29.782
América	29.729	29.782
África	29.840	29.845
Europa	29.821	29.782
Ásia	29.794	29.903
Oceania	29.817	29.813

Ao analisar separadamente cada nacionalidade encontramos tamanho médio do genoma menor na Jordânia, com 29.409 pb, e o maior na Colômbia, África do Sul e no Sri Lanka, com 29.903 pb. Curiosamente, ao observar as nacionalidades em que deram origem às novas estirpes do Sars-Cov-2 (Reino Unido 29.860 pb; África do Sul 29.903 pp; Brasil 29.868 pb; e Índia 29.798 pb) foram encontrados valores acima da média mundial (29.789 pb). Pode-se inferir que o vírus desde o começo da pandemia sofreu mutações, seja ela por deleção, substituição ou recombinação, e é apresentado de maneira diversificada de acordo com a localidade ou continente, fato esse justificando a diversidade dos padrões epidemiológicos ao comparar os países.

Assim, realizando uma análise mais específica, primeiramente na Oceania, construindo a sequência consenso, além de promover um quadro comparativo com a sequência consenso da China (Figura 6) foram encontradas 13 substituições de bases nitrogenadas na Austrália, com incidência de 30,76% mutações de C → T e G → A; sendo que também foi observado a substituição de A → G; A → T; T → C; G → T; e G → C equivalente, individualmente a 7,69%. Ademais, entre essas 13 mutações na Austrália, 53% ocorreram em ORF1ab, 23,07% na proteína estrutural spike (S) e, 23,07% na proteína estrutural nucleocapsídeo (N).

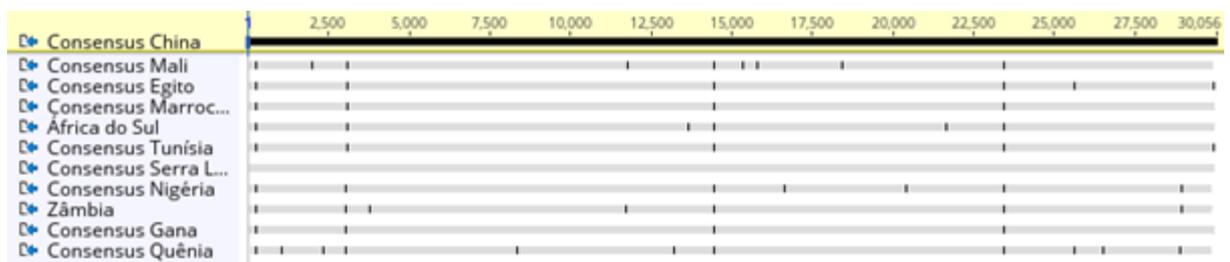
Figura 6 - Mapa comparativo entre as sequências consenso dos países da Oceania e China. Os traços verticais em cinza indicam as respectivas mutações.



Na Nova Zelândia, foram observadas 6 substituições, com a mutação predominante de C → T, equivalente a 50% e, as demais A → G, com 33,33%, e T → C, com 16,66%, sendo que 83,33% ocorriam na ORF1ab, e 16,66% na proteína estrutural spike (S).

Ao analisar a África (Figura 7), construindo a sequência consenso, além de promover um quadro comparativo com a sequência consenso da China (Figura 6) foram encontradas 9 substituições de bases nitrogenadas em Mali, com incidência de 77,77% mutações de C → T, e 11,11%, respectivamente para G → A e G → A. Ademais, entre as 9 substituições, 88,88% ocorreram em ORF1ab e 11,11% na proteína estrutural spike (S).

Figura 7 - Mapa comparativo entre as sequências consenso dos países da África e China. Os traços verticais em cinza indicam as respectivas mutações.



Observando o Egito foram encontradas 6 substituições de bases nitrogenadas, sendo 50% de C → T; 33,33% de A → G; e 16,66% de G → T. Outrossim, 50% mutações ocorrentes em ORF1ab; 16,66% em proteína estrutural spike (S); 16,66% nas unidades ORF3a, responsável por produzir as proteínas acessórias, e 16,66% em sítio 3'UTR.

Nesse contexto, foram encontradas 4 substituições de bases semelhantes em Marrocos, Tunísia e Gana, com 75% de C → T; e 25% de A → G. Outro ponto importante foi que as mutações C → T se concentraram na ORF1ab, e a mutações A → G na proteína estrutural spike (S). Já na África do Sul, berço da variante Beta, foram encontradas 6 trocas

de bases, sendo 83,33% de C → T, e 16,66% de A → G. O qual encontramos 66,66% das mutações em ORF1ab e 33,33% na proteína estrutural spike (S).

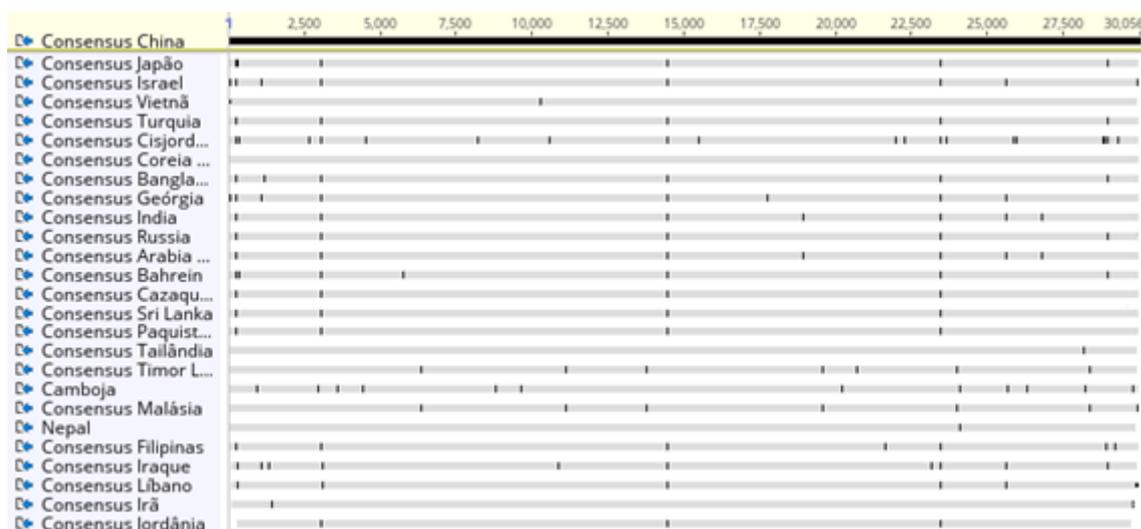
Deve-se ter um olhar atento a Quênia devido às 11 mutações encontradas, sendo 54,54% de C → T; 18,18% de T → C; 9,09% de A → G; 9,09% de G → T; 9,09% de G → A. Ademais, 54,54% das mutações ocorreram em ORF1ab; 18,18% na proteína estrutural nucleocapsídeo (N); 9,09% na proteína estrutural spike (S); e 9,09% na ORF3a; 9,09% na proteína membrana (M).

Já nas sequências obtidas do vírus em Zâmbia, também se deve ficar atendo devido à proximidade com a África do Sul, assim encontramos 10 substituições de base nitrogenada com incidência de 50% de C → T; e 20% de G → A; 10% de G → C; 10% de A → G; e 10% de T → C. Fato esse ocorrendo 60% na ORF1ab; 30% na proteína estrutural nucleocapsídeo (N); e 10% proteína estrutural spike (S).

Na Nigéria, foram identificadas 9 substituições de base nitrogenada, tendo a maior incidência, com 44,44%, em C → T. Também foi observada a substituição de A → G, com 22,22%; de G → A, com 22,22%; e de G → C, com 11,11%. Ademais, mutações ocorrendo em ORF1ab, 55,55%; três mutações consecutivas, assim como, ocorreu em Zâmbia, na proteína estrutural nucleocapsídeo (N), com 33,33% , e 11,11% das mutações ocorrendo na proteína estrutural spike (S). Por fim, não foram identificadas mutações em Serra Leoa.

Ao analisar a Ásia (Figura 8), construindo a sequência consenso, além de promover um quadro comparativo com a sequência consenso da China (Figura 6) foram encontradas 5 substituições de base nitrogenada no Japão, com incidência de 80% mutações de C → T; 20% de A → G. Ademais, entre as 5 substituições, 80% ocorreram em ORF1ab e 11,11% na proteína estrutural spike (S).

Figura 8 - Mapa comparativo entre as sequências consenso dos países da Ásia e China. Os traços verticais em cinza indicam as respectivas mutações.



Em Israel foram encontradas 6 substituições de base nitrogenada, sendo 66,66% de C → T localizadas no ORF1ab; e 16,66% de A → G localizadas na proteína estrutural spike (S); e 16,66% de G → T localizada na ORF3a. Ademais, no Vietnã, apresentaram 3 substituições de base nitrogenada, 66,66% de C → T; 33,33% de A → T, o qual apresentou 33,33% das mutações em 5'UTR; e 66,66% em ORF1ab.

Analisando a Turquia, foram encontradas 7 substituições de base nitrogenada com predominância de 42,85% de C → T, e os demais com 28,57% de G → A; 14,28% de G → C; e 14,28% de A → G. Foram observadas 42,85% das mutações em ORF1ab; 14,28% na proteína estrutural spike (S) e três variações seguidas na proteína estrutural nucleocapsídeo (N) correspondendo a 42,85%.

Foram observadas as mesmas alterações na Geórgia e Índia, este berço da variante Delta, sendo expressas por 71,42% das substituições como C → T; 14,28% de G → T; e 14,28% de A → G. Entretanto, as mutações na Geórgia se localizaram na 5UTR (14,28%); na ORF1ab (57,14%), na proteína estrutural spike (S) (14,28%), e na ORF3a (14,28%), e as mutações da Índia se localizam na ORF1ab (57,14%); na proteína estrutural spike (S) (14,28%); na ORF3a (14,28%) e na pequena membrana (E) com 14,28%.

Na Malásia foram observadas 12 substituições de base nitrogenada, sendo 33,33% de C → T; 33,33% de G → A; 8,33% de C → A; 8,33% de G → T; 8,33% de T → A; e 8,33% de C → A. Ao analisar o genoma, encontraram-se 33,33% das mutações ocorrendo em ORF1ab;

8,33% na proteína estrutural spike (S); 8,33% na proteína estrutural nucleocapsídeo (N) e 50% na terminação 3'UTR. Na Cisjordânia foram identificadas 21 substituições de base nitrogenada, sendo 33,33% de C → T; 28,57% G → T; 14,28% de G → A; 9,52% G → C; 4,76% de T → C; 4,76% de C → A; e 4,76% de A → G. Desse modo, distribuídas em 38,09% na ORF1ab, 19,04% na proteína estrutural spike (S); 9,52% na ORF3a; e 28,57% na proteína estrutural nucleocapsídeo (N).

Na Camboja foram 11 substituições de base nitrogenada, com 45,45% de C → T; 27,27% de G → T; 9,09 de A → G; 9,09 de G → A; e 9,09 de T → C. Sendo divididas em 63,63% para a porção ORF1ab; 9,09% para proteína estrutural spike (S); 9,09% para ORF3a; 9,09% para a pequena membrana (E); e 9,09% para região terminal 3'UTR.

Também foram encontradas semelhanças nas 4 substituições do Cazaquistão, Sri Lanka e Paquistão, sendo 75% de C → T encontradas na ORF1ab; e 25% de A → G encontradas na proteína estrutural spike (S). Já analisando o Timor-Leste, encontraram-se 6 substituições, sendo 66,66% de C → T; 16,66% de C → A; e 16,66% de G → T. Assim, apresentando 83% das mutações na região ORF1ab, e 16,66% na região da proteína estrutural spike (S).

No Iraque foram identificadas 9 substituições, sendo 88,88% de C → T; e 11,11% de G → T. Ao analisar o material genético encontraram-se 66,66% das mutações na ORF1ab; 22,22% na proteína estrutural spike (S); e 11,11% proteína estrutural nucleocapsídeo (N). Na Arábia Saudita encontraram-se 6 substituições de base nitrogenada, sendo 83,33% de C → T; e 16,66% de G → T, o qual tem-se 66,66% das mutações distribuídas no ORF1ab; 16,66% no ORF3a; e 16,66% na pequena membrana (E).

No Irã encontraram-se 3 substituições, sendo C → T; G → A; e G → T, o qual estão localizadas em 66,66% no ORF1ab e 33,33% na região terminal 3'UTR. Em Bangladesh, foram 8 substituições de base nitrogenada, sendo 37,5% de C → T; 25% de G → A; 12,5% de A → T; 12,5% de A → G; e 12,5% de G → C, o qual apresentam 50% na região ORF1ab; 12,5% na proteína estrutural spike (S); e 37,5% na proteína estrutural nucleocapsídeo (N). Ademais, foi observada uma substituição de base nitrogenada na Tailândia, sendo T → C localizado na proteína estrutural nucleocapsídeo (N). Diversos países seguem esse padrão de alteração.

Acerca do exposto, é notória a presença de mecanismos que possibilitam mutações no genoma do Sars-CoV-2. Um ponto interessante encontrado é a grande presença da transição de citosina para timina, fato esse ocorrendo em praticamente todos os países.

Outra mutação comum é a transversão de guanina em timina e, principalmente na Ásia. Em contrapartida, existem poucos mecanismos que transformam alguma base nitrogenada em timina, sendo observado em poucos países. Fato esse bastante importante, pois com o aumento da proporção de timina o genoma do Sars-CoV-2 pode apresentar certa estabilidade, sofrendo menos mutações e podendo se assemelhar aos vírus HCoV-HKU1 e HCoV-OC43, o qual são conhecidos por causar resfriado comum e apresentam alta proporção de timina em seu genoma (OSTROV et al., 2021; Wu et al, 2020).

Ademais, devido ao fato da proteína ORF1ab apresentar tamanho de dois terços do genoma, ela apresenta grande parte das alterações. Apesar disso, também encontrou-se substituição de bases na proteína estrutural spike (S); na proteína estrutural nucleocapsídeo (N); na ORF3a, local responsável pela produção das proteínas acessórias o que apresenta sua função pouco elucidada; pequena membrana (E) e proteína membrana (M). Também foram observadas mutações nas regiões UTR, apesar de serem regiões que não são traduzidas. Entretanto, apesar das mutações, a maioria dessas alterações não consegue expressar alguma alteração significativa na proteína final, pois não foram capazes de alteração de um novo aminoácido (KADAM et al, 2021; WAN et al, 2020).

Contribuindo a essa análise, utilizando enzimas de restrição, encontrou-se predominantemente as enzimas *AatI*, *AgeI*, *AccIII*, *BamHI*, *BbeI*, *BshTI* e *BspEI* que foram capazes de cortar o genoma, além do aparecimento de uma nova enzima que não estava presente na maioria dos genomas, como *Asp718I* na Venezuela; *AatII*, *Acc16I* na Índia e Gana; *AsuII*, na Índia e Itália, *Avill*, na Índia, Jordânia, Malásia e Itália; e *Aval*, na Itália.

Compreende-se que os dados epidemiológicos estão relacionados a diversos fatores, como as medidas de distanciamento social da população; uso de máscaras; aos casos testados e registrados, os quais podem existir diversas subnotificações; tamanho da população; densidade populacional; se existiu ou não colapso na saúde e outros fatores, dessa forma, não expressam com exatidão a manifestação daquele vírus na região, mas a conjuntura de fatores. Entretanto, os dados epidemiológicos podem ser relacionados a mutações específicas dentro do Sars-CoV-2 como, por exemplo, mutações na estrutura proteica S podem induzem aumento da transmissão e infecção; mutações na estrutura ORF1ab podem induzem aumento da carga viral, mutações em ORF3a podem induz a resistência ao sistema imunológico. Assim, estirpes que tiveram um maior potencial de mutações na proteína S, como no caso da África do Sul e Índia, em comparação a estirpe da

Arabia Saudita, podem ser mais transmissíveis e infectantes (LIMA et al, 2020; SOCHER et al., 2021).

A matriz binária produzida com a simulação da digestão de várias enzimas de restrição *in silico* dos 289 genomas dos vírus Sars-Cov-2 obtidos de vários países gerou uma árvore filogenética que permitiu observar a dinâmica entre as estirpes analisadas (Figura 9).

A análise filogenética mostrou que a maioria das estirpes isoladas de 64 países se agruparam em um único clado com similaridade em torno de 100%. Observou-se que as amostras da Argentina, Malta, Estados Unidos e Itália formaram um sub-grupo com similaridade de 70% em relação às demais estirpes desse grupo principal.

Já as estirpes da Índia formaram um segundo agrupamento com as estirpes provenientes da Malásia e Itália apresentando similaridade em torno de 80%.

E, por fim, as estirpes da Argentina e Coreia do Sul formaram um terceiro grupo com 65% de similaridade em relação às demais estirpes analisadas.

O número de enzimas de restrição que foram usadas nas simulações *in silico* não foi suficiente para mapear toda a variabilidade genética entre as estirpes de Sars-Cov-2. Isso se deve ao fato que a maioria das estirpes analisadas, provavelmente, possuem mutações pontuais em locais específicos dos seus genomas, tais como, os genes ORF e o gene que codifica a proteína spike, diferindo em poucos pares de bases, conforme o esperado para a dinâmica de dispersão dos vírus nas populações.

Contudo, essa estratégia foi capaz de discriminar as estirpes provenientes da Itália, Índia e Estados Unidos. Essas estirpes foram características com relação às suas características de infecção e dispersão.

Provavelmente, essa estratégia de simulação poderia ser ampliada para um maior número de enzimas de restrição na tentativa de discriminar outras estirpes do Sars-Cov-2 características.

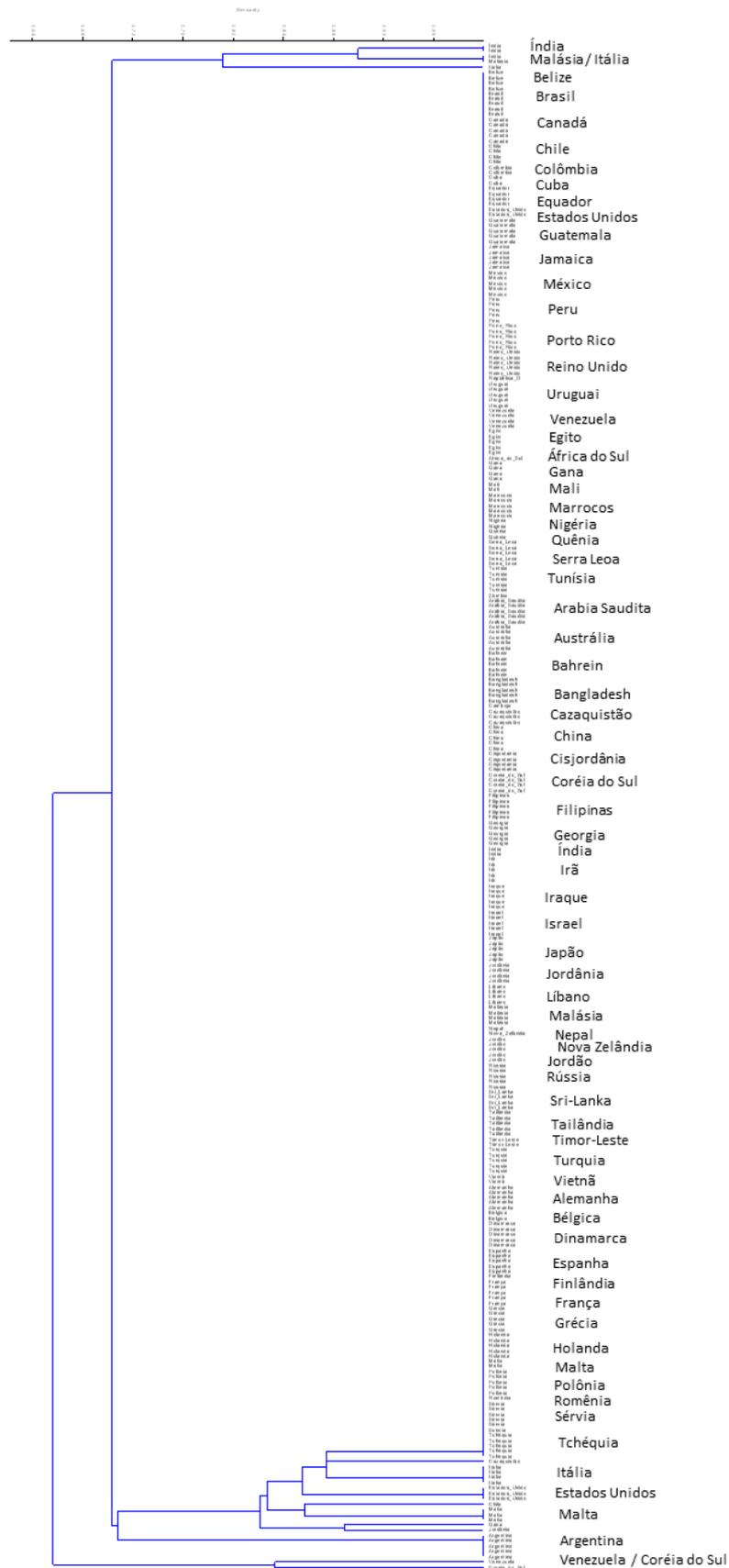


Figura 9 – Análise filogenética das estirpes de Sars-Cov-2 provenientes de 71 países que foram usadas nas simulações *in silico* com enzimas de restrição. Coeficiente de correlação de 0,97.

5 CONSIDERAÇÕES FINAIS

O novo coronavírus passou a ser conhecido em 31 de dezembro de 2019 quando a OMS foi alertada sobre vários casos de pneumonia em Wuhan na China que, inicialmente, tinha sua origem desconhecida e após o sequenciamento e análise do material genético do patógeno o novo coronavírus, foi identificado e denominado de SARS-COV-2, pertencente à família coronaviridae, da ordem Nidovirales e ao gênero Betacoronavírus, que é o maior grupo causador de infecções respiratórias e gastrointestinais. Esse vírus possui dois terços de seu genoma correspondente ao ORF1a e ORF1b que produz duas poliproteínas: PP1a e PP1ab e a partir do seu processamento, surgem dezesseis proteínas não estruturais maduras (NSPs), que são responsáveis por diferentes funções vitais, que inclui a formação do complexo replicase-transcriptase, encarregado pela replicação e transcrição do RNA viral.

O terço final do genoma, que é representado pelo ORFs, gera proteínas estruturais, elas: pequena membrana, membrana, proteína interna, nucleocapsídeo, spike e hemaglutinina esterase. Sendo quatro dessas, fundamentais para a montagem do vírion e para a infecção do coVs, sendo as proteínas spike, as responsáveis pela ligação aos receptores do hospedeiro.

Comprova-se a importância da análise molecular desse vírus, pois partindo dela torna-se possível diferenciar e minuciar as diferentes estirpes existentes, assim destacando suas diferenças e buscando por padrões existentes, que possam mostrar alterações em suas proteínas estruturais e não estruturais, sua etiopatogenia, fisiopatologia, sintomatologia e algum padrão significativo de mutação na tentativa de ajudar no estabelecimento de estratégias para manejo da população em possíveis futuros surtos.

Pode-se observar que as nacionalidades de origem das novas estirpes do Sars-CoV-2, a citar: África do sul, Brasil, Índia e Reino Unido, possuem tamanho de genoma acima da média mundial (29.789 pb), sendo seus respectivos tamanhos: 29.903 pb; 29.868 pb; 29.798 pb; e 29.860 pb. Seguindo com o estudo, utilizamos enzimas de restrição que foram majoritariamente hábeis no corte do genoma, entre elas: *AatI*, *AgeI*, *AccIII*, *BamHI*, *BbeI*, *BshTI* e *BspEI*, além de novas enzimas observadas que não estavam presentes na maioria dos demais genomas, como *AatII*, *Acc16I* na Índia e Gana; *Asp718I* na Venezuela; *AsuII* na Índia e Itália; *AvaI* na Itália; e *AviII* na Índia, Jordânia, Malásia e Itália.

Sendo possível deduzir que desde a origem deste vírus, ele sofre mutações, seja ela por substituição, deleção ou recombinação, sendo apresentada de variadas formas em

combinação com sua respectiva localidade, essa circunstância justifica a diversidade de padrões epidemiológicos existentes quando posto em comparação a diferentes regiões. Entretanto, mais estudos são necessários para ser possível compreender completamente as modificações e mutações sofridas por esse vírus que estejam associadas às complicações que ele pode causar para, assim, identificar e possivelmente traçar um correto perfil epidemiológico.

REFERÊNCIAS

1. ARTIKA, I. M.; DEWANTARI; A. K.; WIYATNO, A. Molecular biology of coronaviruses: current knowledge. *Journal Heliyon* Aug; 6(8): e04743, 2020
2. BAGGIO, Stéphanie et al. "Síndrome Respiratória Aguda Grave Coronavírus 2 (SARS-CoV-2) Carga Viral no Trato Respiratório Superior de Crianças e Adultos com Doença Coronavírus Aguda Inicial 2019 (COVID-19)." *Doenças infecciosas clínicas: uma publicação oficial da Infectious Diseases Society of America* vol. 73,1: 148-150, 2021
3. BELTRÁN-PAVEZ, Carolina et al. "Insights into neutralizing antibody responses in individuals exposed to SARS-CoV-2 in Chile." *Science advances* vol. 7,7 eabe6855. 12 Feb. 2021.
4. BERNAL, J. L. et al. "Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant." *The New England journal of medicine* vol. 385,7: 585-594, 2021
5. BARRETO et al. Colapso na Saúde em Manaus: o fardo de não aderir às medidas não farmacológicas de redução da transmissão da COVID-19. *Ciências da Saúde - Scielo Preprints*, 2021.
6. CHAN, Jasper Fuk-Woo et al. "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan." *Emerging microbes & infections* vol. 9,1 221-236. 28 Jan. 2020.
7. CHEN, Yu et al. "Emerging coronaviruses: Genome structure, replication, and pathogenesis." *Journal of medical virology* vol. 92,4: 418-423, 2020
8. DAVIS, C. et al. Reduced neutralisation of the Delta (B.1.617.2) SARS-CoV-2 variant of concern following vaccination. *MdRxiv - Cold Spring Harbor Laboratory Press*, 2021. Disponível em: <https://www.medrxiv.org/content/10.1101/2021.06.23.21259327v1>. Acesso em 28 de Julho de 2021
9. DECARO, Nicola; LORUSSO, Alessio. "Novel human coronavirus (SARS-CoV-2): A lesson from animal coronaviruses." *Veterinary microbiology* vol. 244, 2020
10. DHAMA, Kuldeep et al. "Coronavirus Disease 2019-COVID-19." *Clinical microbiology reviews* vol. 33,4 e00028-20. 24 Jun. 2020.
11. EXCOFFIER, L. et al. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 10: 564-567.
12. FARIA, Nuno R et al. "Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil." *Science (New York, N.Y.)* vol. 372,6544: 815-821, 2021
13. FRAMPTON, Dan et al. "Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-

- based cohort study." *The Lancet. Infectious diseases*, S1473-3099(21)00170-5. 12 Apr. 2021.
14. GORBALENYA, A. E.; BAKER, S. C. Baric RS. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*; 5:536–544, 2020
 15. Hammer, Ø., Harper, D.A.T., and P. D. Ryan, 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* 4(1): 9pp.
 16. HARRISON, Andrew G et al. "Mechanisms of SARS-CoV-2 Transmission and Pathogenesis." *Trends in immunology* vol. 41,12 1100-1115, 2020
 17. HOFFMANN, Markus et al. "SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies." *Cell* vol. 184,9: 2384-2393.e12, 2021
 18. ISER, Betine Pinto Moehlecke et al. Definição de caso suspeito da COVID-19: uma revisão narrativa dos sinais e sintomas mais frequentes entre os casos confirmados. *Epidemiologia e Serviços de Saúde*, v. 29, n. 3, e2020233, 2020
 19. KADAM, Swapnil B et al. "SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights." *Journal of basic microbiology* vol. 61,3 (2021): 180-202. doi:10.1002/jobm.202000537
 20. KAKU, Yu et al. "Resistance of SARS-CoV-2 variants to neutralization by antibodies induced in convalescent patients with COVID-19." *Cell reports* vol. 36,2: 109385, 2021
 21. LI, B. et al. Infecção viral e transmissão em um grande surto bem traçado causado pela variante SARS-CoV-2 Delta. MDRxiv - Cold Spring Harbor Laboratory Press, 2021. Disponível em: <<https://www.medrxiv.org/content/10.1101/2021.07.07.21260122v2>>. Acesso em 28 de Julho de 2021
 22. LI, Qianqian et al. "SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape." *Cell* vol. 184,9: 2362-2371.e9, 2021
 23. LIMA, C. M. A. O. Information about the new coronavirus disease (COVID-19). *Radiologia Brasileira*. v. 53, n. 2, pp. V-VI, 2020.
 24. LIU, D. X; LIANG, J. Q.; FUNG, T. S. "Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (*Coronaviridae*)." *Encyclopedia of Virology*: 428–440, 2021
 25. MALLAH, Saad I et al. "COVID-19: breaking down a global health crisis." *Annals of clinical microbiology and antimicrobials* vol. 20,1 35. 18 May. 2021.
 26. MALIK, Y. A. "Properties of Coronavirus and SARS-CoV-2." *The Malaysian journal of pathology* vol. 42,1: 3-11, 2020
 27. MORA, E. L et al. Emergencia de variante Delta - B.1.617.2. Su impacto potencial en la evolución de la pandemia por SARS-CoV-2. *Bol. Micol.*; 36(1):12-16, 2021

28. MOYO-GWETE, Thandeka et al. "SARS-CoV-2 501Y.V2 (B.1.351) elicits cross-reactive neutralizing antibodies." *bioRxiv : the preprint server for biology* 2021.03.06.434193. 11 Mar. 2021,
29. OMS. Tracking SARS-CoV-2 variants, 2021. Disponível em: <<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>>. Acesso em 21 de Julho de 2021.
30. OPAS/OMS Brasil. **Folha informativa - COVID-19 (Doença causada pelo novo coronavírus)**, 2020. Disponível em: <https://www.paho.org/bra/index.php?option=com_content&view=article&id=6101:covid19&Itemid=875>. Acesso em: 16 de maio e 2020
31. ORTIZ-PRADO, Esteban et al. "Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review." *Diagnostic microbiology and infectious disease* vol. 98,1: 115094, 2020
32. OSTROV, David A. "Structural Consequences of Variation in SARS-CoV-2 B.1.1.7." *Journal of cell immunology* vol. 3,2: 103-108, 2021
33. PAL, Mahendra et al. "Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): An Update." *Cureus* vol. 12,3 e7423. 26 Mar. 2020.
34. PLANAS, D. *et al.* Sensibilidade reduzida de SARS-CoV-2 variante Delta à neutralização de anticorpos. *Nature* 596, 276–280, 2021
35. RIO, C. D.; MALANI, P. N. COVID-19-New Insights on a Rapidly Changing Epidemic. *JAMA*. 2020
36. SALIAN, Vrishali S et al. "COVID-19 Transmission, Current Treatment, and Future Therapeutic Strategies." *Molecular pharmaceuticals* vol. 18,3: 754-771, 2021
37. SALVATORE, M et al. Resurgence of SARS-CoV-2 in India: Potential role of the B.1.617.2 (Delta) variant and delayed interventions. *MdRxiv - Cold Spring Harbor Laboratory Press*, 2021. Disponível em: <<https://www.medrxiv.org/content/10.1101/2021.06.23.21259405v1.full>>. Acesso em 28 de Julho de 2021
38. SILVA, Rodrigo. **Muitos países erram na luta contra o coronavírus. Ninguém errou mais que o governo chinês.** São Paulo: 2020. Disponível em:<https://www.youtube.com/watch?v=_V4r5ibOm5g> Acesso em: 2 de maio de 2020
39. SOCHER, Eileen et al. "Mutations in the B.1.1.7 SARS-CoV-2 Spike Protein Reduce Receptor-Binding Affinity e Induce a Flexible Link to the Fusion Peptide." *Biomedicines* vol. 9.5 525. 8 de maio. 2021.
40. SOUZA, William M et al. "Neutralização da linhagem SARS-CoV-2 P.1 por anticorpos produzidos por meio de infecção natural de SARS-CoV-2 ou vacinação com uma

- vacina SARS-CoV-2 inativada: um estudo imunológico." *The Lancet. Micróbio* , 10.1016 / S2666-5247 (21) 00129-4. 8 de julho de 2021
41. SUPASA, Piyada et al. "Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera." *Cell* vol. 184,8: 2201-2211.e7, 2021.
 42. The Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Coronavirus COVID-19 (2019-nCoV). Disponível em: <<https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6>>. Acesso em 31 de Julho de 2021.
 43. TUFAN, Z. K.; KAYAASLAN, B.. "Esmagando a curva, o papel das instituições nacionais e internacionais e formuladores de políticas na pandemia COVID-19." *Jornal turco de ciências médicas* vol. 50, SI-1 495-508. 21 de abril de 2020
 44. UZUNIAN, A. Coronavirus SARS-CoV-2 and Covid-19. *Jornal Brasileiro de Patologia e Medicina Laboratorial*. v. 56, e3472020, 2020.
 45. VALENCIA, Damian N. "Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2." *Cureus* vol. 12,3 e7386. 24 Mar. 2020
 46. WAN, Y. et al.. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J Virol*. 94(7):127-20, 2020
 47. WU, Aiping et al. "Composição do genoma e divergência do novo coronavírus (2019-nCoV) originado na China." *Cell host & microbe* vol. 27,3: 325-328, 2020
 48. YANG, W.; SHAMAN, J.. Dinâmica da pandemia de COVID-19 na Índia e impacto da variante SARS-CoV-2 Delta (B.1.617.2). *MdRxiv - Cold Spring Harbor Laboratory Press*, 2021. Disponível em: <<https://www.medrxiv.org/content/10.1101/2021.06.21.21259268v1>>. Acesso em 28 de Julho de 2021
 49. YUKI, Koichi et al. "COVID-19 pathophysiology: A review." *Clinical immunology (Orlando, Fla.)* vol. 215: 108427, 2020
 50. ZHOU, Daming et al. "Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera." *Cell* vol. 184,9: 2348-2361.e6, 2021