UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE MEDICINA
NÚCLEO DE PESQUISA E DESENVOLVIMENTO DE MEDICAMENTOS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS

COVID-19: ASPECTOS DO DIAGNÓSTICO E VIGILÂNCIA GENÔMICA DAS LINHAGENS CIRCULANTES NO ESTADO DO CEARÁ

LUÍNA BENEVIDES LIMA

FORTALEZA-CE

2022

UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE MEDICINA
NÚCLEO DE PESQUISA E DESENVOLVIMENTO DE MEDICAMENTOS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS

COVID-19: ASPECTOS DO DIAGNÓSTICO E VIGILÂNCIA GENÔMICA DAS LINHAGENS CIRCULANTES NO ESTADO DO CEARÁ

LUÍNA BENEVIDES LIMA

Tese apresentada ao Programa de Pós-Graduação em Ciências médicas da Universidade Federal do Ceará como parte dos requisitos para obtenção do título de Doutora em Ciências Médicas.

Orientadora: Professora Dr.ª Raquel Carvalho Montenegro

FORTALEZA-CE

2022

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

L698c Lima, Luina Benevides.
Covid-19: Aspectos do diagnóstico e vigilância genômica das linhagens circulantes no estado do Ceará / Luina Benevides Lima. – 2022.
96 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Faculdade de Medicina, Programa de Pós-Graduação em Ciências Médicas, Fortaleza, 2022.
Orientação: Profa. Dra. Raquel Carvalho Montenegro.

1. genoma. 2. variantes. 3. sars-cov-2. I. Título.

CDD 610

LUÍNA BENEVIDES LIMA

COVID-19: ASPECTOS DO DIAGNÓSTICO E VIGILÂNCIA GENÔMICA DAS LINHAGENS CIRCULANTES NO ESTADO DO CEARÁ

Tese apresentada à Coordenação do Programa de Pós-Graduação em Ciências Médicas da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutora em Ciências Médicas.

Orientadora: Professora Dr.ª Raquel Carvalho Montenegro

Aprovada em 20/05/2022

BANCA EXAMINADORA

_____
Profª Drª Raquel Carvalho Montenegro (Orientadora)
Universidade Federal do Ceará

_____
Profª Drª Silvia Helena Barem Rabenhorst
Universidade Federal do Ceará

_____
Profª Drª Debora Castelo Branco de Souza Collares Maia
Universidade Federal do Ceará

_____
Profª Drª Cristiana Libardi Miranda Furtado
Universidade de Fortaleza

_____
Profª Drª Rommel Mario Rodríguez Burbano
Universidade Federal do Pará

AGRADECIMENTOS

À Deus, por nunca ter soltado a minha mão. Por ter me dado pais maravilhosos, uma irmã incrível e por ter me protegido, cuidado e amado todos os meus dias! Sou grata pela vida, pelo que aprendi, pessoas que conheci, e por tudo que Ele me permitiu vivenciar.

Aos meus pais Paulo e Selma, obrigada por terem me passado bons valores. Por terem priorizado minha educação e estímulos desde a infância, não só na escola em si, mas também na música e no esporte. Obrigada pela infância maravilhosa e pelo cuidado e amor diário.

À minha irmã Joyce, que sempre foi meu braço direito! Me inspiro em você sempre, sister! Obrigada por toda a nossa amizade, mas, especialmente, por dividirmos os desafios da maternidade diariamente. Você sempre foi essencial para mim, e conseguiu ser mais ainda depois que a Luninha nasceu.

Ao meu esposo Felipe, pelo nosso amor, cuidado e parceria. Obrigada meu amor! Você e nossa filha Luna são a concretização do amor de Deus na minha vida! Sou grata por vocês, pela nossa família, pelo nosso dia-dia! Obrigada por dar o seu melhor a cada dia!

À minha filha Luna, pela sua pureza e amor! Você veio no meio de uma pandemia, trazendo esperança e renovando tudo em nós. Você ressignificou minha vida e meus valores, e me faz querer ser uma pessoa melhor a cada dia!

À minha orientadora Raquel Montenegro, agradeço aos ensinamentos, a todo o aprendizado ao longo desses quatro anos e pelo diálogo aberto que sempre tivemos. Obrigada por ter sido uma orientadora maravilhosa!

À Dra. Elisabete e ao Dr. Odorico, pela oportunidade de trabalhar no NPDM, cercada por tantos pesquisadores incríveis. Vocês são exemplos para mim na ciência! Obrigada por tanto aprendizado.

À Laís Brasil, minha amiga e parceira no dia-dia do laboratório. Buscamos juntas a convocação no concurso da UFC e seguimos trabalhando juntas, dividindo não só as tarefas do trabalho, mas também os perrenques da vida pessoal! Obrigada por tudo!!

Ao Emerson Lucena, grande amigo que o laboratório me deu, também sou grata pela amizade sincera! Você é um presente pra mim, amigo! Obrigada pelo companheirismo e carinho!

À Andrea Oliveira, Maísa Holanda e Mariana Brito, meu muito obrigada por toda a colaboração para a realização deste trabalho! Agradeço não só pelo trabalho em equipe que fizemos, mas também pela convivência agradável!

À prof. Vânia Melo, que me orientou durante a graduação e mestrado e que continua presente na minha vida acadêmica compartilhando seus conhecimentos e me ensinando.

Ao Farmagen e ao Núcleo de Pesquisa e Desenvolvimento de Medicamentos, local onde desenvolvo meu doutorado e sou lotada como funcionária da UFC, gratidão! Sigo aprendendo todos os dias algo novo e sendo constantemente desafiada. Agradeço aos amigos que fiz aqui, especialmente: Carlos Payer, Caroline Aquino, Aline Diogo, Jean Silveira e Roberta Taiane.

Ao Programa de Pós-Graduação em Ciências Médicas e todos os seus professores, pela oportunidade de cursar o doutorado e também pelos aprendizados que tive durante esses anos. À secretária do programa, Ivone, por toda a atenção e prontidão em solucionar os problemas que surgiram ao longo desses quatro anos.

Ao CNPQ e à FUNCAP, órgãos de fomento que tornaram possível realizar esse trabalho.

RESUMO

O vírus SARS-CoV-2 foi identificado em dezembro de 2019 e se espalhou rapidamente pelo mundo. Em janeiro de 2020, o primeiro genoma estava disponível, com aproximadamente 30000 bases. Desde então, o vírus acumulou diversas mutações, levando ao surgimento de diferentes linhagens. Neste trabalho, foi realizada uma revisão sobre o método padrão ouro para a detecção do SARS-CoV-2, abordando diversos fatores que podem contribuir para a obtenção de resultados falsos negativos ou falsos positivos. Uma análise *in silico* identificou a formação de homodímeros de iniciadores e/ou sondas nos kits de diagnóstico China CDC (ORF1ab target), Charité (gene E) e HKU (gene N); e heterodímeros nos kits China CDC (gene N), Charité (gene E), e US CDC (genes N2 e N3), sendo esses os kits mais passíveis de gerarem resultados falsos. Neste trabalho, também foi realizado o sequenciamento de 34 genomas do SARS-CoV-2, isolados entre julho/20 e julho/21 no Ceará. Essas sequencias, analisadas em conjunto com sequencias depositadas no banco de dados GISAID neste período para o Ceará, revelaram a ocorrência de 8 linhagens entre março e dezembro de 2020. A linhagem mais prevalente foi a B.1.1.33 (n=45), seguida por B.1 (n=21), B.1.212 (n=19) e P.2 (n=14). Além da identificação das linhagens circulantes, esse trabalho proporcionou a identificação de 202 SNVs (Variantes de Nucleotídeo Único), entre os 34 genomas sequenciados. Mutações com >40% de prevalência incluem p.F924F e p. P4715, ambos em ORF1ab; p.D614G no gene S; p.I33T em ORF6; p.R203K, p.R203R, p.G204R e p.I292T no gene N. A única mutação detectada em todos os 34 genomas sequenciados foi a D614G no gene S. A vigilância genômica das linhagens circulantes nos permite monitorar a pandemia, auxiliando na tomada de decisões do governo e no rastreamento do vírus. No banco de dados GISAID, poucos genomas de SARS-CoV-2 são provenientes de amostras coletadas nos primeiros meses da pandemia no Ceará. Nossos dados visam preencher essas lacunas e contribuir para o melhor entendimento da diversidade genética das variantes e sua prevalência no estado do Ceará.

Palavras-chaves: SARS-COV-2; genoma; variantes.

ABSTRACT

SARS-CoV-2 virus emerged in December 2019 and has rapidly spread around the world. In January 2020, the first genome sequence was available, with approximately 30000 bases. Since then, the virus accumulated diverse mutations, leading to the emergence of different lineages. In this work, a review was carried out on the gold standard method for SARS-CoV-2 detection, addressing the factors that may contribute to the achievement of false negative or false positive results. An in-silico analysis identified the formation of primer and/or probe homodimers in the China CDC (ORF1ab target), Charité (E gene) and HKU (N gene) diagnostic kits; and heterodimers in the China CDC (N gene), Charité (E gene), and US CDC (N2 and N3 genes) kits, these being the kits more likely to generate false results. In this work, 34 genomes of SARS-CoV-2 were sequenced, from samples obtained between July/20 and July/21 in Ceará. These sequences, analyzed together with sequences deposited in the GISAID database in this period for Ceará, revealed the occurrence of 8 lineages between March and December 2020. The most prevalent strain was B.1.1.33 (n=45), followed by B.1 (n=21), B.1.212 (n=19) and P.2 (n=14). In this work, 202 SNVs (Single Nucleotide Variant) were identified among the 34 sequenced genomes. Mutations with >40% prevalence include p.F924F and p. P4715, both in ORF1ab; p.D614G in the S gene; p.I33T in ORF6; p.R203K, p.R203R, p.G204R and p.I292T in the N gene. The only mutation detected in 100% of our sequenced genomes was D614G in the S gene. Genomic surveillance of circulating lineages allows us to monitor the pandemic, assisting in government decision-making and in monitoring the virus. In GISAID database, few SARS-CoV-2 genomes come from samples collected in the first months of the pandemic in Ceará. Our data aim to fill these gaps and contribute to a better understanding of SARS-CoV-2 variants genetic diversity and their prevalence in Ceará.state.

Key-words: SARS-CoV-2; genome, variant.

LISTA DE ILUSTRAÇÕES

Capítulo I: Fundamentação teórica

Capítulo II: True or False: What are the factors that influence COVID-19 diagnosis by RT-qPCR?

Capítulo III: Genomic Surveillance: Circulating lineages and genomic variation of the SARS-CoV-2 in Ceara state, Northeast Brazil

LISTA DE ABREVIATURAS E SIGLAS

OMS	Organização Mundial da Saúde
WHO	World Health Organization
EPIs	Equipamentos de Proteção Individual
CoVs	Coronavírus
SARS-CoV-1	Severe Acute Respiratory Coronavirus-1
MERS-CoV	Middle East Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Coronavirus-2
ORFs	Open Reading Frame
RBD	Receptor Binding Domain
ACE2	Angiotensin Converting Enzyme 2
RT-qPCR	Reverse Transcription-Polymerase Chain Reaction
RT-LAMP	Reverse Transcription Loop-Mediated Isothermal Amplification
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
SHERLOCK	Specific High-sensitivity Enzymatic Reporter unLOCKing
cDNA	DNA complementar
ECDC	Centro Europeu para Controle e Prevenção de Doenças
VOCs	Variants of Concern
VOIs	Variants of Interest
VUM	Variants under Monitorin
CDC	Centers for Disease Control and Prevention
FDA	Food and Drug Administration
SNV	Single Nucleotide Variant
Cq	Quantification Cycle
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
PANGOLIN	Phylogenetic Assignment of Named Global Outbreak Lineages

SUMÁRIO

CAPÍTULO I

Fundamentação teórica

1. Introdução

Em dezembro de 2019, surgia o SARS-CoV-2, vírus responsável pela doença COVID-19, relatado pela primeira vez em Wuhan, China. O vírus rapidamente se tornou uma questão de saúde pública mundial, devido à sua severidade clínica e sua fácil transmissibilidade (YANG *et al*, 2020). Apenas três meses depois do primeiro caso relatado, em 11 de março de 2020, a Organização Mundial da Saúde (OMS) declarou a pandemia do COVID-19, com casos registrados em diversos países. Até abril de 2022, o SARS-CoV-2 já havia infectado mais de 497 milhões de indivíduos e causado mais de 6,1 milhões de mortes (WHO, 2022).

A pandemia provocou um grande impacto na sociedade, levando a medidas como distanciamentos sociais, uso de máscaras, quarentenas, proibição de vôos, além de ter impactado a economia mundial, afetando diversos negócios. A demora pela implementação de medidas de segurança provocou, em diversos países, a superlotação em hospitais. Um esforço mundial foi realizado para a fabricação e comercialização de equipamentos para o suprimento de oxigênio a pacientes, Equipamentos de Proteção Individual (EPIs), leitos de hospitais e obras emergenciais foram realizadas para a construção de hospitais temporários para que se conseguisse atender um maior número de pacientes nos períodos de picos da pandemia (AQUINO *et al*, 2020).

Uma vez que o mundo percebeu a dimensão e gravidade da pandemia da COVID-19, a busca por um tratamento eficaz ou por medidas preventivas (vacinas) passou a ser a prioridade mundial. Diversos grupos desenvolveram, em paralelo, vacinas eficazes e capazes de conter o avanço do vírus (KRAMER, 2020) e, apenas 272 dias após a WHO (World Health Organization) declarar a pandemia do COVID-19, a fabricante Pfizer iniciou a aplicação de vacinas no Reino Unido (CASTRO & SINGER, 2021).

Em todos os países, programas de vacinação foram elaborados criando grupos prioritários, utilizando critérios com base na idade ou em condições patológicas, por exemplo, de forma a contribuir para a contenção do vírus, diminuição da sobrecarga de leitos dos hospitais, diminuição de casos e de óbitos causados pela doença, além da retomada das atividades econômicas. Apesar dos avanços na conscientização da população mundial e nos programas de vacinação, o surgimento de novas variantes é um desafio permanente, uma vez que erros na replicação viral, constantemente, trazem à tona novas mutações para o SARS-CoV-2, que podem vir a favorecer mecanismos de escape ao sistema imune ou aumentar a patogenicidade das linhagens circulantes (SIQUEIRA *et al*, 2022).

Considerando as últimas duas décadas, o SARS-CoV-2 é o terceiro vírus a ameaçar a população mundial com características pandêmicas. Além do SARS-CoV-2, tivemos o SARS-CoV e o MERS-CoV, todos pertencentes à mesma família de vírus e com similaridades genômicas entre si (SOUZA *et al*, 2021).

1.1 Família Coronavírus

A família dos Coronavirus (CoVs) compreende vírus de RNA fita simples positiva e envelopados. Os vírus pertencentes a essa família podem ser classificados em quatro gêneros: *Alpha, Beta, Gamma* e *Delta.* Os gêneros mais conhecidos são os *Alpha* e *Beta*, devido à sua habilidade de atravessar a barreira de espécies (SOUZA *et al*, 2021; JIAO *et al*, 2021).

Animais como morcegos abrigam uma grande diversidade de coronavírus, e a transmissão de um animal vertebrado para os humanos constitui a fonte da atual infecção do coronavírus, podendo ser também a fonte de potenciais futuros surtos (BEAN & SAGAR, 2021).

Em abril de 2002, o vírus causador da doença *Severe Acute Respiratory Coronavirus-1* (SARS-CoV-1) emergiu em Guangsdong Province/China, causando infecções com uma taxa de mortalidade de aproximadamente 10%. Com período de incubação de aproximadamente 5 dias, cerca de 95% dos pacientes desenvolviam a doença após 13 dias de exposição. Até março de 2003, o vírus já tinha atingido 29 países. Esse vírus apresenta uma patogênese única, por ser capaz de causar infecções nos tratos respiratórios superiores e inferiores. Medidas de contenção rapidamente foram tomadas, e a transmissão foi interrompida. Em 2012, outra síndrome respiratória surgia em Zarqa-Jordan, chamada *Middle East Respiratory Syndrome* e causada pelo MERS-CoV, também pertencente à família dos Coronavírus (BEAN & SAGAR, 2021; JIAO *et al*, 2021). Em setembro de 2012, o vírus já havia atingido 27 países. Diferente do SARS-CoV-1 e do SARS-CoV-2, que utilizam o receptor ACE2 para entrar na célula hospedeira humana, o MERS-CoV faz uso do receptor DPP4. Pacientes infectados por este vírus eram, em sua maioria, homens (65%), e tinham uma idade média de 49 anos. Além disso, o vírus possui um tempo de incubação de 5.2 dias, atingindo sua maior carga viral com 12 dias de incubação. A infecção pelo MERS-CoV permaneceu geograficamente restrita, porém ainda há relatos de casos até os dias atuais. (SOUZA *et al*, 2021)

Em relação ao SARS-CoV, os morcegos devem ser os hospedeiros naturais do vírus, e as civetas (*Paguma larvata*), os hospedeiros intermediários. Já para o vírus MERS-CoV, os dromedários (*Camelus dromedarius*) foram os hospedeiros intermediários (CUI *et al*, 2019).

O surgimento do vírus SARS-CoV-2 teve um impacto mundial bem maior do que os demais vírus acima citados, dada a sua rápida transmissão, além da demora da implementação de medidas de segurança pública. Os desfechos da doença variam entre casos assintomáticos a óbitos, sendo influenciados por outras doenças que ocorram simultaneamente no paciente. Os três vírus citados pertencem ao gênero β-coronavírus, sendo este gênero conhecido por serem patógenos de animais, causando doenças respiratórias e gastroenterites (JIAO *et al*, 2021).

Vírus pertencentes à família Coronavírus representam, portanto, um grupo de vírus com potencial para provocar novos surtos e/ou doenças de impacto global. Estudos genéticos são de supra importância, pois permitem a identificação de mutações (*Single Nucleotide Variant* - SNV) e de regiões conservadas compartilhadas, importantes para a vigilância genômica das linhagens circulantes.

1.2 SARS-CoV-2

O vírus SARS-CoV-2 é um vírus de RNA pertencente ao gênero beta-coronavírus (linhagem 2B), o mesmo ramo evolutivo do SARS-CoV-1 e MERS-CoV. Apesar disso, o SARS-CoV-2 compartilha apenas 79% de identidade genética com o SARS-CoV-1 e apenas 50% com o MERS-CoV. As diferenças genéticas entre os genomas desses vírus estão localizadas principalmente no domínio N-terminal e na sequência de nucleotídeos na região do RBD (*Receptor Binding Domain*), que estão envolvidos no reconhecimento do receptor celular (SOUZA *et al*, 2021).

Devido à ausência de um mecanismo de reparo, seu processo de replicação tem uma alta taxa de mutação, propiciando o surgimento de novas mutações, que podem vir a se estabelecer nas linhagens circulantes, trazendo vantagens competitivas ao vírus (ZHOU *et al*, 2021). Apesar das diferenças entre os genomas, os sintomas clínicos da infecção por SARS-CoV-2 são semelhantes às infecções por SARS-CoV-1 e MERS-CoV, sendo caracterizada tipicamente por sintomas respiratórios. Além disso, reações inflamatórias profundas das vias aéreas e danos alveolares também são comuns (XU *et al*, 2020)

O vírus responsável pela COVID-19 atinge os hospedeiros através de gotículas respiratórias do ar, aerossol, contato direto com superfícies contaminadas e transmissão fecal-oral. Os alvos principais de entrada do vírus compreendem o trato respiratório, células epiteliais alveolares, células vasculares endoteliais e macrófagos alveolares (SABIONI *et al*, 2021).

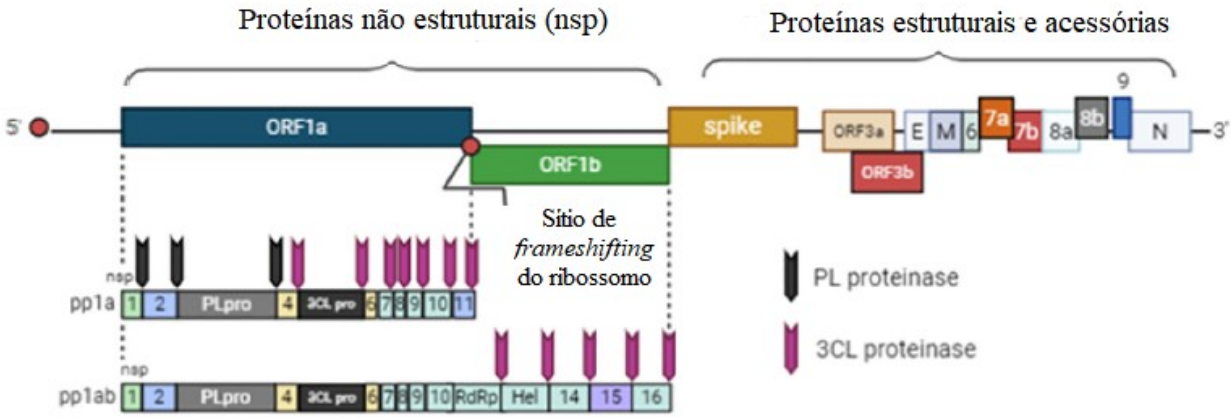
Diferente do SARS-CoV-1 e do MERS-CoV, o vírus SARS-CoV-2 atingiu todos os continentes rapidamente, demonstrando uma capacidade de contágio elevada. A sua alta

transmissibilidade é influenciada em grande parte por mutações acumuladas na proteína S, o que levaram essa proteína a ser 20 vezes mais eficaz em reconhecer os receptores humanos, quando comparada à proteína S dos vírus SARS-CoV-1 e do MERS-CoV (SOUZA *et al*, 2021).

1.2.1 Estrutura genômica do SARS-CoV-2

O genoma dos coronavírus é composto pela ORF1ab (*Open Reading Frame* ou Fase de Leitura Aberta), responsável por 16 proteínas não estruturais (NSP), seis proteínas acessórias, glicoproteína de superfície (S), proteína envelope (E), proteína de membrana (M) e proteína do nucleocapsídeo (N) (TIMMERS *et al*, 2021), conforme pode ser visualizado na Figura 1.

Figura 1. Estrutura genômica dos Coronavírus.



Fonte: Adaptado de Souza *et al* (2021).

As proteínas codificadas pela ORF1ab desempenham um papel importante na síntese de RNA viral. A ORF1 funciona como replicase, replicase/transcriptase ou polimerase e é traduzida em ORF1a (~486 kDa, produto principal) e ORF1b (~306KDa) na célula hospedeira. A ORF1a compreende 10 NSPs, com importante papel em situações de estresses celulares e manutenção da integridade funcional dos componentes celulares, além de atuar na replicação viral. Já a ORF1b, codifica RNA polimerase dependente de RNA viral (NSP 12), helicase (NSP 13), exonuclease (NSP 14), uma endonuclease específica de poliU (Uridilato) (NSP 15) e metiltransferase (NSP 16). Essas SNVs são, portanto, fundamentais na patogênese viral e representam também um promissor alvo para fármacos antivirais/síntese de vacinas (GORDON *et al*, 2020; SHUVAN *et al*, 2020)

No que tange às proteínas estruturais, os coronavírus apresentam quatro principais: Proteína de membrana (M), Envelope (E) e Nucleocapsídeo (N) e Spike (S).

A proteína M é composta por 221 aminoácidos ou 25-30kDa em seu estado pré-glicosilação, estando relacionada com o processo de formação de novos vírus. Essa proteína apresenta baixa variabilidade entre os Beta Coronavírus, podendo estar envolvida em funções *house keeping* (BIANCHI *et al*, 2020; VALE et al, 2021; SOUZA *et al*, 2021).

A proteína E é composta por 76-109 aminoácidos ou 8.5 kDa sendo a menor de todas as proteínas estruturais (PARK *et al*, 2021). É altamente conservada entre os coronavírus e funciona como uma viporina canalizadora de íons, rompendo as membranas do hospedeiro e permitindo a liberação dos vírus. Ela também interage com outras proteínas tanto do vírus como do hospedeiro, participando do processo de formação de novas partículas virais e liberação (BIANCHI *et al*, 2020; TIMMERS *et al*, 2021).

A proteína N é composta por 419 aminoácidos e tem peso molecular de 43 a 50 kDa, sendo maior que a proteína M e E, porém menor que a proteína S. Essa proteína está relacionada aos processos de transcrição do RNA e replicação, sendo altamente conservada (TIMMERS *et al*, 2021; VALE et al, 2021).

A proteína S é a maior proteína dos coronavírus, tendo 1273 aminoácidos e um peso molecular de 180-200 kDa, além de um alto nível de glicosilação em sua estrutura. Essa é a proteína mais abundante do envelope do vírus. Supõe-se que a função biológica dessa proteína é a evasão do sistema imune do hospedeiro, além de ser fundamental no processo de entrada do vírus na célula do hospedeiro (TIMMERS *et al*, 2021).

No processo de entrada do vírus em uma célula hospedeira, a proteína S se liga aos receptores da célula humana denominados ACE2 e TMPRSS2. A proteína S é composta por duas subunidades (domínios S1 e S2). O SARS-CoV-2 utiliza a região RBD do domínio S1 para se ligar ao receptor ACE2 (*angiotensin-converting enzyme 2*), o que estimula o TMPRSS2 a clivar a proteína S. Esses eventos culminam na fusão das membranas celulares e virais, possibilitando a entrada do vírus na célula hospedeira. Portanto, a ligação da proteína S ao receptor ACE2 é a primeira etapa-chave que permite que o vírus entre nas células alvo (SOROKINA *et al*, 2020; WALLS *et al*, 2020).

Após a fusão das membranas celulares, ocorre a tradução do gene da replicase do RNA genômico do virus, e a tradução e montagem dos complexos de replicase viral. Em seguida, novos vírus maduros são formados, passando pela etapa de encapsidação. Após a montagem, os vírus são transportados para a superfície celular em vesículas e liberados por exocitose (MALIK, 2020).

O conhecimento da estrutura genômica do SARS-CoV-2, bem como de suas regiões mais conservadas e mais variáveis, tornou possível desenvolver métodos moleculares de diagnósticos específicos, sendo fundamental para o monitoramento da pandemia (LIU *et al*, 2020).

1.2.2 Variabilidade genética no SARS-CoV-2

A taxa evolutiva média para o SARS-CoV-2 é de aproximadamente $8x10^{-4}$ substituições de nucleotídeos por sítio por ano, ou seja, uma mutação a cada duas semanas (SOFI *et al*, 2020). Diversas mutações já foram descritas ao longo de todo o genoma do vírus, sendo as mutações na proteína S as mais relevantes do ponto de vista clínico e epidemiológico, devido à participação dessa proteína no início do processo de infecção viral (SOUZA *et al*, 2021).

O código genético é lido em códons (conjunto de três bases nucleotídicas). Dessa forma, é possível que a leitura desse código seja feita em três janelas distintas, a depender da base nucleotídica considerada como a base inicial na leitura. Dito isso, temos também diferentes mutações que podem afetar os códons de leitura e, consequentemente, os aminoácidos traduzidos.

As mutações genéticas podem ser benéficas, maléficas ou neutras. Elas são classificadas em: *missense*, que ocasionam a mudança de um aminoácido na composição da proteína, ou *nonsense*, quando a alteração de aminoácido produz um códon de terminação, interrompendo a síntese proteica de forma prematura. Existem ainda as mutações silenciosas, quando a mutação não altera o aminoácido na proteína produzida. Dito isso, mutações do tipo inserções ou deleções podem causar mudança nas janelas de leituras (*frames*), se os nucleotídeos inseridos ou deletados não forem múltiplos de três. Nos casos em que não há mudança da janela de leitura, a mutação é denominada *inframe*. Quando há mudança na janela de leitura, é dito que essa alteração causa um *frameshift* (DASHTI & GAMIELDIEN, 2018)

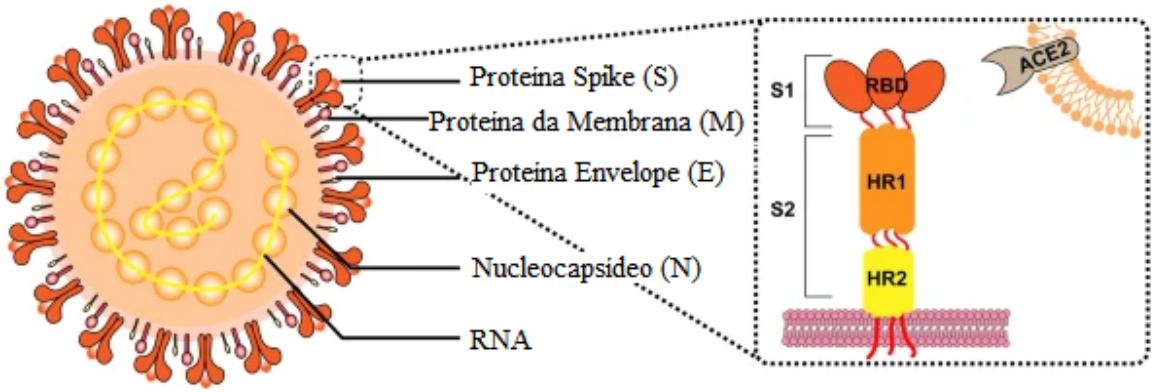
O acúmulo de mutações pode levar ao surgimento de novas linhagens. Linhagens são definidas como organismos que compartilham um ancestral comum e apresentam mutações similares. Dentro de uma mesma linhagem, é possível ter diferentes variantes, que constituem, portanto, vírus com pequenas variações genéticas (uma ou mais mutações) (SINGH & YI, 2021).

Desde o início da pandemia do SARS-CoV-2, grande esforço tem sido feito para o conhecimento e constante atualização dos genomas das linhagens circulantes em todo o

mundo. A primeira sequência do genoma completo do SARS-CoV-2 foi depositada no banco de dados GISAID em janeiro de 2020. Desde então, mutações no gene S ganharam bastante atenção, sendo alvo de constante monitoramento e estudo, devido ao seu importante papel no processo de entrada na célula hospedeira (SOUZA *et al*, 2021).

Mutações na proteína S podem causar o aumento da afinidade entre o RBD e o receptor ACE2, o que facilitaria a transmissão do vírus (Figura 2). Por essa proteína ser uma peça-chave no processo de infecção viral, ela é, muitas vezes, utilizada como alvo tanto no desenvolvimento de vacinas, como nas tentativas de obtenção de fármacos para o tratamento da COVID-19 (ZHOU *et al* 2021).

Figura 2. Estrutura do vírus SARS-CoV-2, detalhando a proteína S.



Fonte: Adaptado de ZHOU *et al* (2021).

Variantes do vírus SARS-CoV-2 que possuem as mutações N501Y, por exemplo, têm maior afinidade ao se ligar ao ACE2. Além disso, alguns resíduos de aminoácidos da proteína S são determinantes nessa interação, incluindo L455, F486, Q493, S494, N501 e Y505. Mutações que afetem esses aminoácidos tem, portanto, potencial para afetar a transmissibilidade do vírus e a carga viral nos pacientes acometidos com a doença (ZHOU *et al* 2021).

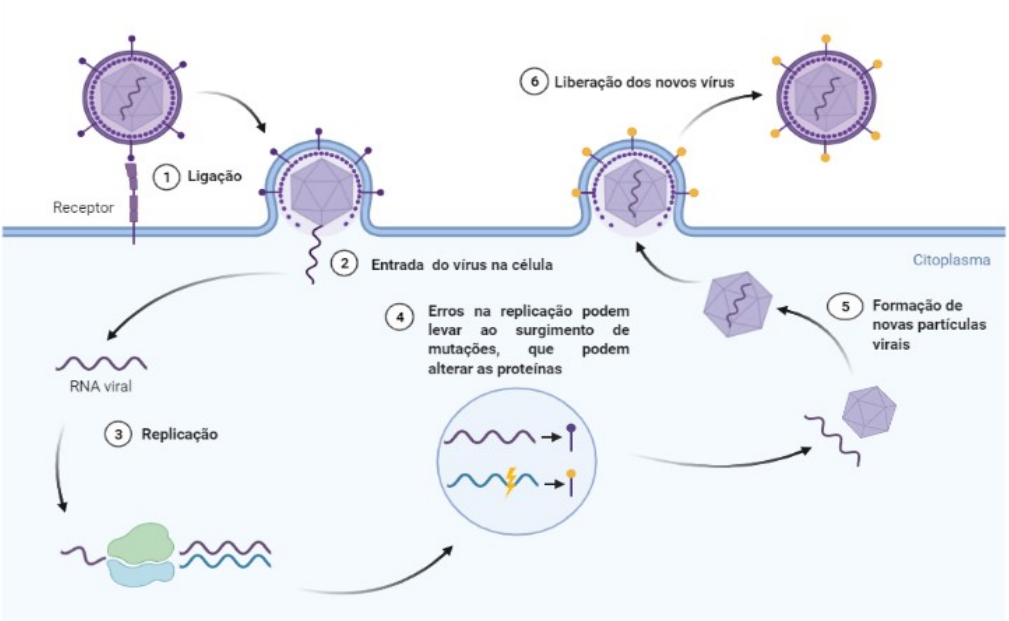
Estudos de monitoramento da variabilidade genética do SARS-CoV-2 possibilitaram a classificação dos vírus circulantes em diferentes variantes ou linhagens, com base em suas mutações e suas características epidemiológicas (HE *et al*, 2021).

1.2.3 Classificação das linhagens do SARS-CoV-2

O surgimento de novas variantes de um vírus ocorre devido ao acúmulo de mutações, causadas por erros na replicação viral dentro de uma célula hospedeira. Com o acúmulo de

algumas mutações, há o surgimento de novas variantes, com o potencial de gerarem vantagens competitivas, quando comparada às linhagens circulantes anteriormente (Figura 3). Essas vantagens podem ser: maior capacidade de dispersão, podem afetar a capacidade de performance das vacinas, ferramentas de diagnóstico, entre outras medidas preventivas (ISLAM *et al*, 2022).

Figura 3. Surgimento de mutações durante a replicação viral. Feito com BioRender.com.



Fonte: autoria própria.

Há diferentes sistema de nomenclatura para as linhagens do SARS-CoV-2: GISAID, Nextstrain e Pango. Esses sistemas são utilizados no meio científicos, mas são nomenclaturas não ideais para ambientes não científicos. Dessa forma, a OMS reuniu cientistas de vários países e agências para estabelecer uma nomenclatura de fácil entendimento e uso em ambientes não científicos, o que se concretizou na adoção de letras do alfabeto grego (WHO,2022).

Desde o início da pandemia do novo coronavírus (SARS-CoV-2), diversas novas variantes já foram descritas. Em decorrência das mutações características de cada uma dessas variantes e dos efeitos por elas causados, algumas delas foram classificadas como *variants of concern* (VOCs), *variants of interest* (VOIs) ou *variants under monitorin* (VUMs) pelo CDC (*Centers for Disease Control and Prevention*) (HE *et al*, 2021; ISLAM *et al*, 2022).

As VOCs são definidas como linhagens que, num estudo comparativo, tenham demonstrado estar associadas a uma ou mais das seguintes mudanças: (1) Aumento na transmissibilidade ou mudança prejudicial na epidemiologia do COVID-19; (2) Aumento na virulência ou alteração na apresentação clínica da doença; (3) Diminuição da eficácia das

medidas de saúde pública e sociais disponíveis ou diagnósticos, vacinas e terapias (WHO, 2022; ISLAM *et al*, 2022).

No que tange às VOCs, podemos citar as linhagens: 1) *Alpha* ou B.1.1.7; 2) *Beta* ou B.1.351; c) *Gamma* ou P.1.; d) *Delta* ou B.1.617.2 e e) *Ômicron* ou B.1.1.529. Cada uma dessas linhagens apresenta assinaturas múltiplas (deleções e mutações características), o que justifica a sua classificação como VOC e a necessidade de um monitoramento mais criterioso para essas linhagens (WHO, 2022).

A linhagem B.1.1.7 (*Alpha*) foi identificada em setembro de 2020 no Reino Unido e é detentora de 14 mutações *missense*, nas quais há alterações dos aminoácidos na proteína resultante. Dessas 14 mutações, sete estão localizadas na proteína S (FREITAS *et al*, 2021). Além dessas, há três deleções na proteína Spike, o que favorece a entrada do vírus na célula hospedeira. Foi estimado que a transmissibilidade desta variante era cerca de 43-90% maior que as demais variantes que surgiram antes dela (DAVIES *et al*, 2021).

Já a variante B.1.617.2 (*Delta*), foi descrita em outubro de 2020 na Índia, e contém diversas mutações no domínio N-terminal e no RBD da proteína S (FREITAS *et al*, 2021; PLANAS *et al*, 2021). Essa linhagem demostrou ser 60% mais transmissível do que a linhagem *Alpha* e, pouco tempo depois, deu origem a uma sublinhagem, chamada *Delta plus* (B.1.617.2.1 ou AY.1). Comparada à *Delta*, a linhagem *Delta plus* apresenta apenas uma mutação adicional na proteína S, de forma que demonstrou ter taxas de transmissibilidade semelhantes (ISLAM *et al*, 2022).

Em novembro de 2020, uma nova variante de SARS-CoV-2, a P1 (*Gamma*), foi detectada em quatro viajantes provenientes do Brasil. Essa linhagem apresentou 17 substituições de aminoácidos, sendo 10 dessas, na superfície da proteína S. As mutações mais preocupantes dessa linhagem foram: N501Y, E484K e K417T. Além dessas, duas outras mutações foram encontradas: na ORF8 e na proteína N (COG, 2021).

Em 18 de dezembro de 2020, autoridades da África do Sul anunciaram a detecção da B.1.351 (*Beta* ou 20H/501Y.V2). Três mutações receberam maior atenção, presentes na superfície da proteína S: K417N, E484K e N501Y. 25. Comparada às variantes pré-existentes, essa variante apresentou uma capacidade de disseminação maior e mostrou-se causadora de uma condição severa em pacientes jovens acometidos com o vírus (GALLOWAY *et al*, 2021)

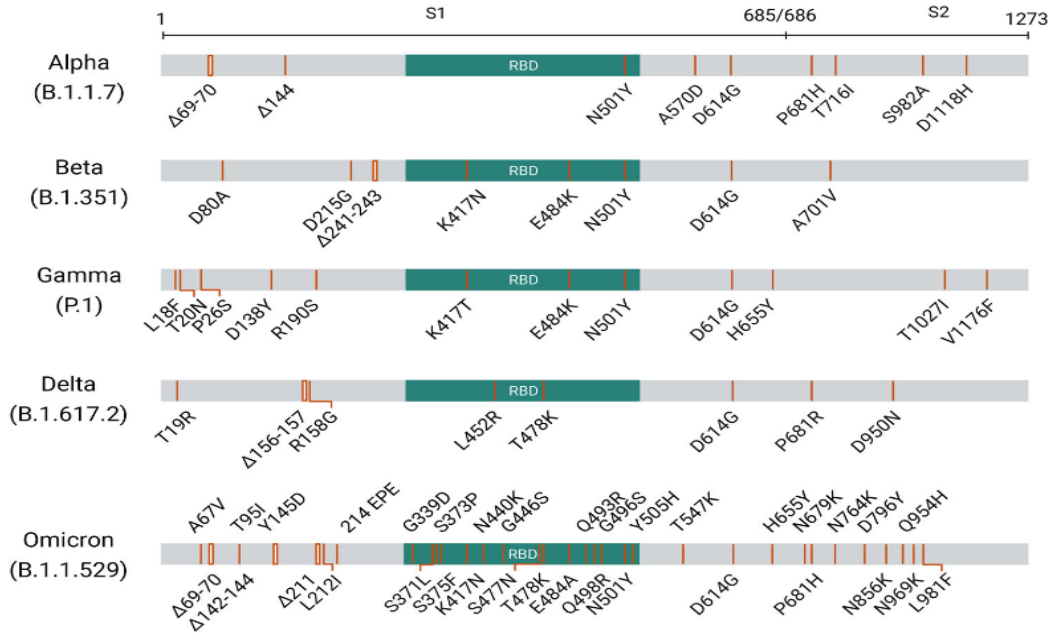
Por último, tivemos a identificação da variante *Ômicron* (B.1.1.529), em novembro de 2021, na África do Sul. A variante Ômicron tem mais de 60 substituições/deleções/inserções, o que a torna a variante com o maior número de mutações dentre todas as variantes conhecidas até o momento (MAHASE, 2022). Além disso, a *Ômicron* também compartilha

diversas mutações com outras variantes classificadas como VOCs, como *Alpha, Beta* e *Gamma*, como podemos observar na Figura 4 (HE *et al*, 2021; ISLAM *et al*, 2022).

Pouco tempo após o surgimento da *Ômicron*, surgiram suas sublinhagens: BA.1, BA.2 e BA.3. A sublinhagem BA.1, até o momento, já foi detectada em mais de 130 países e apresenta 37 mutações apenas na proteína Spike. Ela foi associada com uma menor severidade, apesar da maior transmissibilidade. A sublinhagem BA.2 apresenta menos mutações que a BA.1, totalizando 31 na proteina Spike, e ainda permanece como uma subvariante pouco conhecida, mas que tem se espalhado rapidamente no mundo. Quanto à subvariante BA.3, ela apresenta 31 mutações na proteína Spike e já foi reportada em pelo menos 19 países. (MAHASE, 2022).

Dadas todas as diferenças e especificidades inerentes a cada uma das linhagens acima citadas, é importante também atentar para as mutações genéticas em comum, ditas convergentes, compartilhadas entre as linhagens classificadas como VOCs e demonstradas na Figura 4.

Figura 4. Diagrama esquemático mostrando as mutações no gene S dentre as cinco linhagens classificadas atualmente como VOCs.



Fonte: HE *et al*, 2021.

Dentre as mutações descritas na Figura 4, três mutações localizadas na região RBD, merecem destaque: 1) Mutação N501Y, presente nas variantes *Alpha, Beta, Gamma* e *Ômicron*, responsável pelo aumento da afinidade da proteína S (Spike) e o receptor humano ACE2; 2) Mutação E484Δ, presente nas variantes *Beta, Gamma* e *Ômicron*, também

localizada na proteína S e responsável por um leve impacto positivo na ligação ao receptor ACE2. Contudo, se essas duas mutações ocorrem juntas no genoma viral, elas passam a ter um efeito sinérgico, aumentando a afinidade da ligação; 3) Mutação no sítio K417Δ da Spike (K417N ou K417T), presente nas linhagens *Beta, Gamma* e *Ômicron*, com um impacto positivo na expressão do gene S, além de uma maior proteção contra a neutralização por soros convalescentes (MARTIN *et al*, 2021). O surgimento dessas mutações de forma independente nas linhagens demonstra de forma muito interessante o efeito da pressão seletiva do ambiente *versus* a vantagem competitiva adquirida pelas linhagens detentoras da mutação.

Além das variantes acima descritas, em março de 2022, surgiu uma nova vante resultante de uma combinação da Ômicron com a Delta: a variante Deltacron. Essa variante conta com 36 alterações de aminoácidos apenas na proteína S, quando comparada à linhagenm original do SARS-CoV-2. Por conta disso, a Organização Mundial da Saúde (OMS) está monitorando sua ocorrência, mas ainda não a enquadro como VOC (MAULUD *et al*, 2022).

As VOIs compreendem variantes com mutações que são preditas ou conhecidas por afetar características do vírus tais como transmissibilidade, severidade da doença, escape imune e escape diagnóstico ou terapêutico, além de terem sido identificadas por ter uma transmissão comunitária significativa ou múltiplos clusters de COVID-19 em múltiplos países, com uma crescente prevalência relativa (WHO, 2022). São classificadas como VOIs as linhagens: *Lambda* (B.1.1.1 ou C.37) e *Mu* (B.1.621) (CHAKRABORTY *et al*, 2021).

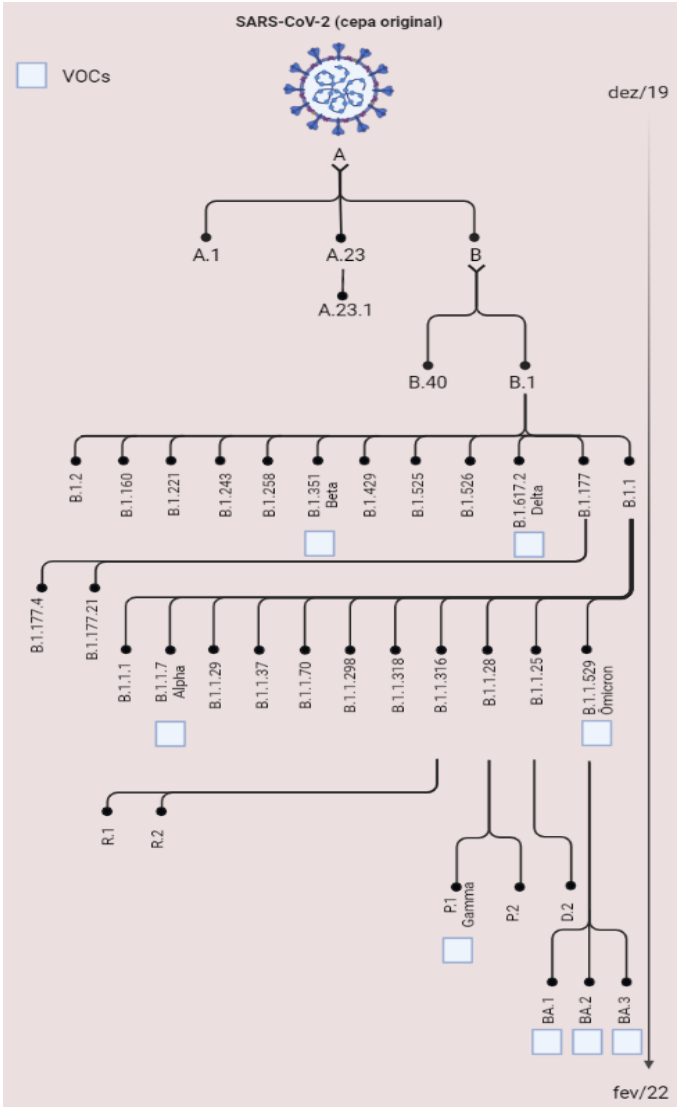
A linhagem *Lambda* foi identificada pela primeira vez no Peru, em dezembro de 2020, e apresenta uma nova deleção (S: Δ246 to 252, localizada no domínio N-terminal), além de sete mutações não-sinônimas no gene S (duas situadas na região RBD). Essa linhagem tem mutação L452Q, exclusiva dela, enquanto L452R está presente na variante *Delta*, estando associada com um aumento na afinidade com o receptor ACE2. (ROMERO *et al*, 2021)

A linhagem *Mu* foi a última linhagem classificada como VOI. Essa linhagem foi detectada na Colômbia em janeiro de 2021 e apresenta as mutações T95I e YY144-145TSN no domínio N-terminal; R346K, E484K e N501Y no domínio de ligação ao receptor e as mutações D614G P681H e D950N em outras regiões da proteína S. Em agosto de 2021, essa linhagem já havia sido detectada em 39 países (URIU *et al*, 2021).

Em outubro de 2021, o CDC criou a categoria VUMs. Este grupo reúne variantes com mutações que são suspeitas de afetar características do vírus, que podem vir a ser um risco no futuro, mas que ainda estão associadas a um impacto fenotípico e epidemiológico incerto (WHO, 2022). Atualmente, as linhagens B.1.1.318, C.1.2 e B.1.640 compõem este grupo.

O sequenciamento genético e conhecimento das linhagens circulantes permite a identificação de regiões conservadas (comuns a várias linhagens), bem como de regiões únicas em cada linhagem. Essas informações contribuem para o desenvolvimento de vacinas eficazes e de testes diagnósticos específicos, sendo essenciais no combate à pandemia (JIAO *et al*, 2021). A Figura 5 elucida os caminhos evolutivos tomados no decorrer do surgimento das linhagens acima descritas, além de outras não mencionadas.

Figura 5. Evolução das linhagens do SARS-CoV-2, destacando as cinco linhagens classificadas atualmente como VOCs. Feito com BioRender.com.



Fonte: autoria própria.

1.3 Métodos de detecção do SARS-CoV-2

Diversos métodos de diagnóstico estão disponíveis atualmente para a detecção do SARS-CoV-2. Dentre os principais métodos possíveis de serem utilizados para esse fim,

podemos citar: RT-qPCR (*Reverse Transcription-Polymerase Chain Reaction*), RT-LAMP (*Reverse Transcription Loop-Mediated Isothermal Amplification*), *Microarray,* CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats),* entre outros (SOFI *et al*, 2020; JALANDRA *et al*, 2020).

O método considerado como o "padrão ouro" é a técnica de PCR em tempo real (RT-qPCR). Essa técnica tem uma alta sensibilidade, requerendo apenas uma pequena quantidade do RNA viral, porém demanda algumas horas para que se obtenha o resultado (SOFI *et al*, 2020). Primeiro, o RNA viral é convertido a cDNA e é amplificado utilizando iniciadores específicos. A amplificação é então detectada pelo uso de sinais fluorescentes emitido por sondas, que são usados para a medição do número de cópias da sequência alvo, uma medida indireta da carga viral. Apesar de ser o método padrão ouro utilizado, o RT-qPCR, como qualquer outra técnica, está sujeito a erros técnicos que podem levar a resultados falso negativos, além de requerer um preparo de amostras e condições laboratoriais específicos (LIU *et al*, 2020).

O RT-LAMP é uma técnica alternativa, sendo um método rápido de amplificação de DNA, realizada em uma temperatura constante, e não requerendo, portanto, equipamentos específicos como o termociclador. A técnica é normalmente realizada em 30 minutos, o que representa uma grande vantagem, considerando a alta demanda de testes que temos em momentos de pico da pandemia. Essa técnica emprega 4 ou 6 iniciadores, que se ligam a diferentes regiões do DNA alvo, conferindo assim uma alta especificidade. A adição de uma etapa inicial de transcrição reversa possibilita utilizar essa técnica para a detecção do SARS-CoV-2, visto que o vírus é de RNA (HUANG *et al*, 2020).

O *Microarray* é uma técnica de alta performance utilizada na detecção de patógenos e que garante resultados rápidos. A técnica utiliza oligonucleotídeos fixados em uma fase sólida marcados com fluorescência, que são então incubados com o material genético a ser analisado. Havendo a hibridização do material genético com os oligos fixados em fase sólida, o resultado é visualizado pela emissão da fluorescência. Apesar da alta especificidade e rapidez da técnica, ela tem um alto custo, tornando inviável seu uso em larga escala (MUKHOPADHYAY *et al*, 2021; DAMIN *et al*, 2021)

A técnica CRISPR é um método bastante utilizado para edição genética, mas que foi adaptado com sucesso para a detecção do SARS-CoV-2. Inicialmente, o RNA alvo viral é amplificado, o que pode ser feito por uma PCR convencional ou RPA (*Recombinase Polymerase Amplification*, que atua em uma reação isotérmica), aumentando assim a quantidade de material genético do vírus. Na segunda etapa, a enzima Cas13a é reprogramada

com RNAs, permitindo o reconhecimento de fragmentos alvo virais. Havendo o reconhecimento (no caso de amostras positivas), a enzima Cas13a é ativada e cliva RNAs fluorescentes repórteres, agindo de forma específica. A técnica, chamada SHERLOCK (*Specific High-sensitivity Enzymatic Reporter unLOCKing*), foi desenvolvida para a detecção do zika virus e da dengue, mas foi facilmente adaptada para a detecção do SARS-CoV-2 (LIU *et al*, 2020, MUSTAFA *et al*, 2021).

Diferentemente das técnicas acima mencionadas, que detectam infecções ativas, buscando a presença viral, há também técnicas baseadas na sorologia (detecção de anticorpos), que servem como indicativos de indivíduos que já foram expostos ao vírus (JALANDRA *et al*, 2020). Contudo, como testes sorológicos não detectam a fase inicial da infecção, o Centro Europeu para Controle e Prevenção de Doenças (ECDC) aprovou seu uso apenas para finalidades epidemiológicas e de vigilância (MILLER *et al*, 2019).

1.4 Vigilância genômica do SARS-CoV-2

A vigilância genômica tem papel importante no controle de qualquer doença, especialmente quando nos referimos a uma pandemia. Dados de vigilância constribuem para estimar a proporção entre as variantes circulantes regional e nacionalmente. Essas informações, juntamente com os esforços para caracterizar o impacto clínico e na saúde pública das variantes do SARS-CoV-2, auxiliam as tomadas de decisões por parte do governo quanto às medidas de segurança (PAUL et al, 2021).

No Brasil, algumas decisões governamentais contribuiram para a rápida disseminação do SARS-CoV-2, como a longa demora na imposição de restrições de viagens aéreas e a falta de controle em viagens domésticas. Além disso, o relaxamento no distanciamento, a crescente mobilidade da população intermunicipal e o ritmo lento da vacinação também contribuíram para o surgimento e disseminação de linhagens pelo país (SANTOS *et al*, 2021). Por outro lado, medidas de controle eficazes contribuem para uma menor diversidade genética do SARS-CoV2 e para uma diminuição nas taxas de transmissão. Estudos de vigilância genômica são fundamentais para esse monitoramento, bem como para acompanhar a eficácia da vacinação na população (DENG et al, 2021; PAUL et al, 2021).

Com base no exposto, se faz necessária uma vigilância genômica constante para que possamos compreender os cursos da evolução do vírus SARS-CoV-2 e também para que possamos tomar medidas adequadas de combate à pandemia. Essa vigilância se faz através de duas frentes: diagnósticos moleculares que permitam o rápido manejo do paciente e sequenciamento contínuo de amostras virais, que permitam a identificação de novas mutações

e/ou linhagens. A identificação das mutações e/ou linhagens circulantes, somado ao desenvolvimento de vacinas (que devem ser constantemente melhoradas, considerando as contínuas mutações sofridas pelo vírus), contribuem para a tomada de decisão do poder público sobre as medidas adequadas a serem implementadas na população para o controle e combate à pandemia.

Considerando essas duas frentes (diagnóstico e sequenciamento viral), esse trabalho se propôs a: 1) trazer uma visão geral acerca do diagnóstico molecular do SARS-CoV-2, considerando os fatores que podem afetar a sensibilidade e/ou acurácia da técnica RT-qPCR; 2) realizar o sequenciamento de 34 genomas de amostras do SARS-CoV-2 provenientes do estado do Ceará, obtidas durante o período de junho/2020 a março/2020.

2. Objetivos

2.1 Objetivo geral

Realizar um levantamento dos fatores que afetam o diagnóstico do SARS-CoV-2 através da RT-qPCR e identificar as variantes do vírus SARS-CoV-2 circulantes no estado do Ceará através do sequenciamento do genoma viral.

2.2 Objetivos específicos

- Descrever as possíveis causas da obtenção de resultados falso-positivos e falso-negativos no diagnóstico molecular da COVID-19;
- Realizar o sequenciamento do genoma completo dos vírus obtidos nas amostras selecionadas;
- Classificar as amostras de acordo com as variantes já descritas mundialmente ou identificar novas;
- Realizar a chamada de variantes dos genomas.

REFERÊNCIAS

AQUINO, E.M.L.; SILVEIRA, I.H.; PESCARINI, J.M.; AQUINO, R. *et al*. Social distancing measures to control the COVID-19 pandemic: potential impacts and challenges in Brazil. **Ciência saúde coletiva**, v. 25, suppl 1, 2020.

BEAN, D. J.; SAGAR, M. Family matters for coronavirus disease and vaccines. **The Jornal of Clinical Investigation**, 2021.

BIANCHI, M.; BENVENUTO, D.; GIOVANETTI, M.; ANGELETTI, S.; CICCOZZI, M.; PASCARELLA, S. Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics? **BioMed Research International**, 2020.

CASTRO, M. C.; SINGER, B. Prioritizing COVID-19 vaccination by age. **PNAS** April 13, 118 (15) e2103700118, 2021.

COG. Covid-19 Genomic UK Consortium. COG-UK report on SARS-CoV-2 Spike mutations of interest in the UK 15th January 2021. 2021. Acessado em dezembro, 2021. https://www.cogconsortium.uk/wp-content/uploads/2021/02/Report-2_

CHAKRABORTY, C.; SHARMA, A.R.; BHATTACHARVA, M.; AGORAMOORTHY, G.; LEE, S.S. Evolution, mode of transmission and mutational landscape of newly emerging SARS-CoV-2 variants. **ASM Journals, mBio**, Vol. 12, No. 4, 2021.

CUI J, LI F, SHI Z-L. Origin and evolution of pathogenic coronaviruses. **Nat Rev Microbiol**. 17(3): 181-192, 2019.

DAMIN, F.; GALBIATI, S.; GAGLIARDI, S.; CEREDA, C.; DRAGONI, F.; FENIZIA, C.; SAVASI, V.; SOLA, L.; CHIARI, M. CovidArray: A Microarray-Based Assay with High Sensitivity for the Detection of Sars-Cov-2 in Nasopharyngeal Swabs. **Sensors**, 21, 2490, 2021.

DASHTI, M. J .S.; GAMIELDIEN, J. A practical guide to filtering and prioritizing genetic variants. **Biotechniques**, v. 62, nº 1, 2018.

DAVIES, N.G.; ABBOTT, S.; BARNARD, R.C., *et al*. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. **Science**. 2021.

DENG, X.; GU, W.; FEDERMAN, S.; DU PLESSIS, L.; PYBUS, O. G.; FARIA, N. R.; *et al* Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California, **Science** v. 369, p. 582, 2020.

FREITAS, A. R. R.; GIOVANETTI, M.; ALCANTARA, L. C. J. Variantes emergentes do SARS-CoV-2 e suas implicações na saúde coletiva. **Interamerican Journal of Medicina and Helth**, 2021.

GALLOWAY S.E.; PAUL P.; MACCANNELL, D. R., *et al*. Emergence of SARS-CoV-2 B.1.1.7 lineage – United States, December 29, 2020–January 12, 2021. **MMWR Morb Mortal Wkly Rep**. v. 70: p.95-99, 2021.

GÓMEZ-CARBALLA, A.; BELLO, X.; PARDO-SECO, J.; MARTINÓN-TORRES, F.; SALAS, A. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. **Cold Spring Harbor Laboratory Press**, 2020.

GORDON D. E.; JANG G.M.; BOUHADDOU M., et al. A SARS-CoV-2-human protein-protein interaction map reveals drug targets and potential drug-repurposing. **Nature**, 2020; V. 583: 459- 468, 2020.

HE, X.; HONG, W.; PAN, X.; LU, G.; WEI, X. SARS-CoV-2 Ômicron variant: Characteristics and prevention. **MedComm**, 2021.

HUANG, W. E.; LIM, B.; CHIA-CHEN, H.; XIONG, D.; WU, W. *et al*. RT-LAMP for rapid diagnosis of coronavirus SARS-CoV-2. **Microbial Biotechnology**, v. 13 (4), p. 950– 961, 2020.

ISLAM, S.; ISLAM, T.; ISLAM, R. New Coronavirus Variants are Creating More Challenges to Global Healthcare System: A Brief Report on the Current Knowledge. **Clinical Pathology** Volume 15: 1–7, 2022.

JALANDRA, R.; YADAV, A. L.; VERMA, D.; DALAL, N.; SHARMA, M.; SINGH, R.; KUMAR, A.; SOLANKI, P. R. Strategies and perspectives to develop SARS-CoV-2 detection methods and diagnostics, **Biomedicine & Pharmacotherapy**, Volume 129, 2020.

JIAO, D.; DONG, X.; YU, Y.; WEI, C. Gene Presence/Absence Variation analysis of coronavirus family displays its pan-genomic diversity. **International Journal of Biological Sciences.** v. 17(14), p.3717–37272021, 2021.

KRAMER, F. SARS-CoV-2 vaccines in development. **Nature**, v. 586, p. 516-527, 2020.

LIU, R, FU, A, DENG, Z, LI, Y, LIU, T. Promising methods for detection of novel coronavirus SARS-CoV-2. **VIEW**, 1:e4.; 2020.

MAHASE, E. COVID-19: What we know about omicron sublineages? : **BMJ** 376:o358, 2022.

MALIK, Y.A. Properties of Coronavirus and SARS-CoV-2. **Malaysian J Pathol** v. 42(1): p. 3 – 11, 2020.

MARTIN, D. P.; WEAVER, S.; TEGALLY, H.; SAN, J.E.; SHANK, S. D.; WILKINSON, E.; […] The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. **Cell**, Volume 184, Issue 20, Pages 5189-5200.e7, 2021.

MAULUD, S.Q.; HASAN, D. A.; ALI, R. K., et al. Deltacron: Apprehending a new phase of the COVID-19 pandemic. **International Journal of Surgery** (London, England). May;102:106654, 2022 .

MILLER, S.; CHIU, C.Y.; RODINO, K.G.; MILLER, M.B. Point-Counterpoint: Should We Be Performing Metagenomic Next-Generation Sequencing for Infectious Disease Diagnosis in the Clinical Laboratory? **J. Clin. Clin. Microbiol.** 58, 2019.

MUKHOPADHYAY, C.; SHARMA, P.; SINHA, K.; RAJARSHI, K. Recent trends in analytical and digital techniques for the detection of the SARS-Cov-2, **Biophysical Chemistry**, Volume 270, 2021.

MUSTAFA, M. I.; MAKHAWI, A. M. SHERLOCK and DETECTR: CRISPR-Cas Systems as Potential Rapid Diagnostic Tools for Emerging Infectious Diseases. **Journal of Clinical Microbiology**, 2021.

PARK, S. H.; SIDDIQI, H.; CASTRO, D.V.; DE ANGELIS, A.A.; OOM, A.L.; STONEHAM, C. A.; et al. (2021) Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. **PLoS Pathog** 17(5): e1009519.

PAUL, P., FRANCE, A. M., AOKI, Y., BATRA, D., BIGGERSTAFF, M., DUGAN, V., GALLOWAY, S., HALL, A. J., JOHANSSON, M. A., KONDOR, R. J., HALPIN, A. L., LEE, B., LEE, J. S., LIMBAGO, B., MACNEIL, A., MACCANNELL, D., PADEN, C. R., QUEEN, K., REESE, H. E., RETCHLESS, A. C., … SILK, B. J. Genomic Surveillance for SARS-CoV-2 Variants Circulating in the United States, December 2020-May 2021. **MMWR. Morbidity and mortality weekly report**, 70 (23), 846–850, 2021.

PLANAS, D., VEYER, D., BAIDALIUK, A. *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. **Nature**, v. 596, p.276–280, 2021.

ROMERO, P.E.; DÁVILA-BARCLAY, A.; SALVATIERRA, G. *et al*. The Emergence of Sars-CoV-2 Variant Lambda (C.37) in South America. **Genomics and Proteomics**, 2021.

SABIONI, L.; LORENZO, A.; LAMAS, C.; MUCCILLO, F.; CASTRO-FARIA-NETO, H. C.; ESTATO, V.; TIBIRICA, E. Systemic microvascular endothelial dysfunction and disease severity in COVID-19 patients: evaluation by laser Doppler perfusion monitoring and cytokine/chemokine analysis. **Microvasc. Res.**, v.134, 2021.

SANTOS, C.A.; BEZERRA, G.V.B.; DE AZEVEDO MARINHO, A.R.R.A.; ALVES, J.C.; TANAJURA, D.M.; MARTINS-FILHO, P.R. SARS-CoV-2 genomic surveillance in Northeast Brazil: timing of emergence of the Brazilian variant of concern P1. **Journal of Travel Medicine**, V. *28*, pp.1–3, 2021.

SHUVAN, B. ; SOHAN, S.; RIJU, D.; KOUSIK, K.; PRITHA, B. Mutational spectra of SARS-CoV-2 orf1ab polyprotein and signature mutations in the United States of America. **Journal of Medical Virology**, V. 93, p. 1428-1435, 2020.

SINGH, D.; YI, S. On the origin and Evolution of SARS-CoV-2. **Experimental & Molecular** Medicine, v. 53, pp. 537-547, 2021

SIQUEIRA, P. C.; COLA, J. P.; COMERIO, T.; SALES, C. M. M.; MACIEL, E. L. Limiar de imunidade de rebanho para SARS-CoV-2 e efetividade da vacinação no Brasil. **J Bras Pneumol**, v. 48(2) : e20210401, 2022.

SOFI, M. S.; HAMID, A.; BHAT, S. U. SARS-CoV-2: A critical review of its history, pathogenesis, transmission, diagnosis and treatment. **Biosafety and Health**, v. 2, Issue 4, pp. 217-225, 2020.

SOROKINA, M., M. C. TEIXEIRA, J., BARRERA-VILARMAU, S. *et al.* Structural models of human ACE2 variants with SARS-CoV-2 Spike protein for structure-based drug design. **Sci Data** 7, 309, 2020.

SOUZA, P.F.N.; MESQUITA, F. P.; AMARAL, J. L.; LANDIM, P. G. C.; LIMA, K. R. P.; COSTA, M. B.; FARIAS, I. R.; LIMA, L. B.; MONTENEGRO, R. C. The human pandemic coronaviruses on the show: The spike glycoprotein as the main actor in the coronaviruses play. **International Journal of Biological Macromolecules**, Volume 179, p. 1-19, 2021.

URIU, K.; KIMURA, I.; SHIRAKAWA, K.; TAKAORI-KONDO, A.; NAKADA, T.; KANEDA, A.; NAKAGAWA, S.; SATO, K. Neutralization of the SARS-CoV-2 Mu Variant by Convalescent and Vaccine Serum. **New England Journal of Medicine**, 2021.

VALE, F. F.; VÍTOR, J. M. B.; MARQUES, A. T.; AZEVEDO-PEREIRA, J. M.; ANES, E.; GONCALVES, J. Origin, phylogeny, variability and epitope conservation of SARS-CoV-2 worldwide. **Virus Res**, 304, Article 198526, 2021.

WALLS, A. C.; PARK, Y.J.; TORTORICI, M.A.; WALL, A.; MCGUIRE, A. T.; VEESLER, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein **Cell**, 181, pp. 281-292.e6, 2020.

WHO (The World Health Organization). Fonte: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants; Acesso em: 17 de fevereiro de 2022.

XU, Z.; SHI, L.; WANG, Y.; ZHANG, J.; HUANG, L.; LIU, S.; et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. **Lancet Respir. Med.**, p. 8, pp. 420-422, 2020.

YANG, J.; CHEN, X.; DENG, X.; CHEN, Z.; GONG, H. YAN, H.; WU, Q.; SHI, H.; LAI, S.; AJELLI, M.; VIBOUD, C.; YU, H. Disease burden and clinical severity of the first pandemic wave of COVID-19 in Wuhan, China. **Nature Communications**, volume 11, Article number: 5411, 2020.

ZHOU, W., WANG, W. Fast-spreading SARS-CoV-2 variants: challenges to and new design strategies of COVID-19 vaccines. **Signal Transduction and Targeted Therapy** 6, 226, 2021.

CAPÍTULO II

True or False: What are the factors that influence COVID-19 diagnosis by RT-qPCR?

True or False: What are the factors that influence COVID-19 diagnosis by RT-qPCR?

Luina Benevides Lima [1‡], Felipe Pantoja Mesquita[1‡], Pedro F. N. Souza[2], Lais Lacerda Brasil de Oliveira[1], Francisca Andréa da Silva Oliveira[1], Maria Elisabete Amaral de Moraes[1], Raquel Carvalho Montenegro[1*].

[1]Laboratory of Pharmacogenetics, Drug Research and Development Center (NPDM), Federal University of Ceará, Fortaleza, CE 60430-275, Brazil
[2]Department of Biochemistry and Molecular Biology (DBBM), Federal University of Ceará, Fortaleza, CE 60430-275, Brazil

[‡]These authors contributed equally.
*Corresponding author.
E-mail address: rmontenegro@ufc.br (Raquel Carvalho Montenegro)

Artigo publicado na revista *Expert Review of Molecular Diagnostics*, 2022.

https://doi.org/10.1080/14737159.2022.2037425

Abstract

Introduction

The Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) disease has had a catastrophe impact on the world resulting in several deaths. Since World Health Organization declare the pandemic status of the disease, several molecular diagnostic kits have been developed to help the tracking of viruses spread.

Areas Covered

This review aims to describe and evaluate the currently reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) diagnosis kit. Several processes used in COVID-19 diagnostic procedures are detailed in further depth to demonstrate optimal practices. Furthermore, sampling methods, primer-probe dimerization occurrence and other influencing factors were discussed. Therefore, we debate the main factors that influence the viral detection of SARS-COV-2 and how they can affect the diagnosis of patients.

Expert Opinion

Here is highlighted and discussed several factors can interfere in the RT-QPCR analysis, such as the viral load of the sample, collection site, collection methodology, sample storage, transport, and different brand kits. This is a pioneer study to discuss the factor that could lead to wrong interpretation of RT-QPCR diagnosis of SARS-CoV-2. Moreover, the presence of mismatch between primer and sample target, as well as the primer dimerization might be a key factor for the sensitivity of real-time PCR. The failure in any one of these points can lead to obtaining false-positive or false-negative results. This study aimed to help the readers to understand what very likely is behind a bad result of SARS-CoV-2 detection by RT-QPCR and what could be done to reach a reliable diagnosis.

Keywords: Coronavirus, Diagnosis, COVID-19, RT-QPCR, Sensitivity.

1 Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) belongs to the *Coronaviridae* family and contains a positive sense nonsegmental single-stranded RNA with around 30 Kb length. In December 2019, it was discovered that infections by SARS-CoV-2 led to coronavirus disease, later named COVID-19 [1]. COVID-19 has spread rapidly worldwide and became pandemic in March 2020 [2]. According to WHO, until august 2021, approximately 207,784,507 cases of COVID-19 were reported worldwide with 4,370,424 deaths.

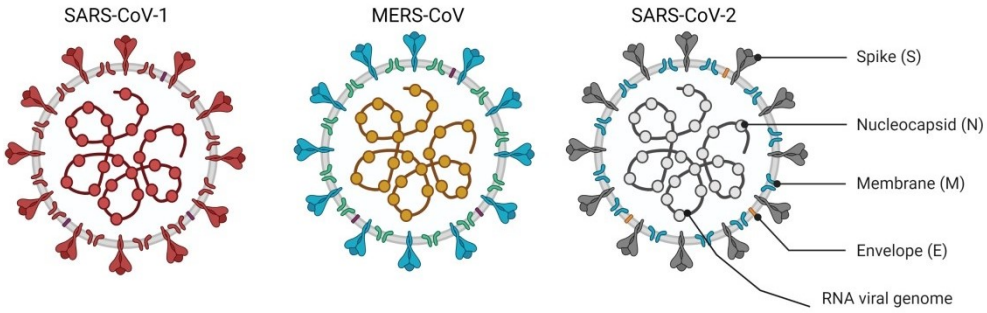
The gold standard for SARS-CoV-2 infection diagnosis is Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR), also named real-time PCR, which is based on the use of probes and primers to specifically amplify a targeted region of viral genetic material [3]. Despite the high specificity and sensitivity of RT-qPCR, the method may be affected by several factors, such as the diagnostic kit used, viral loads, site of collection, and time of infection [4]. Indeed, false results from RT-qPCR analyses were reported from Wuhan hospitals which several factors can influence during the sample collection and processing [5,6].

Thus, the objective of this work was to list all factors that can contribute to false-negative or false-positive results in the RT-qPCR test for SARS-COV-2. However, before going more in-depth in the discussion, a brief background on coronaviruses and RT-qPCR is required.

2 Coronaviruses

The Coronaviridae family is divided into four main genera: alpha, beta, gamma, and delta-CoVs [7,8]. Before the ongoing SARS-CoV-2 outbreak in December 2019, SARS-CoV (2002-2003) and MERS-CoV (2012) smaller outbreaks lead to severe respiratory illnesses [7,9,10]. SARS-CoV and MERS-CoV reached several countries, infecting and killing thousands of individuals by causing respiratory and neurological diseases with a high case fatality rate [9]. Coronaviruses share similar structures with genomes enclosed in a lipid envelope (Figure 1). Regarding the proteome, they have a nucleocapsid protein (N) physically attached to RNA. Additionally, the viral gene encodes a small membrane envelope protein (E), spike protein (S), and a membrane protein (M) (Fig. 1) [8].

Figure 1. Morphology of the three relevant coronaviruses with epidemic/pandemic impact on human health. SARS-CoV-1, MERS-CoV and SARS-CoV-2. The structural proteins are shown in the figure, such as spike (S), membrane (M), nucleocapsid (N) and envelope (E). Their genes are commonly used as target for real time PCR detection. Created in BioRender.com.



Among those proteins, given its importance to viruses entering the cell, spike protein has become a target for drug development and antibody neutralization. However, spike protein is the most variable protein in coronaviruses making it a bad target for diagnosis [11,12]. Despite S protein that present high mutational rate, the other structural proteins in which are more conserved among the human coronaviruses including SARS-CoV, MERS-CoV, and SARS-CoV-2 were used to develop or adapt the existing diagnosis technologies for be able to detect COVID-19.

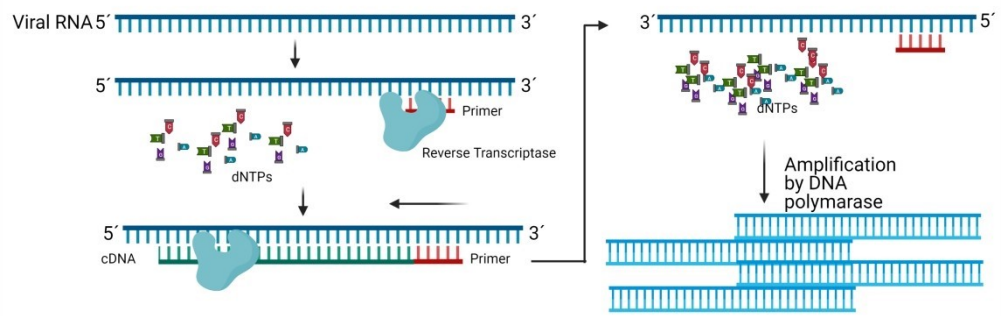
3 RT-qPCR for SARS-CoV-2 detection

RT-qPCR is the most employed technique to identify the presence or absence of SARS-CoV-2, including for early diagnosis of COVID-19 disease (Fig. 2) [13–16]. SARS-CoV-2 detection is similar to that employed for other acute respiratory infections caused by viruses [14,16]. The sample collection for diagnostics can be performed from several points from human body, such as nasopharyngeal and oropharyngeal swabs, human fluids such as blood, blood serum, saliva, urine, and anal (Fig. 3) [17,18]. Furthermore, SARS-CoV-2 could be found in peripheral blood specimens, although variable results have been reported [19].

Overall, PCR reactions are applied to samples composed of DNA, allowing direct amplification by Taq polymerase activity and detection by the machine. However, to detect RNA viruses, like SARS-CoV-2, the process is a bit different (Fig. 2). In this context, a previous step for viral mRNA conversion to DNA is required. Then, the RT-qPCR detection for RNA viruses occurs in two steps: 1) a reverse transcription reaction to produced complementary DNA (cDNA) using copies of mRNA as primer catalyzed by an RNA-

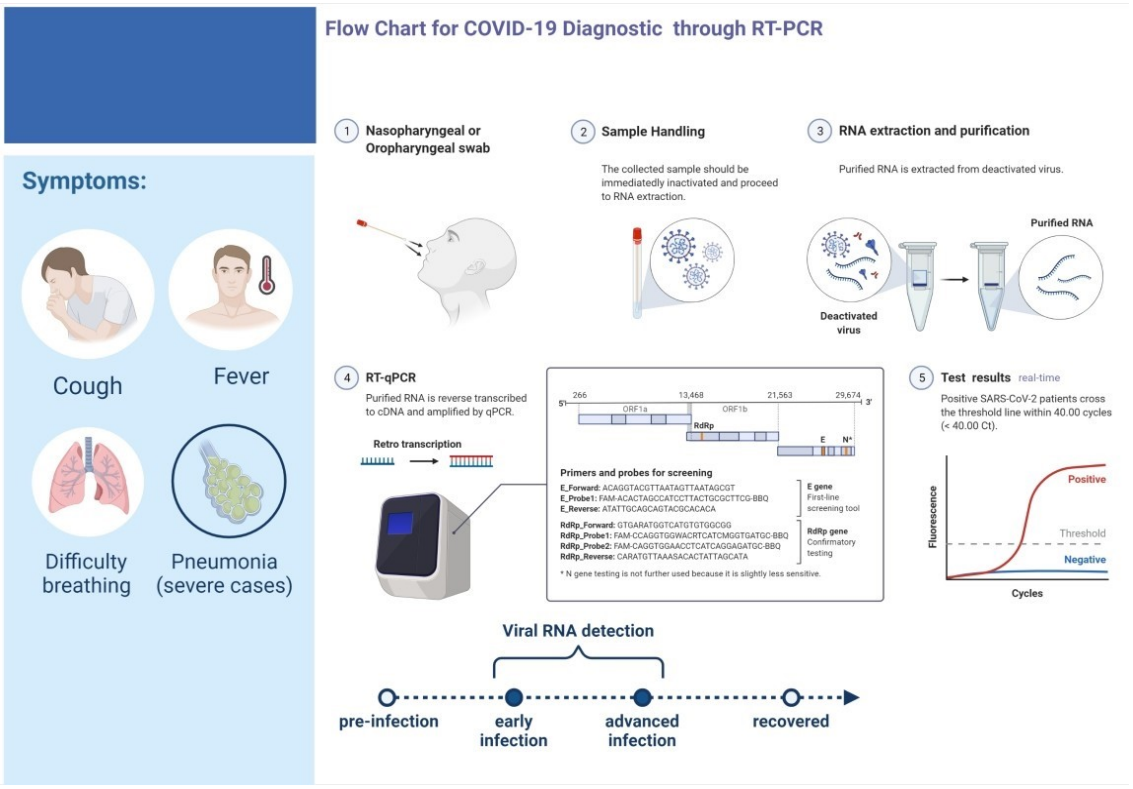
dependent DNA polymerase (reverse-transcriptase) Taq polymerase is applied to amplify the specific segment of genome which provide result about virus presence (Fig. 2) [20]. Most RT-qPCR tests for SARS-CoV-2 are quantitative by using fluorescence measurements that are sometimes referred to as RT-qPCR. Briefly, cDNA hybridizes with a probe targeted with both fluorescent and quencher labels. After polymerization into double-stranded DNA (ds-DNA), the quencher and fluorescent probes are separated and light emission from the fluorophore is observed upon light excitation [20].

Figure 2. Scheme showing gene amplification in RT-qPCR process. In the RT-qPCR the first step is the construction of complementary DNA (cDNA) using RNA as model by reverse transcriptase. Therealter, the stable double-strand DNA is used as template for the exponentially amplification of the product. Created in BioRender.com.



SARS-CoV-2 detection by RT-qPCR is quite simple, as summarized in Fig. 3. Patients are eligible to be tested once they present symptoms. The most common symptoms are cough, dyspnea, chest pain, myalgia/arthralgia, diarrhea, nausea, vomiting, and common systemic symptoms observed: fever, chills, and fatigue [21]. First, the health professional performs the sample collection from patients primarily by nasopharyngeal or oropharyngeal swab method. Second, after the collection, sample handling is involved in virus transport, which should occur at a controlled temperature (2-8 °C) to a specialized laboratory, followed by virus inactivation which could be by physical (heat and exposure to UV light) or chemical (chlorinated disinfectants). The third step is regarding the RNA extraction and purification, and then the fourth step is the RT-qPCR itself. At this point, the RNA purified is first converted cDNA, and the amplification starts (Figure 2 and 3) [22].

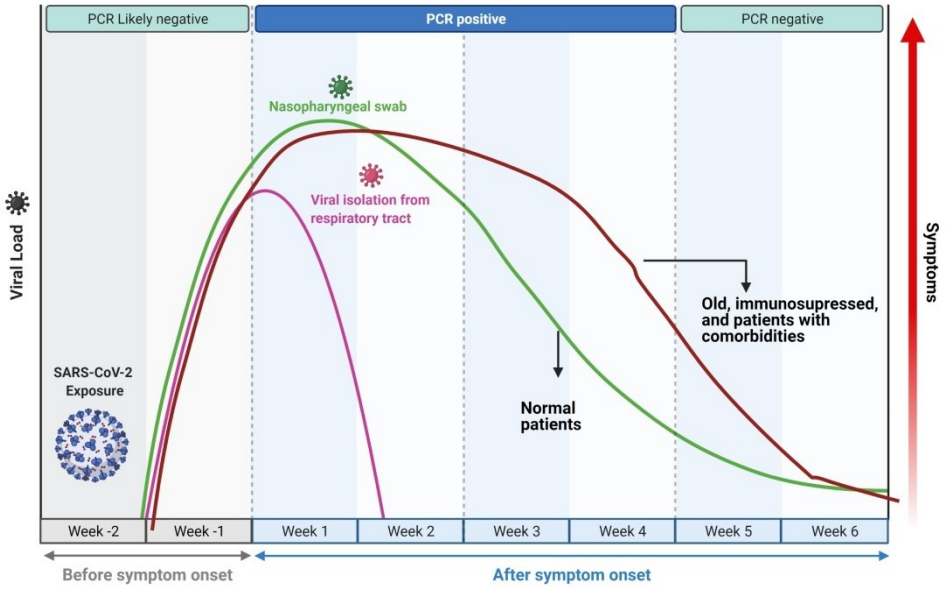
Figure 3. Flow chart for COVID-19 diagnostic by RT-qPCR. Patients presenting symptoms of COVID-19 are subjected to the test. (1) occurs the sample collection by either naso- or oropharyngeal. (2) The collected sample is immediately processed. (3) RNA extraction. (4) RT-qPCR process and (5) Data analysis. Created in BioRender.com.



For the amplification process, the common targets employed are E, ORF1ab, and N genes, which are considered stable genes. In this step, two primers and one probe are used for each gene. The probe is involved in fluorescence releasing, used for reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) detection (Fig. 3). The fifth is probably the most crucial step in data analysis. At this point, the data analysis will reveal if the patient is positive or negative for SARS-CoV-2. Many factors could affect this process, including poor sample collection and handling, RNA extraction, and RT-qPCR runs [13,20].

Another important point that could affect the results is the time of sample collection (Figure 4).

Figure 4. Flow chart suggesting the best time to perform COVID-19 diagnostics. After exposition to SARS-CoV-2, symptomatic stage is going until 2 weeks after contact, which is the best time to SARS-CoV-2 detection. In normal patients and compromised patients, low viral loads could still be detected in 3-4 weeks. Created in BioRender.com.



If the sample is collected too early, in a time called pre-infection, the results could be a false-negative because the amount of the virus is too low and is not detected yet. At this point, the virus is still in the replication process. Likewise, if the collection is made too late, the patient could present a negative result because now, it is in the recovery process, where the body already eliminates the virus. The optimal time for collection comprehends three days after the first symptoms come up until the fourteen-day [23]. This timeline provides a more reliable true negative or positive result. In this manuscript we discussed the factors that may interfere in SARS-CoV-2 detection by RT-qPCR tests.

4 Influencing Factors on SARS-CoV-2 Detection

4.1 Disease staging

Several data reported the positive correlation between viral loads and disease staging [4]. Thus, if the patient decides to perform an RT-qPCR test to diagnose COVID-19 on the first day of symptoms, the result probably would be a false negative because the viral load is still extremely low [24]. In contrast, in normal conditions and in patients with no comorbidities, if the test is taken at 13-14 days of the first symptoms, it can also lead to a false negative result (Fig. 4). However, in old or immunosuppressed patients with

comorbidities, the infection and positive RT-qPCR results could still be positive even after 3-5 weeks later [25].

Here, we reinforce the diagnostic window importance on SARS-CoV-2 detection, to ensure a correct diagnostic. Overall, 5 to 6 days after symptoms come up, high loads of SARS-CoV-2 in their upper and lower respiratory tracts are detected (Fig. 4) [26–28].

4.2 Sample collection methodology and sample storage

According to WHO (2021), laboratory tests in COVID-19-suspected patients should be performed using samples collected from the upper respiratory tract, such as nasopharyngeal and oropharyngeal swab (Fig. 3) as well as lower respiratory specimens [29]. However, depending on the sample collection site, different viral loads are recovered [4]. Also, the effect of specimen collection time on the detection rate of novel coronavirus is important for the diagnosis success. Liu *et al*. (2020) demonstrated the nasopharyngeal swab detection rate, nasal swab, and oropharyngeal swab are higher before washing in the morning and lower after washing morning and during the afternoon. The study suggests that this is probably due to the human body resting during the night, increasing virus propagation. During the day, the activity state of the human body might affect the virus accumulation [30].

Some collection sites provide different features. For instance, saliva samples have been reported as a low cost-effective and non-invasive alternative, since it has been used to detect other respiratory viruses [31]. The collection made by swabs taken from the throat immediately upon symptom onset is 6.4% less effective than nasal swabs to yield positive results in nasal swabs [4].

A nasopharyngeal and oropharyngeal swab are often recommended for screening or diagnosis and provide a great sensitivity to early infection [16,28,32]. A single nasopharyngeal swab has become preferred because it is well accepted by the patient and is safer for the operator. Nasopharyngeal swabs have an inherent quality control reached by the correct area to be tested in the nasal cavity. W. Wang *et al*. (2020b) have just reported that oropharyngeal swabs in China are frequently employed ($n = 398$) than nasal swabs ($n = 8$) to diagnose COVID-19 outbreak; however, the SARS-CoV-2 was only detected in 32% of oropharyngeal swabs. That result is 50% lower than those for nasal swabs (63%) [33].

The collection by both nasopharyngeal and oropharyngeal swabs, either as an independent sample or a single aliquot of viral transport medium, is an attractive option in normal circumstances. As such, institutions should consider the potential outbreak effect on national/international supply chains. Despite that, there are not only nasopharyngeal and

oropharyngeal swabs to collect and diagnose. It was reported that sputum specimens or bronchoalveolar lavage fluid specimens have a remarkably high detection rate for novel coronavirus [34]. However, it is not possible to perform these kinds of collection samples in some patients. Other sample collection sites such as anal swabs, serum, stool, urine, feces, or ocular secretions have also shown to be viable sources of SARS-COV-2 detection [35,36]. Although the sample size has a high influence on the viral load recovered, it directly influences test sensitivity. For example, according to Mattioli *et al* (2020), SARS-CoV-2 could be found in 78% of serum but only in 50% of plasma samples. Therefore, it is clear that the collection site is determinant to viral load measurements.

The standardization of the collection method is one of the most important steps in diagnosing infectious diseases. Pondaven-Letourmy *et al* (2020) described two techniques for evolving the nasopharyngeal region: nasopharyngeal swab or nasopharyngeal wash/aspiration. The chosen technique could influence the viral load recovered, as well as the sample collection correct execution. Well-trained testing teams should also help increase the sensitivity of the test, which would avoid false-negative results [37].

On 8th May 2020, the salivary test was approved by the Food and Drug Administration, but we also found that it can be done in different ways: collecting through the spit, expectorated saliva, saliva collected directly from the salivary gland. Auto sampling methodology, where the patient collects himself, usually is not standardized, which can generate more bias for studies [38,39].

Those different methods may also interfere with the viral load recovered from the patient and, subsequently, on its detection accuracy (Table 1). Indeed, there is not a consensus yet if saliva samples are accurate as a nasopharyngeal sample for SARS-CoV-2 detection. Landry *et al* (2020) have shown that saliva samples are more sensitive than nasopharyngeal samples. Wyllie *et al* (2020) revealed that saliva samples allowed higher detection of SARS-CoV-2 RNA copies than samples collected by nasopharyngeal swab from the same patients at the same time. Interestingly, the authors discuss that saliva samples presented more positive results than nasopharyngeal samples up to 10 after COVID-19. 1 to 5 days after COVID-19 diagnosis, from 70 patients, 81% of the saliva samples were positive compared to 71% positive for nasopharyngeal samples [40].

Table 1. Differences in sample collection sites for SARS-CoV-2 detection.

Sample collection site	Characteristics (advantages and disadvantages)	Reference
Nasopharyngeal swab	Great sensitivity to early infection. Gold standard on SARS-CoV-2 detection. Although, this method depend on the collector expertise.	29, 37
Oropharyngeal swab	Non-invasive alternative. Lower detection rate when compared to nasopharyngeal samples	33
Sputum specimens / bronchoalveolar lavage fluid	High detection rate. Although, it is not possible to collect from this site in some patients	29, 34
Saliva	Low cost-effective, non-invasive alternative, and not dependent on the expertise of the collector.	31, 38, 39, 40
Serum, plasma	Low detection rate. These sample sites have a low detection rate and are not recommended.	6
Anal swabs, urine, feces, ocular secretions or semen	These sample sites have a low detection rate and are not recommended.	29, 33, 35, 36

This mentioned study with 13 health care workers who are completely asymptomatic was carried out by testing samples from saliva and nasopharyngeal swabs. All 13 have SARS-CoV-2 detected on saliva samples. When tested by nasopharyngeal samples, only 6 had SARS-CoV-2 detected. In conclusion, nasopharyngeal samples provide more false negative results when compared to saliva samples. This could be explained by the variation in nasopharyngeal sampling. It is known nasopharyngeal sampling is harder than saliva and this could be responsible for the variations and false results provided by nasopharyngeal samples.

In contrast, saliva sampling is too easy and can be done by the patient providing solid results [40].

WHO (2021) has recommended that after collection, the samples must be stored at 2-8 °C for no longer than 72 hours. Samples with a delay in testing or shipping must be stored at -70 °C or below. Yet, due to the high demand for tests in labs during the most critical periods of the SARS-CoV-2 outbreak, there was a delay in the deadline to process samples and release results. That scenario required sample storage for a time higher than 72 hours at 2-8 °C. A failure in storage can lead to RNA degradation contributing to a false-negative test [41]. Furthermore, if sample transportation is required, the material collected should be transported in an ideal transport medium under cold conditions in a triple-layered packaging which consists of a leak-proof receptacle [42].

Until now, none of the variants has posed any difficult on the identification by either nasopharyngeal or salivary samples. However, the newest variant Ômicron require some adjustments during collection to provide a reliable diagnostic. In case of Ômicron, the best site of collection is by saliva swabs than nasal swabs. Marais *et al*. (2021) revealed that saliva samples present a positivity of 100% (95% CI: 90-100%). In contrast, saliva sample presented a positivity of 86% (95% CI: 71-94%) to Ômicron. This fact is explained because Ômicron is more adapted to accumulate in saliva than in nasopharyngeal samples Marais *et al*. (2021).

4.3 RNA extraction methodologies

Several RNA extraction strategies and kits have emerged or been adapted as alternatives to accelerate the sample processing step and increase the viral load recovered after extraction (Table 2). A study showed the combination of heat treatment and proteinase K to improve the RNA yield after extraction and SARS-CoV-2 detection by RT-qPCR (Table 2) [43]. The authors found that proteinase K and heat pre-treatment combination led to a higher yield of RNA collected and the obtention of lower cycles quantification in RT-qPCR reaction for SARS-CoV-2 compared with heat only or no pre-treatment.

Other studies performed a comparative analysis of different RNA extraction methods, including Qiamp DSP Virus Spin Kit (Qiagen, Hilden, Germany, Cat.61704), Total RNA Purification Kit (Norgen, Rome, Italy, Cat. # 17200), Viral Nucleic Acid (DNA/RNA) Extraction Kit I (Fisher Molecular Biology, Rome, Italy, Cat. DR-003), BSA-based method, acid pH-based-method, High temperature-based method, TRIzol (Ambion da Life Technologies®) and EXTRAzol (BLIRT S.A., Gdańsk, Poland, Cat. EM30-100) [44,45]. The results were pertinent in demonstrating the RNA isolation efficiency by EXTRAzol was lower

than that extracted by column-based methods and this low yield is likely to affect the RT-qPCR performance. Moreover, the acid- pH-based method was considered an excellent alternative to commercial systems.

Table 2. RNA extraction methods.

RNA extraction method/ extra steps	Characteristics	Reference
EXTRAzol	less efficient when compared to column methods	44, 45
column-based methods	Great efficiency	44, 45
magnetic bead extraction	Great efficiency	46
proteinase K and heat pre-treatment combination + any extraction method	Promotes a higher yield of RNA collected	43

Other extraction upgrades or methods are constantly mentioned in literature to improve the test sensitivity. Klein *et al* (2020) provide an alternative method for RNA extraction based on magnetic bead extraction [46]. Another study proposed workable COVID-19 testing which might be implemented by sampling directly into a lysis buffer and RT-qPCR master mix without intermediate steps such as extraction processing [47].

4.4 SARS-CoV-2 Detection by RT-qPCR

When it comes to performing the maximum SARS-CoV-2 detection tests in a minimum period of time, there are some methods we can rapidly think of, like reverse transcription loop-mediated isothermal amplification (RT-LAMP) and RT-qPCR [46]. Although there are other SARS-CoV-2 detection protocols, the real-time qPCR is the gold-standard method recommended by the WHO [48]. Nowadays, the challenge is the detection specificity and sensitivity, which are variable and occasionally low [33]. Collection site, period of collection, sample conservation and transportation, low patient viral load, sporadic shedding, and variation in detection kits from different manufacturers contribute to the low sensitivity of SARS-CoV-2 detection [49].

In RT-qPCR, primers/probes act as biorecognition elements for different target genes, such as ORF1ab, N, and E gene [6]. Different diagnostic kits may use different SARS-CoV-2 targets to identify the virus presence/absence on the sample. Different laboratories around the world have developed several modifications of these assays.

CDC designed FDA EUA 2019-nCoV CDC kit (IDT, USA) searches for N1 and N2 targets, two regions on virus nucleocapsid gene (N) [50]. In contrast, the kit developed by Fiocruz (SARS-Cov2 (E) - Bio-Manguinhos (according to the Berlin protocol) searches for one region on the E gene [14,51]. Some other protocols also may use the RNA polymerase gene (RdRp/Helicase) or even the Spike gene (S) [14].

All possible chosen targets are susceptible to nucleotide substitution in SARS-CoV-2, affecting the oligonucleotide hybridization efficiency if mutation occurs in the primer on probes annealing regions. Yet, it is known that some regions of the SARS-CoV-2 genome are more likely to undergo mutations than others [52,53].

As observed by several studies which critically compared the efficiency and sensitivity of widely used RT-qPCR kits, the primer-probe set, and variability of SARS-CoV-2 genome have a clear participation on the reaction limit of detection. One of the key factors for detection sensitivity is the primer/probe efficiency in binding target [54]. Therefore, the accumulated mutations in SARS-CoV-2 genome during its pandemic outbreak, if it occurs in the primer region of the target DNA, implies that mismatches may affect the detection of the target [55].

Table 3 shows the results from a few studies which demonstrated the mismatch frequency of four RT-qPCR kits for different targets that had occurred in a specific number of analyzed SARS-CoV-2 genomes. The most problematic mismatch is on China CDC kit (targeting N gene), with a frequency between 12.7% - 85.3% [56,57], and Charité (targeting ORF1b) with a frequency reaching 100% [58]. As explained by Corman and Drosten (2020), a plausible reason for mismatch presence observed in some detection kits was the incomplete genomic information available at the point of designing [59].

According to Corman *et al* (2020) data, PCR assays using the N gene were slightly less sensitive than assays using E and RdRp genes. Indeed, ORF1ab and N genes were not recommended for RT-qPCR testing by institutions worldwide [60] and the S gene is more susceptible to mutations, which could affect diagnosis. Ramírez *et al* (2021) affirmed that sensibility on detection lineage B.1.1.7 could be affected if PCR kit is directed to the Spike (S) gene [61]. Considering that mutations in the S gene are present in different lineages, this is not a good target for diagnosis assays.

Also, Buchta *et al* (2021) showed that the same patient sample could alter Cq (Quantification Cycle) values if run with different diagnostic PCR kits with different targeting genes. This is critical, once Cq values are used as a reference to define clinical decisions and

to guide patient care. Therefore, the choice of a SARS-CoV-2 gene target on diagnosing and patient monitoring is crucial [62].

Table 3. The real-time qPCR kit for SARS-CoV-2 detection and their mismatch frequency.

Source	**Target gene**	**Mismatch analysis**		**Refs**
		Frequency*	**Total Samples**	
CDC (China)	ORF1ab	0.4% 0.05 - 0.39% 1.1% 0.03%	992 ~16,000 177 2,569	56, 57, 58, 78
	N	12.7% 18.8% 85.3% 13.9%	992 16,662 177 2,569	56, 57, 58, 78
Charité	E	0.4% 0.03% - 0.14% 1.1% 0%	992 ~16,000 177 2,569	56, 57, 58, 78
	ORF1b	99.8% 100% 0.03%	992 17,004 2,569	57, 58, 78
HKU	N	0.5% 0.3% 58.2% 0.07%	992 16,667 177 2,569	56, 57, 58, 78
	ORF1b	0.2% 0%	16,932 2,569	58, 78
CDC (US)	N	0.2 - 3.9% 1.6% 1.7% 0.3%	992 16,920 177 2,569	56, 57, 58, 78

Another important aspect of SARS-CoV-2 detection by RT-qPCR is the endogenous internal control. To guarantee uniformity, reproducibility, and the extraction process quality, the "Minimum Information for publication of Quantitative real time PCR Experiments" guideline recommend that the choice of reference endogenous genes, also called endogenous housekeeping gene, should be essential part of RT-qPCR experiments [13]. FDA and other authors have reported the optimal human endogenous genes in the SARS-CoV-2 RT-qPCR detection. The RNAse P and Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) are

among the genes that act as an excellent internal control by excluding the possibility of false results due to the presence of low quality and integrity of RNA samples [63–65].

4.5 Data analysis

The result interpretation depends on detection kit guides that recommend a maximum Quantification Cycle (Cq) to be considered a positive result. The Cq represents the amplification cycle that trespasses the specific fluorescence intensity (named threshold line) programmed automatically or manually on equipment. The variation of Cq values reported in few studies was critically discussed [13]. In diagnostic laboratories day-to-day, some situations can impact the result analysis. For example, suppose a patient has a Cq strictly within the limit recommended by the kit guide. In that case, some analysts may consider this sample as positive, while others may consider automatically or manually as negative.

It has been reported that Cq values of 25-28 are considered appropriate as an indicator of SARS-CoV-2 positivity, and higher values (> 28) could be due to the inactivation of Taq-polymerase [57]. Generally, the Cq value below 40 is recommended as a SARS-CoV-2 positivity in different protocol RT-qPCR kits. However, some conditions mentioned before, such as collection site, period of collection, sample conservation, and transportation, might affect the sensitivity of the method. Vogels *et al* (2020) highlighted the possibility of Cq values >35 could be considered negative. Therefore, it is required from the analyst's experience to interpret false-positive and -negative samples.

Also, inconclusive results are also quite common. It can be obtained, for example, in diagnostic kits with two gene targets, where only one target amplifies. This common situation reinforces the influencing factors on SARS-CoV-2 diagnostic here discussed, once the same methodology for genetic material assessment was performed, including collecting site, extraction and RT-qPCR methodologies, sample storage and manipulation. Besides, it elucidates the concern on a reliable result in diagnostic kits that use only one gene target.

4.6 Primer-probe dimerization occurrence

Currently, many companies have developed diagnostic RT-qPCR kits without proper validation testing. One of the influencing parameters of sensitivity and specificity that might alter the data analysis is the primers/probe's dimer occurrence. Won *et al*. (2020) and Jaeger *et al*. (2021) showed that the US-CDC kit (N gene) had unexpected amplification during RT-qPCR in negative samples and no-template controls [66,67]. This was also demonstrated by another study with Charité RT-qPCR kit (E gene) that found high amounts of unspecific

signals in late cycles in no-template control [68]. These examples are categorical in describing the primer-dimer as a cause of the unspecific amplification.

In this study, we provide an in-silico analysis to confirm the dimer possibility of primer-probe sets. Self- and heterodimer formation was performed using OligoAnalyzer v3.1 (Integrated DNA Technologies®). To determine the likelihood of a primer/probe binding to each other, the software calculates the Gibbs free energy (ΔG) of nucleic acid hybridization as an indicator of dimer formation [69]. Our analysis displayed the binding energy variation of self and hetero-dimer among RT-qPCR kits for SARS-CoV-2 detection.

It has been established as the optimal standard properties for a primer set, including primer size, product size, melting temperature, GC content, and binding energy [70]. The most important property is the thermodynamic parameter that guarantees the non-occurrence of primer/probe dimerization for the designer. In other words, larger negative values of free energy binding value imply a higher probability of self and hetero hybridization [71]. The optimal free binding energy required to break the dimerization well tolerated is usually ≥ - 9 kcal/mole [72–74]. Among of primer/probe set analyzed, three primer/probe with < - 9 kcal/mole binding energy in the self-dimer analysis were detected: China CDC (ORF1ab target), Charité (E target), and HKU (N target) (Table 4). Moreover, the probe of RT-qPCR kit HKU for N target detection had the lowest self-dimer energy (-14.35). In the hetero-dimer analysis, the probe-reverse primer dimerization was detected with < -9 kcal/mole for China CDC (N target), Charité (E target), and US CDC (N2 and N3 target) (Table 4).

Jaeger *et al*. (2021) and Park *et al*. (2020) proposed several points to avoid primer/probe dimer formation and to optimize qPCR performance [67,75]. Among them, the reduction of primer set concentration, probe concentration, $MgSO_4$ concentration, annealing and extension temperature, and reverse transcription time. Therefore, it is well known the importance of primer design and optimization of qPCR reaction for proper validation and commercial distribution of diagnostic kits [76].

Table 4. Binding energy of self- and hetero dimer occurrence for different RT-qPCR kit for SARS-CoV-2 detection.

Source	Target gene	Sequence (5`- 3`)	Self-dimer Binding Energy (kcal/mole)	Heterodimer Binding Energy (kcal/mole)
CDC (China)	ORF1ab	F – CCCTGTGGGTTTTACACTTAA P – CCGTCTGCGGTATGTGGAAAGGTTATGG R - ACGATTGTGCATCAGCTGA	-6.14 -6.68 -13.39	Probe-Reverse -5.09
	N	F – GGGGAACTTCTCCTGCTAGAAT P – TTGCTGCTGCTTGACAGATT R - CAGACATTTTGCTCTCAAGCTG	-5.12 -3.55 -6.34	Probe-Reverse -10.21
Charité	E	F – ACAGGTACGTTAATAGTTAATAGCGT P – ACACTAGCCATCCTTACTGCGCTTCG R - ATATTGCAGCAGTACGCACACA	-6.3 -9.89 -7.05	Probe-Reverse -9
	ORF1b	F – GTGAAATGGTCATGTGTGGCGG P – CAGGTGGAACCTCATCAGGAGATGC R - CAAATGTTAAAAACACTATTAGCATA	-5.38 -6.01 -5.24	Probe-Reverse -6.57
HKU	N	F – TAATCAGACAAGGAACTGATTA P – GCAAATTGTGCAATTTGCGG R - CGAAGGTGTGACTTCCATG	-9.51 -14.35 -5.38	Forward-Reverse -6.59
	ORF1b	F – TGGGGTTTTACAGGTAACCT P – TAGTTGTGATGCAATCATGACTAG R - AACACGCTTAACAAAGCACTC	-6.36 -8.53 -6.68	Probe-Reverse -5.24
CDC (US)	N1	F – GACCCCAAAATCAGCGAAAT P – ACCCCGCATTACGTTTGGTGGACC R - TCTGGTTACTGCCAGTTGAATCTG	-3.61 -6.3 -6.62	Forward-Probe -8.91
	N2	F – TTACAAACATTGGCCGCAAA P – ACAATTTGCCCCCAGCGCTTCAG R - GCGCGACATTCCGAAGAA	-9.28 -13.09 -10.36	Probe-Reverse -9.89
	N3	F – GGGAGCCTTGAATACACCAAAA P – AYCACATTGGCACCCGCAATCCTG R - TGTAGCACGATTGCAGCATTG	-3.9 -5.37 -7.05	Probe-Reverse -10.09

4.7 Other influencing factors

Besides the factors mentioned above, Bentivegna *et al* (2021) discussed the possibility of false negative results in RT-qPCR in patients who have tested negative between two positive exams [77]. According to the authors, the negative result could be obtained due to a

prolonged viral clearance, which raises another point of discussion to patients who have been qualified as reinfection ones.

After all of the factors mentioned, it is also important to highlight the patients with pulmonary CT finding but who tested negative for SARS-CoV-2 (Table 5). These patients usually stay in isolation for long periods, considering the clinical condition, even with a negative molecular test. Many of the factors here cited may be the main cause for the negative result, or even more than one factor, combined. The point is, after more than one year that COVID-19 became pandemic, there is still much to learn about the immunity response to the virus, and its fast mutation rate, which has a direct impact on molecular tests and disease control.

Table 5. Factors that has a influence on SARS-CoV-2 diagnose and recommendations to minimize false results.

Factor	Authors recommendation
Disease staging	Collection centers should give orientation about the correct collection time (3-10 days after first symptoms). Samples outside this window should not be accepted.
Sample collection methodology	Ideally, the collection should always be performed by a trained team. Self-collection might impact the viral load recovered.
Storage sample	Samples should be stored at 2-8ºC/3days at maximum. If diagnostic centers cannot keep the samples in these conditions, the samples should be discharged and collected again.
RNA extraction methodologies	Column-based methods and extraction methods upgrades are recommended to improve the test sensitivity.
RT-qPCR detection kit	The same kit must be used to a patient if there's an intention to compare or monitoring the infection. Among the possible targets, E and RdRp genes has shown to has a higher sensibility.
Data analysis	We recommend the use of kits with two gene targets, if possible.
Primer-probe dimerization occurrence	Primers testing should be performed again, if any bias occurs on reactions.

5 Conclusion

The target gene standardization used in SARS-CoV-2 detection, as well as the method and collection site would be the ideal measure to enable comparison between patients results and CTs. However, considering the diversity in fabricants kits worldwide, it is impracticable to standardize all of these features. Therefore, we reinforce the importance of COVID-19 diagnosis being concluded based on combined tests, so that the correct diagnosis can be reached, even with the occurrence of false-negative or false-positive molecular tests. Additionally, the association of clinical-epidemiological information and complementary exams would help to avoid false-positive or false-negative results.

6 Expert opinion

COVID-19 massive diagnosis is a critical method for effectively monitoring and controlling its spread. Furthermore, in the absence of a fully vaccinated population, increasing COVID-19 monitoring capability with trustworthy results for large-scale sampling seems to be the most promising option for to understand, contain, and defeat this epidemic. The COVID-19 pandemic emphasizes the importance of establishing a robust and long-term mechanism for the accelerated growth, dissemination, and implementation of adapted diagnostic tests against the virus.

The rapid and intensive manufacture of molecular kits by numerous laboratories have significantly assisted countries' testing capacities. A large number of kits are now commercially available while others are still being developed. The kits that have gained federal agency of the Department of Health and Human Services approvals are the most preferred to use or proceed with when implementing the massive diagnosis. However, several findings, even some related in this study, revealed the limitations and weak points of the available COVID-19 diagnostic kits. As mentioned in this study, multiple influence factors in sample selection, nucleic acid extraction, and RT-qPCR, may be decisive for a successful diagnosis. Therefore, the knowledge about these factors is essential to provide a reliable diagnostic and even understand what could led to a wrong diagnosis.

It is clear in our review that the sample collection methodology and storage can directly influence in the viral loads recovered. This is a quite important point because a bad storage sample could lead to false-negative results, which is a results of virus degradation and not essentially the absence of the virus. Nasopharyngeal swabs are still a good alternative for almost all SARS-CoV-2 variants (Alpha, Beta, Gamma and Delta), with exception for the Ômicron variant that is preferable detected by the oropharyngeal swab. In case of Ômicron,

this happens because Ômicron replicates more efficiently in oropharyngeal region. Therefore, we claim attention to the need of best practices in this context.

In addition, the RNA extraction of sample possible infected with SARS-COV-2 is a sensible step with important reflection in the result and diagnostic. Based on that, the professionals must be informed the possible variation of RNA viral yield due to the RNA extraction methodology. The best extraction method identified was the acid-Ph-based method.

The interpretation of molecular result can be a postanalytical issue and must be relevant during the training sections of the professionals. For example, quantification cycle (Cq) line when incorrect determined could represent an inconclusive result that can contribute for the clinical misconduct. Another issue identified in this review was the mismatches and primer/probe dimerization that, for example, can threaten the precision of COVID-19 diagnosis. Diagnostic kit vendors, as well as diagnostic laboratories, must be aware of these concerns in order to avoid more consequences for the public. As a result, additional tests and studies are urgently needed to improve the production of a SARS-CoV-2 detection kit with more sensitivity and specificity.

Since there are several emergent SARS-CoV-2 variants, it is necessary the appropriate supervision and regulation to avoid the inaccuracy issues in the diagnosis kits due to the high range of genetic variability of SARS-CoV-2 variants.

Acknowledgments

The authors are grateful to the Brazilian Agencies CNPq, CAPES, and the Federal University of Ceará for fellowships and financial support.

Conflicts of interest

None of the authors has a conflict of interest.

Funding

This work was supported by the Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (number: 03195011/2020) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (number: 88881.505364/2020-01 - edital Nº 9/2020).

Author contributions statement

Conceptualization: Luina Benevides Lima and Felipe Pantoja Mesquita; Data curation: Felipe Pantoja Mesquita; Formal analysis: Luina Benevides Lima, Pedro F. N. Souza and Felipe Pantoja Mesquita; Funding acquisition: Maria Elisabete Amaral de Moraes and Raquel

Carvalho Montenegro; Methodology: Luina Benevides Lima, Pedro F. N. Souza and Felipe Pantoja Mesquita; Supervision: Raquel Carvalho Montenegro; Writing–original draft: Luina Benevides Lima, Pedro F. N. Souza and Felipe Pantoja Mesquita; Writing, review and editing: Luina Benevides Lima, Pedro F. N. Souza and Felipe Pantoja Mesquita

References

* of importance

** of considerable importance

[1] Zhou P, Yang X-L, Wang X-G, *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–273.

*** This manuscript provides important information about the origin of SARS-CoV-2**

[2] WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data. Available from: https://covid19.who.int/.

[3] Okamaoto K, Shirato K, Nao N, *et al*. Assessment of Real-Time RT-QPCR Kits for SARS-CoV-2 Detection. Jpn J Infect Dis. 2020;73:366–368.

**** This manuscript higlights the kits for SARS-CoV-2 detection and make comparisons among them.**

[4] Wikramaratna PS, Paton RS, Ghafari M, *et al*. Estimating the false-negative test probability of SARS-CoV-2 by RT-QPCR. Eurosurveillance. 2020;25.

**** This manuscript reports what are the chances of result of COVID-19 being a wrong.**

[5] Wang Y, Kang H, Liu X, *et al*. Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. J Med Virol . 2020 ;92:538–539.

[6] Mattioli IA, Hassan A, Oliveira ON, *et al*. On the Challenges for the Diagnosis of SARS-CoV-2 Based on a Review of Current Methodologies. ACS Sensors. 2020;5:3655–3677.

**** This manuscript reports the challengings in COVID-19 diagnosis**

[7] Corman VM, Muth D, Niemeyer D, *et al*. Hosts and Sources of Endemic Human Coronaviruses. Adv Virus Res. 2018;163–188.

[8] Weiss SR, Leibowitz JL. Coronavirus Pathogenesis. Adv Virus Res. 2011;85–164.

[9] Andersen KG, Rambaut A, Lipkin WI, *et al*. The proximal origin of SARS-CoV-2. Nat Med. 2020;26:450–452.

[10] Wu D, Wu T, Liu Q, *et al*. The SARS-CoV-2 outbreak: What we know. Int J Infect Dis. 2020;94:44–48.

[11] Lan J, Ge J, Yu J, *et al*. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581:215–220.

[12] Yan R, Zhang Y, Li Y, *et al*. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020;367:1444–1448.

[13] Bustin SA, Benes V, Garson JA, *et al*. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem. 2009;55:611–622.

*** This manuscript is important to understand the protocols for experiments involving qRT-QPCR.**

[14] Corman VM, Landt O, Kaiser M, *et al*. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-QPCR. Eurosurveillance. 2020;25.

*** This manuscript is reports some regions of SARS-CoV-2 that could be employed in the detection of SARS-CoV-2 by qRT-QPCR.**

[15] Freeman WM, Walker SJ, Vrana KE. Quantitative RT-QPCR: Pitfalls and Potential. Biotechniques. 1999;26:112–125.

[16] Kim S, Hwang YJ, Kwak Y. Prolonged SARS-CoV-2 detection and reversed RT-QPCR results in mild or asymptomatic patients. Infect Dis (Auckl). 2021;53:31–37.

[17] Ling Y, Xu S-B, Lin Y-X, *et al*. Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. Chin Med J (Engl). 2020;133:1039–1043.

[18] Peng L, Liu J, Xu W, *et al*. SARS-CoV-2 can be detected in urine, blood, anal swabs, and oropharyngeal swabs specimens. J Med Virol. 2020;92:1676–1680.

[19] Azghandi M, Kerachian MA. Detection of novel coronavirus (SARS-CoV-2) RNA in peripheral blood specimens. J Transl Med. 2020;18:412.

*** This manuscript reports a methodology to detect SARS-CoV-2 in blood samples.**

[20] Bustin S, Coward A, Sadler G, *et al*. CoV2-ID, a MIQE-compliant sub-20-min 5-plex RT-QPCR assay targeting SARS-CoV-2 for the diagnosis of COVID-19. Sci Rep. 2020;10:22214.

[21] He X, Cheng X, Feng X, *et al*. Clinical Symptom Differences Between Mild and Severe COVID-19 Patients in China: A Meta-Analysis. Front Public Heal. 2021;8.

[22] Petruzzi G, De Virgilio A, Pichi B, *et al*. <scp>COVID</scp> -19: Nasal and oropharyngeal swab. Head Neck. 2020;42:1303–1304.

[23] Islam KU, Iqbal J. An Update on Molecular Diagnostics for COVID-19. Front Cell Infect Microbiol. 2020;10.

[24] Kanji JN, Zelyas N, MacDonald C, *et al*. False negative rate of COVID-19 PCR

testing: a discordant testing analysis. Virol J. 2021;18:13.

*** This manuscript tries to understand the False-negative results of routinely used assays are important to confirm adequate clinical performance.**

[25] Mancuso P, Venturelli F, Vicentini M, *et al*. Temporal profile and determinants of viral shedding and of viral clearance confirmation on nasopharyngeal swabs from SARS-CoV-2-positive subjects: a population-based prospective cohort study in Reggio Emilia, Italy. BMJ Open. 2020;10:e040380.

[26] Pan L, Mu M, Yang P, *et al*. Clinical Characteristics of COVID-19 Patients With Digestive Symptoms in Hubei, China: A Descriptive, Cross-Sectional, Multicenter Study. Am J Gastroenterol. 2020;115:766–773.

[27] Wölfel R, Corman VM, Guggemos W, *et al*. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020;581:465–469.

[28] Zou L, Ruan F, Huang M, *et al*. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med. 2020;382:1177–1179.

[29] Diagnostic testing for SARS-CoV-2. Available from: https://www.who.int/publications/i/item/diagnostic-testing-for-SARS-CoV-2.

[30] Liu M, Li Q, Zhou J, *et al*. Value of swab types and collection time on SARS-COV-2 detection using RT-QPCR assay. J Virol Methods. 2020;286:113974.

[31] Muniz I de AF, Linden L Van der, Santos ME, *et al*. SARS-CoV-2 and Saliva as a Diagnostic Tool: A Real Possibility. Pesqui Bras Odontopediatria Clin Integr. 2020;20:1–7.

[32] Chen JH-K, Yip CC-Y, Poon RW-S, *et al*. Evaluating the use of posterior oropharyngeal saliva in a point-of-care assay for the detection of SARS-CoV-2. Emerg Microbes Infect. 2020;9:1356–1359.

****This manuscript reports an study that evalaute case-by-case of sampling from oropharyngeal saliva to perform diagnosis.**

[33] Wang W, Xu Y, Gao R, *et al*. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. JAMA. 2020;323:1843–1844.

[34] Han H, Luo Q, Mo F, *et al*. SARS-CoV-2 RNA more readily detected in induced sputum than in throat swabs of convalescent COVID-19 patients. Lancet Infect Dis. 2020;20:655–656.

[35] Chen C, Gao G, Xu Y, *et al*. SARS-CoV-2–Positive Sputum and Feces After Conversion of Pharyngeal Samples in Patients With COVID-19. Ann Intern Med. 2020;172:832–834.

[36] Xia J, Tong J, Liu M, *et al*. Evaluation of coronavirus in tears and conjunctival secretions of patients with SARS-CoV-2 infection. J Med Virol . 2020 ;92:589–594. Available from: https://pubmed.ncbi.nlm.nih.gov/32100876/.

[37] Pondaven-Letourmy S, Alvin F, Boumghit Y, *et al*. How to perform a nasopharyngeal swab in adults and children in the COVID-19 era. Eur Ann Otorhinolaryngol Head Neck Dis . 2020 ;137:325–327. Available from: https://pubmed.ncbi.nlm.nih.gov/32646750/.

[38] Landry ML, Criscuolo J, Peaper DR. Challenges in use of saliva for detection of SARS CoV-2 RNA in symptomatic outpatients. J Clin Virol . 2020 ;130:104567. Available from: https://pubmed.ncbi.nlm.nih.gov/32750665/.

[39] Williams E, Bond K, Zhang B, *et al*. Saliva as a Noninvasive Specimen for Detection of SARS-CoV-2. McAdam AJ, editor. J Clin Microbiol . 2020 ;58. Available from: https://pubmed.ncbi.nlm.nih.gov/32317257/.

[40] Wyllie AL, Fournier J, Casanovas-Massana A, *et al*. Saliva or Nasopharyngeal Swab Specimens for Detection of SARS-CoV-2. N Engl J Med . 2020 ;383:1283–1286. Available from: https://pubmed.ncbi.nlm.nih.gov/32857487/.

[41] Olivieri EHR, de Andrade Franco L, Pereira RG, *et al*. Biobanking Practice: RNA Storage at Low Concentration Affects Integrity. Biopreserv Biobank . 2014 ;12:46–52. Available from: https://pubmed.ncbi.nlm.nih.gov/24620769/.

[42] Lippi G, Adeli K, Ferrari M, *et al*. Biosafety measures for preventing infection from COVID-19 in clinical laboratories: IFCC Taskforce Recommendations. Clin Chem Lab Med . 2020 ;58:1053–1062. Available from: https://pubmed.ncbi.nlm.nih.gov/32396137/.

[43] Chu AW-H, Chan W-M, Ip JD, *et al*. Evaluation of simple nucleic acid extraction methods for the detection of SARS-CoV-2 in nasopharyngeal and saliva specimens during global shortage of extraction kits. J Clin Virol . 2020 ;129:104519. Available from: https://pubmed.ncbi.nlm.nih.gov/32629187/.

****This manuscript compares the methods of RNA extraction for detection by nasopharyngeal and saliva samples.**

[44] Ambrosi C, Prezioso C, Checconi P, *et al*. SARS-CoV-2: Comparative analysis of different RNA extraction methods. J Virol Methods . 2021 ;287:114008. Available from: https://pubmed.ncbi.nlm.nih.gov/33160015/.

[45] Wozniak A, Cerda A, Ibarra-Henríquez C, *et al*. A simple RNA preparation method for SARS-CoV-2 detection by RT-qPCR. Sci Rep . 2020 ;10:16608. Available from:

https://pubmed.ncbi.nlm.nih.gov/33024174/.

[46] Klein S, Müller TG, Khalid D, *et al*. SARS-CoV-2 RNA Extraction Using Magnetic Beads for Rapid Large-Scale Testing by RT-qPCR and RT-LAMP. Viruses . 2020 ;12:863. Available from: https://pubmed.ncbi.nlm.nih.gov/32784757/.

[47] Smyrlaki I, Ekman M, Lentini A, *et al*. Massive and rapid COVID-19 testing is feasible by extraction-free SARS-CoV-2 RT-QPCR. Nat Commun . 2020 ;11:4812. Available from: https://pubmed.ncbi.nlm.nih.gov/32968075/.

[48] Sule WF, Oluwayelu DO. Real-time RT-QPCR for COVID-19 diagnosis: challenges and prospects. Pan Afr Med J . 2020 ;35:121. Available from: https://pubmed.ncbi.nlm.nih.gov/33282076/.

[49] Tahamtan A, Ardebili A. Real-time RT-QPCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn . 2020 ;20:453–454. Available from: https://pubmed.ncbi.nlm.nih.gov/32297805/.

[50] Freire-Paspuel B, Garcia-Bereguiain MA. Analytical sensitivity and clinical performance of a triplex RT-qPCR assay using CDC N1, N2, and RP targets for SARS-CoV-2 diagnosis. Int J Infect Dis . 2021 ;102:14–16. Available from: https://pubmed.ncbi.nlm.nih.gov/33115681/.

[51] Araujo DB, Machado RRG, Amgarten DE, *et al*. SARS-CoV-2 isolation from the first reported patients in Brazil and establishment of a coordinated task network. Mem Inst Oswaldo Cruz . 2020 ;115:1–8. Available from: https://pubmed.ncbi.nlm.nih.gov/33111751/.

[52] Arndt AL, Larson BJ, Hogue BG. A Conserved Domain in the Coronavirus Membrane Protein Tail Is Important for Virus Assembly. J Virol . 2010 ;84:11418–11428. Available from: https://pubmed.ncbi.nlm.nih.gov/20719948/.

[53] Wakida H, Kawata K, Yamaji Y, *et al*. Stability of RNA sequences derived from the coronavirus genome in human cells. Biochem Biophys Res Commun . 2020 ;527:993–999. Available from: https://pubmed.ncbi.nlm.nih.gov/32446559/.

[54] Jung Y, Park G-S, Moon JH, *et al*. Comparative Analysis of Primer–Probe Sets for RT-qPCR of COVID-19 Causative Virus (SARS-CoV-2). ACS Infect Dis . 2020 ;6:2513–2523. Available from: https://pubmed.ncbi.nlm.nih.gov/32786273/.

[55] Valesano AL, Rumfelt KE, Dimcheff DE, *et al*. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. Pekosz A, editor. PLOS Pathog . 2021 ;17:e1009499. Available from: https://pubmed.ncbi.nlm.nih.gov/33826681/.

[56] Santos R da S, Bret RSC, Moreira AC de OM, *et al*. In silico analysis of mismatches in RT-qPCR assays of 177 SARS-CoV-2 sequences from Brazil. Rev Soc Bras Med Trop . 2020 ;53:1–5. Available from: https://pubmed.ncbi.nlm.nih.gov/33263691/.

[57] Vogels CBF, Brito AF, Wyllie AL, *et al*. Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. Nat Microbiol . 2020 ;5:1299–1305. Available from: https://pubmed.ncbi.nlm.nih.gov/32651556/.

[58] Khan KA, Cheung P. Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome. R Soc Open Sci . 2020 ;7:200636. Available from: https://pubmed.ncbi.nlm.nih.gov/32742701/.

[59] Corman VM, Drosten C. Authors' response: SARS-CoV-2 detection by real-time RT-QPCR. Eurosurveillance . 2020 ;25. Available from: https://pubmed.ncbi.nlm.nih.gov/32489177/.

[60] Yan Y, Chang L, Wang L. Laboratory testing of SARS-CoV, MERS-CoV, and SARS-CoV-2 (2019-nCoV): Current status, challenges, and countermeasures. Rev Med Virol . 2020 ;30. Available from: https://pubmed.ncbi.nlm.nih.gov/32302058/.

[61] Ramírez JD, Muñoz M, Patiño LH, *et al*. Will the emergent SARS-CoV2 B.1.1.7 lineage affect molecular diagnosis of COVID-19? J Med Virol . 2021 ;93:2566–2568. Available from: https://pubmed.ncbi.nlm.nih.gov/33506970/.

[62] Buchta C, Görzer I, Chiba P, *et al*. Variability of cycle threshold values in an external quality assessment scheme for detection of the SARS-CoV-2 virus genome by RT-QPCR. Clin Chem Lab Med . 2021 ;59:987–994. Available from: https://pubmed.ncbi.nlm.nih.gov/33554519/.

[63] Ñique AM, Coronado-Marquina F, Mendez Rico JA, *et al*. A faster and less costly alternative for RNA extraction of SARS-CoV-2 using proteinase k treatment followed by thermal shock. Aboelhadid SM, editor. PLoS One . 2021 ;16:e0248885. Available from: https://pubmed.ncbi.nlm.nih.gov/33760876/.

[64] COVID-19 Coronavirus Real Time PCR Kit - Instructions for Use | FDA . . Available from: https://www.fda.gov/media/139279.

[65] Miranda RL, Guterres A, de Azeredo Lima CH, *et al*. Misinterpretation of viral load in COVID-19 clinical outcomes. Virus Res . 2021 ;296:198340. Available from: /pmc/articles/PMC7881726/.

[66] Won J, Lee S, Park M, *et al*. Development of a Laboratory-safe and Low-cost Detection Protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19). Exp Neurobiol . 2020 ;29:107–119. Available from:

https://pubmed.ncbi.nlm.nih.gov/32156101/.

[67] Jaeger LH, Nascimento TC, Rocha FD, *et al*. Adjusting RT-qPCR conditions to avoid unspecific amplification in SARS-CoV-2 diagnosis. Int J Infect Dis . 2021 ;102:437–439. Available from: https://pubmed.ncbi.nlm.nih.gov/33130201/.

[68] Konrad R, Eberle U, Dangel A, *et al*. Rapid establishment of laboratory diagnostics for the novel coronavirus SARS-CoV-2 in Bavaria, Germany, February 2020. Eurosurveillance . 2020 ;25. Available from: https://pubmed.ncbi.nlm.nih.gov/32156330/.

[69] Mann T, Humbert R, Dorschner M, *et al*. A thermodynamic approach to PCR primer design. Nucleic Acids Res . 2009 ;37:e95–e95. Available from: https://pubmed.ncbi.nlm.nih.gov/19528077/.

[70] Hendling M, Barišić I. In-silico Design of DNA Oligonucleotides: Challenges and Approaches. Comput Struct Biotechnol J . 2019 ;17:1056–1065. Available from: https://pubmed.ncbi.nlm.nih.gov/31452858/.

[71] Meagher RJ, Priye A, Light YK, *et al*. Impact of primer dimers and self-amplifying hairpins on reverse transcription loop-mediated isothermal amplification detection of viral RNA. Analyst . 2018 ;143:1924–1933. Available from: https://pubmed.ncbi.nlm.nih.gov/29620773/.

[72] Rachlin J, Ding C, Cantor C, *et al*. Computational tradeoffs in multiplex PCR assay design for SNP genotyping. BMC Genomics . 2005 ;6:102. Available from: https://pubmed.ncbi.nlm.nih.gov/16042802/.

[73] Rozen S, Skaletsky H. Primer3 on the WWW for General Users and for Biologist Programmers. Bioinforma Methods Protoc . New Jersey: Humana Press; 2000 . p. 365–386. Available from: https://pubmed.ncbi.nlm.nih.gov/10547847/.

[74] Shen Z, Qu W, Wang W, *et al*. MPprimer: a program for reliable multiplex PCR primer design. BMC Bioinformatics . 2010 ;11:143. Available from: https://pubmed.ncbi.nlm.nih.gov/20298595/.

[75] Park M, Won J, Choi BY, *et al*. Optimization of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using PCR and real-time PCR. Exp Mol Med . 2020 ;52:963–977. Available from: https://pubmed.ncbi.nlm.nih.gov/32546849/.

[76] Thornton B, Basu C. Real-time PCR (qPCR) primer design using free online software. Biochem Mol Biol Educ . 2011 ;39:145–154. Available from: https://pubmed.ncbi.nlm.nih.gov/21445907/.

[77] Bentivegna E, Sentimentale A, Luciani M, *et al*. New IgM seroconversion and positive RT-PCR test after exposure to the virus in recovered COVID-19 patient. J Med Virol . 2021 ;93:97–98. Available from: https://pubmed.ncbi.nlm.nih.gov/32525558/.

CAPÍTULO III

Genomic surveillance: Circulating lineages and genomic variation of SARS-CoV-2 in early pandemics in Ceará state, northeast Brazil

Genomic Surveillance: Circulating lineages and genomic variation of SARS-CoV-2 in early pandemics in Ceará state, Northeast Brazil

Francisca Andréa da Silva Oliveira[a, b, c,1], Maísa Viana de Holanda[b], Luína Benevides Lima[a], Mariana Brito Dantas[b], Igor Oliveira Duarte[c], Luzia Gabrielle Zeferino de Castro[c], Laís Lacerda Brasil de Oliveira[a], Carlos Roberto Koscky Paier[a], Caroline de Fátima Aquino Moreira-Nunes[a], Nicholas Costa Barroso Lima[d], Vânia Maria Maciel Melo[b,c], Raquel Carvalho Montenegro[a,b]

[a]Laboratório de Farmacogenética, Núcleo de Pesquisa e Desenvolvimento de Medicamentos, Universidade Federal do Ceará, Rua Coronel Nunes de Melo, n° 1000, 60430-270, Fortaleza, CE, Brazil.
[b]Central de Genômica e Bioinformática, Núcleo de Pesquisa e Desenvolvimento de Medicamentos, Universidade Federal do Ceará, Rua Coronel Nunes de Melo, n° 1000, 60430-270, Fortaleza, CE, Brazil.
[c]Laboratório de Ecologia Microbiana e Biotecnologia, Departamento de Biologia, Universidade Federal do Ceará, Campus do Pici, s/n, Bloco 909, Pici, 60455-760, Fortaleza, CE, Brazil.
[d]Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Ceará, Campus do Pici, s/n, Bloco 907, Pici, 60440-900, Fortaleza-CE, Brazil.

Corresponding author at: Central de Genômica e Bioinformática, Núcleo de Pesquisa e Desenvolvimento de Medicamentos, Universidade Federal do Ceará, Rua Coronel Nunes de Melo, n° 1000, 60430-270, Fortaleza-CE, Brazil.
E-mail address: andreasilvaoli@gmail.com

Abstract

Ceará state was considered the second highest in number of cases and death rate in the Northeast of Brazil by COVID-19. Despite that, the early dynamics of the pandemic in the Ceará was not yet well understood due the low genomic surveillance of SARS-CoV-2 in 2020. In this study, we analyze the circulating lineages and the genomic variation of the virus in Ceará state. Thirty-four genomes were sequenced and combined with sequences available in GISAID database from March to December 2020. The lineages most prevalent detected were B.1.1.33 (39.1%) followed by B.1 (18.3%), B.1.212 (16.5%) and P.2 (12.2%). Analyzing the mutations, a total of 202 single-nucleotide variant (SNVs) were identified among the 34 genomes, of which 127 were missense, 74 synonymous, and 1 nonsense mutation. Some of the detected mutations were associated with increased transmission rate and severity. ORF1ab was the gene with more accumulated SNVs, followed by the spike protein with 70 and 23 missense mutations, respectively. Although our results have limitations by low sampling numbers, they provide insight to better understanding and description of circulating lineages and genetic diversity, helping to elucidate how the pandemic course of the SARS-CoV-2 in Ceará happened.

Keywords: COVID-19; Genome sequences; Mutations

1. Introduction

In December 2019, the disease known as COVID-19, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), was detected in Wuhan, China (Wang et al., 2020; Zhu et al., 2020), and was later declared as a pandemic by the World Health Organization (WHO) in March 2020. Since then, the virus has spread rapidly resulting in more than five million deaths worldwide. In Brazil, the first confirmed contagion of SARS-CoV-2 was in late February 2020 in the state of São Paulo (de Jesus et al., 2020). After this, the country rapidly became one of the epicenters of the pandemic with lineages B.1.1.28 and B.1.1.33 being the most prevalent in the early epidemic phase (Resende et al., 2021a). In late 2020, two variants, Zeta (P.2) (Voloch et al., 2020) and Gamma (P.1) (Faria et al., 2021a), descendants of lineage B.1.1.28, emerged and were associated with the second phase of the pandemic. With a population of over 9.2 million people, Ceará is an economically relevant state in Brazil, with a strong travel industry and a high-traffic airport located in its capital, Fortaleza. Ceará has the second highest number of cases in the Northeast of Brazil, with 26,204 deaths registered (http://covid.saude.gov.br, accessed on 25 February, 2022). The first reports of COVID-19 in Ceará dated from March 2020. From the early pandemic, genomic surveillance

has been an efficient tool to trace variants of SARS-CoV-2 and study the virus import and spread. Furthermore, in 2020, before a strong genomic surveillance service was established in Ceará, only few SARS-CoV-2 genomes were sequenced to infer the lineages that were circulating in the first months of pandemic. After the first SARS-CoV-2 genome sequence was available, in January 2020 (Wu et al., 2020), about nine million genomes were sequenced and shared on the Global Initiative on Sharing All Influenza Data (GISAID) database, allowing identification of the virus lineages worldwide. Mutations have emerged throughout the virus genome, but those related to gene S are more relevant, once its product, the Spike protein, is directly involved in the host cell entrance process (Fung and Liu, 2019). For example, a single amino acid change from aspartic acid to glycine at position 614 of protein Spike (D614G) became dominant in a short time, and was associated with increased transmission of the virus (Korber et al., 2020). Knowledge of new mutations and circulating lineage is essential for decision making on measures to contain the pandemic, since each variant may influence in the pathogenicity and transmissibility of the virus (Lauring and Hodcroft, 2021; Saito et al., 2021; Wang et al., 2021). In this study, we used thirty-four SARS-CoV-2 genome sequences to investigate the circulating lineages and detect mutation patterns to better understand the dispersion and evolution of the SARS-CoV-2 in the early phase of the epidemic in Ceará. Sequences from 2020 were used to determine the circulating lineages in this year, before the strike of the second wave, during which Gamma lineages were most prevalent. In addition, we highlight the importance of monitoring SARS-CoV-2 lineages through genomic surveillance as a measure to contain the pandemic.

2. Material and methods

2.1. Ethical Aspects

This research was approved by the Federal University of Ceará (UFC) Ethics Committee (CEP/CAAE: 31453320.7.0000.5054) and the Brazilian Ministry of Health SISGEN (A29A4F4).

2.2. Sample selection and viral detection by RT-qPCR

From 5,449 samples used to perform SARS-CoV-2 detection for diagnosis from July 2020 to June 2021, thirty-four clinical RT-qPCR positive samples with the lowest CT in each month were chosen for this study. The diagnostic procedure was performed at the Laboratory of Pharmacogenetics in the Drug Research and Development Center (NPDM), Federal

University of Ceará (UFC). Nasopharyngeal swabs were confirmed as positive for SARS-CoV-2 using iTaq Universal Probes One-Step Kit (Bio-Rad, USA) on a QuantStudio 5 instrument (Thermo Fisher Scientific, USA). The protocol used was established by the Centers for Disease Control and Prevention (CDC, Atlanta, USA), and to detect the presence of SARS-CoV-2 was used N1 and N2 genes from viral Nucleocapsid, and the human RNase P gene as an internal control.

2.3. Nucleic acid extraction, library preparation, and sequencing

The viral RNA was extracted from 140 µL of clinical samples using QIAamp Viral RNA Mini kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. The libraries were prepared using AmpliSeq Plus or COVIDSeq kits (Illumina, San Diego, USA), according to the manufacturer's protocols. AmpliSeq Plus libraries were purified with AMPure XP magnetic beads (Beckman Coulter, Brea, USA). Libraries were quantified using High Sensitivity dsDNA quantification kit with Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, USA), and mean fragment size was analyzed by TapeStation 4150 with DNA HS D1000 kit (Agilent, Santa Clara, USA). Library concentration was calculated and diluted to 4 nM. Libraries were pooled, denatured, and diluted to a final concentration of up to 12 pM and sequenced on MiSeq platform with MiSeq Reagent kit v2 (300-cycle) (Illumina, San Diego, USA) to generate reads of 2 x 150 bp. The DNA sequencing was performed at the Genomics and Bioinformatics Center (CeGenBio) of Drug Research and Development Center (NPDM) of the Federal University of Ceará, Brazil. The sequences are available in the GISAID database (https://gisaid.org/) (Supplementary Table S1).

2.4. Data analysis

Sequencing data were inspected for overall quality, adapter and size trimming with FastQC v0.11.9 (Andrews, 2010). Good quality reads were then processed and analyzed to variant calling and lineage classification. Reads were aligned to SARS-CoV-2 reference genome (NCBI acc. ID NC_045512.2) using BWA v0.7.17-r1188 (Li and Durbin, 2009). SortSam v2.18.29 from Picard Tools (http://broadinstitute.github.io/picard/) was used to sort the alignment, and samtools v1.11 (Danecek et al., 2021) was used to create an index of the alignment. GATK v.4.1.9.0 (Van der Auwera and O'Connor, 2020) called, selected and filtered variants according to the alignment of samples' reads to the reference genome. Variants were annotated according to the coding sequences of the reference using snpEff v5.0e (Cingolani et al., 2012a) and the resulting VCF file was converted into a table using

SnpSift v.4.3t (Cingolani et al., 2012b). In order to classify the samples into pango lineages we assembled samples' reads with *Skesa* v.2.1-SVN_551987:557549M (Souvorov et al., 2018) generating contigs for each genome sample. We used *RagTag* v.2.0.0 (Alonge et al., 2019) to order and orient contigs generated by *Skesa* using the reference genome and generated a superscaffold. Where no contigs aligned on the reference genome, 'N's were used to complete the superscaffold sequence. This step is necessary because *Pangolin* v.3.1.11 (github.com/cov-lineages/pangolin) was used to assign a pango lineage to each genome. Secondary analysis was performed with DRAGEN COVID Lineage app v.3.5.8 from Illumina`s BaseSpace (basespace.illumina.com). The sequences from 2020 were used to determine the distribution of lineages circulating in the state of Ceará. These sequences were combined to all 90 SARS-CoV-2 genome sequences available from GISAID (accession numbers are shown in Supplementary File S1) with collection date in the year 2020 and collection site within the state of Ceará to increase sample size for the analysis (Supplementary Table S2). Additionally, all 34 genomes sequenced were used to derive a count table with the mutations found in the dataset.

3. Results

We sequenced 34 samples of SARS-CoV-2 from Ceará, which were later classified into pango lineages (Supplementary Table S1). Among these, 25 were collected from July to December 2020 and nine from January to June 2021. Other 90 sequences from GISAID from the state of Ceará, and from the year 2020, were added to the dataset to determine the circulating SARS-CoV-2 lineages from March to December. Analyzing these sequences, eight viral lineages were identified in 2020, being B.1.1.33 (n=45, 39.1%) the most prevalent one, followed by B.1 (n=21; 18.3%), B.1.212 (n=19, 16.5%) and P.2 (n=14, 12.2%). Other lineages were found, although in less proportion: B.1.1.28 (n=9, 7.8%), B.1.1 (n=5, 4.3%), B.1.1.371 (n=1, 0.9%), and P.1 (n=1, 0.9%). Lineage B.1.1.33 was observed in all months from 2020, but more frequently from April to December (Figure 1). Lineages B.1 and B.1.212 were represented mainly up to September 2020. P.2 lineage (VOC Zeta) was first observed in April, though was not again sampled until November, but presented higher prevalence in December 2020, with 10 genomes (40% for this month). According to the analyzed data, the lineage B.1.1.28 did not have a high prevalence in Ceará, during the year 2020, being sampled only in July, November and December. Notably, we have found one genome sequence belonging to Gamma (P.1) variant in October 2020.

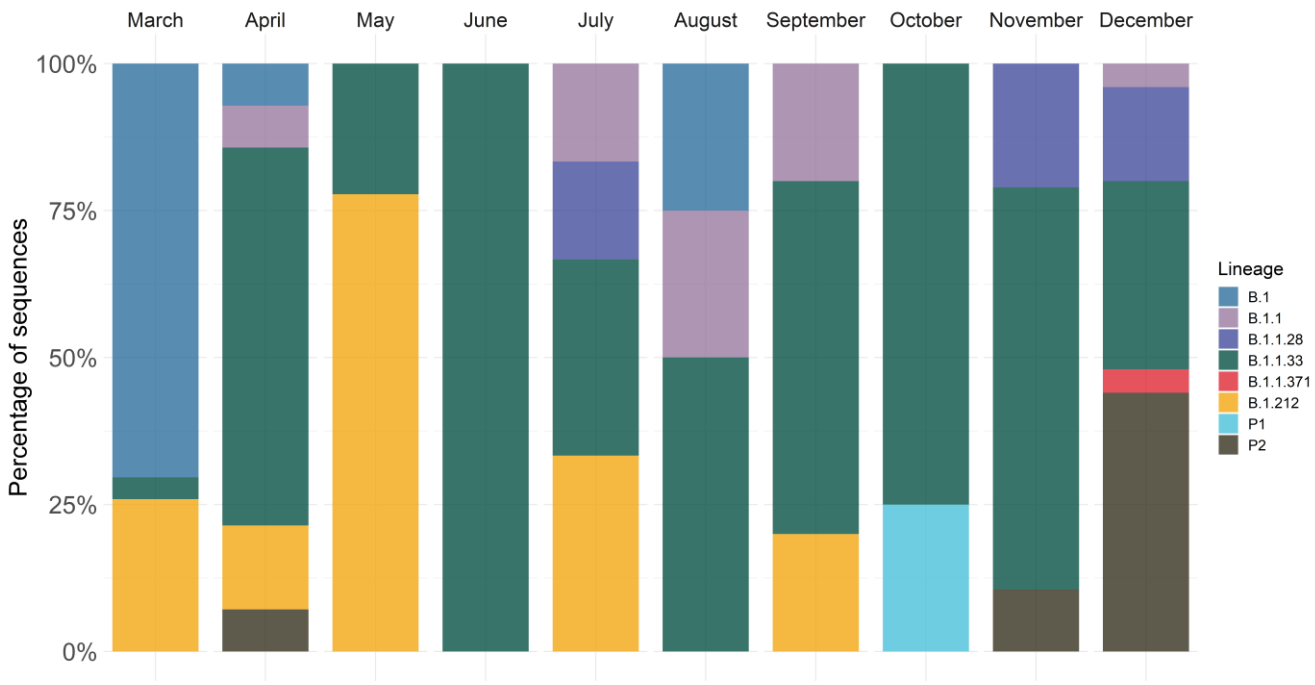


Figure 1. Prevalence of SAR-CoV-2 genome sequences for each month from Ceará, from present study and GISAID database as relative values

The occurrence of mutations in the 34 SARS-CoV-2 genomes sequenced were also assessed. A total of 202 single-nucleotide variants (SNVs) were found, among which 127 were missense (non-synonymous), 74 synonymous and 1 nonsense (Supplementary Table S3). Missense mutations with over 40% of prevalence within the dataset were C14408T (P4715L) in ORF1ab; A23403G (D614G) in gene S; T27299C (I33T) in ORF6; and G28881A (R203K), G28882A (R203R), G28883C (G204R) and T29148C (I292T) in gene N (Figure 2a). The gene with the highest concentration of SNVs was ORF1ab, with 70 missense mutations. Also, 23 missense SNVs were identified in the gene S and, among which, the mutation D614G was detected in 100% of our sequenced genomes (Figure 2b). Missense SNVs in the gene S with a frequency of at least 17% among the genomes analyzed included C21614T (L18F), C21621A (T20N), C21638T (P26S), G21974T (D138Y), G22132T (R190S), A22812C (K417T), G23012A (E484K), A23063T (N501Y), C23525T (H655Y), C24642T (T1027I), and G25088T (V1176F). The only deletion found within the dataset was the VOC Gamma synapomorphic deletion in position 11,287 (S3675-F3677_), which is a conservative 9-bp in-frame deletion. As expected, this mutation was found in the six genomes identified as lineage P.1. We found only one nonsense mutation, located in ORF7a, at position 27,673 (Q94*).

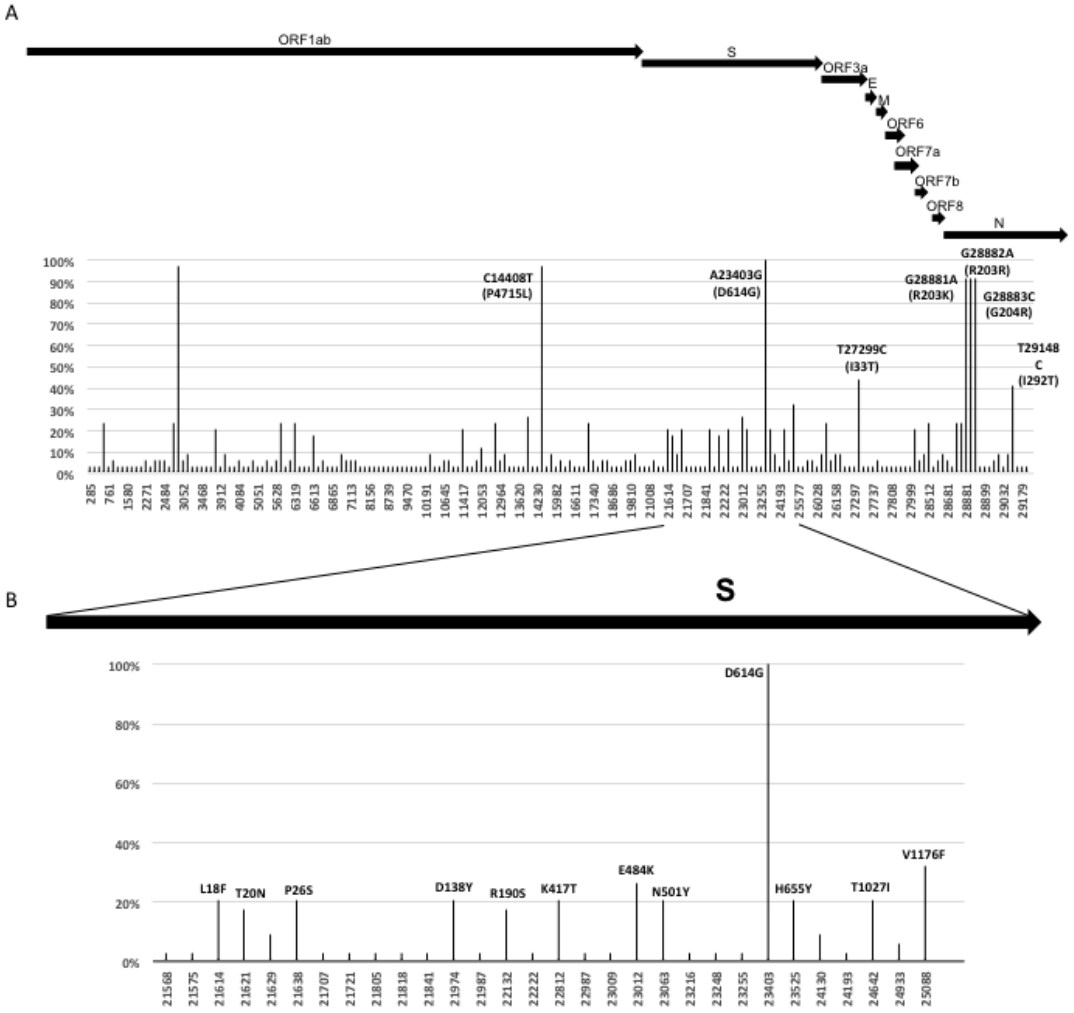


Figure 2. Mutations of SARS-CoV-2 genome sequences from Ceará state, Northeast Brazil. a) Frequency of SNVs per SARS-CoV-2 genome position among the 34 genome sequences (missense SNVs with prevalence >40% were represented). b) Frequency of SNVs in Spike protein (S) among the 34 genome sequences (SNVs with prevalence >17% were represented).

4. Discussion

In the present study, we showed the most prevalent circulating lineages and the distribution of SARS-CoV-2 mutations in Ceará state. Genome sequencing is an essential step to understand dispersion and to detect mutations. The first viral genomes sequenced available in GISAID from Ceará were classified as B.1 (Candido et al., 2020), and similar results have been reported for other Brazilian states (Botelho-Souza et al., 2021; dos Santos et al., 2021). Lineage B.1 was predominant worldwide, especially in Europe, and emerged around January 2020 (Rambaut et al., 2020), contributing to the early viral epidemic dynamics in Brazil (Candido et al., 2020). Within our dataset, B.1.1.33 was the most prevalent lineage in Ceará in 2020, which was in accordance with another study, also conducted in the Northeast region of

Brazil, that has shown a higher prevalence of B.1.1.33 lineage (dos Santos et al., 2021). In a different study, a low prevalence of B.1.1.33 lineage in Ceará in the first two months of the pandemic was reported (Resende et al., 2021a). However, divergences of results are expected due to sampling size bias. The results showed occurrence of lineage P.2 in April, with the highest frequency in December 2020. In Brazil, lineage P.2 was first identified in Rio de Janeiro, in October 2020 (Voloch et al., 2020), even though Lamarca et al. (2021) estimated that the origin of P.2 lineage took place in February 2020. From the results, lineage P.1 was detected in Ceará in mid-October, 2020. P.1 (VOC Gamma) was first detected in Manaus in November-December 2020, and was quickly found in other Brazilian states (Faria et al., 2021b). Recently, Lamarca et al. (2021) inferred that P.1 had its origin around August 2020, which is in accordance with our results. This suggests that lineage P.1 emerged and was not noticed earlier due to poor genomic surveillance in the country at that time. Our analysis identified genome mutations within a dataset comprising 34 samples collected from patients in Ceará, using as reference the genome of the first SARS-CoV-2 isolated in Wuhan, China. In the Spike protein, the mutation D614G was most prevalent, with 100% of frequency within the dataset. D614G was first detected in January 2020, in samples from China and Germany, but quickly became the dominant genotype throughout the world (Korber et al., 2020; Yurkovetskiy et al., 2020). Moreover, D614G has been associated with lower Cq values in infected patients, possibly indicating a higher upper respiratory tract viral load. Despite that, this mutation was not associated with increased disease severity or case fatality rates (Korber et al., 2020). D614G has been previously associated with ORF1ab P4715L, with a strong allelic association, therefore, they possibly confer a fitness gain (Yang et al., 2020). Indeed, these mutations had a high prevalence among the sequenced genomes, and this may confer increased transmissibly of the virus. Moreover, mutations P4715L and D614G were detected in South America, including Brazil, and have been correlated with higher mortality rates (Fang et al., 2021; Toyoshima et al., 2020). Among our samples, we also detected mutations that were located in the Spike protein that have been associated with a higher infectivity and evasion of immune system, such as K417T, E484K and N501Y (Harvey et al., 2021; Khan et al., 2021). E484K has also been detected in lineage P.2 (Voloch et al., 2020), and in one of the genomes classified as B.1.1.33 from July 2020, within the dataset. Resende et al. (2021b) reported a new SARS-CoV-2 considered as Variant of Interest (VOI), N.9, that probably emerged in August 2020, that descended from lineage B.1.133. However, according to the authors, VOI N.9 has other three mutations that were not present in our sample. Some Clade 2 signature mutations, which showed a wide spread in Brazil (Candido et al., 2020), were also

found in the dataset: I33T, located in ORF6, and I292T, in the N gene. Structural proteins encoded for the two regions are involved in the degradation of interferon-induced antiviral proteins (Li et al., 2020). Finally, other important mutations found with a higher prevalence among our dataset were R203K and G204R, located in the N gene. The occurrence of these SNVs was reported by Laamarti et al. (2020), who detected them in samples from all continents except Africa and Asia. Our results highlight the importance of genomic surveillance as a tool for monitoring and understanding the evolution of SARS-CoV-2 and other viruses. Furthermore, this work fills a gap in the knowledge about SARS-CoV-2 as it reports the early imports of lineages and the prevalence of mutations in the state of Ceará in the year when COVID-19 struck Brazil.

5. Conclusions

We have showed that the lineage B.1.1.33 was the most prevalent in early epidemic phase in Ceará state in 2020. The mutations reported in this present study also brought up the genetic diversity of SARS-CoV-2 variants and provided evidence associated with higher transmissibility and disease severity. Our results confirm the need to sustain continuous genomic surveillance through SARS-CoV-2 sequencing in order to identify circulating lineages and to monitor the pandemic.

6. Authorship contribution statement

Francisca Andréa da Silva Oliveira: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft. **Maísa Viana de Holanda:** Conceptualization, Methodology, Writing – original draft. **Luína Benevides Lima:** Conceptualization, Formal analysis, Writing – original draft. **Mariana Brito Dantas:** Conceptualization, Methodology. **Igor Oliveira Duarte:** Formal analysis, Investigation. **Luzia Gabrielle Zeferino de Castro:** Formal analysis, Investigation. **Laís Lacerda Brasil de Oliveira**: Methodology, Investigation. **Carlos Roberto Koscky Paier**: Methodology, Investigation. **Caroline de Fátima Aquino Moreira-Nunes**: Conceptualization, Investigation. **Nicholas Costa Barroso Lima**: Conceptualization, Formal analysis, Investigation, Writing – original draft. **Vânia Maria Maciel Melo:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. **Raquel Carvalho Montenegro:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was financially supported by The Brazilian Ministry of Education (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – 88887.512361/2020-00) by CAPES-EPIDEMIAS-Edital de Seleção Emergencial-Prevenção e Combate a Surtos, Epidemias e Pandemia provided a scholarship to FASO.

Funding: This study was supported by Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP – grant number 03195011/2020); Coordenação de Aperfeiçoamento de Pessoal de Nível (CAPES 23038.001015/2020-15).

Supplementary materials

Supplementary Table S1 – SARS-CoV-2 genome sequences from the state of Ceará, Northeast Brazil, collected until June 2021

Supplementary Table S2 – Number of SARS-CoV-2 genome sequences used in the analysis from March 2020 to December 2020

Supplementary Table S3 – Prevalence of SNVs identified in the 34 genomes sequenced

Supplementary File S1 – SARS-CoV-2 genome sequences available from GISAID (accession numbers)

Reference

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., Schatz, M.C., 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20, 1–17. https://doi.org/10.1186/s13059-019-1829-6

Andrews, S., 2010. A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Botelho-Souza, L.F., Nogueira-Lima, F.S., Roca, T.P., Naveca, F.G., de Oliveria dos Santos, A., Maia, A.C.S., da Silva, C.C., de Melo Mendonça, A.L.F., Lugtenburg, C.A.B., Azzi, C.F.G., Fontes, J.L.F., Cavalcante, S., de Cássia Pontello Rampazzo, R., Santos, C.H.N., Di Sabatino Guimarães, A.P., Máximo, F.R., Villalobos-Salcedo, J.M., Vieira, D.S., 2021. SARS-CoV-2 genomic surveillance in Rondônia, Brazilian Western Amazon. Sci. Rep. 11, 1–12. https://doi.org/10.1038/s41598-021-83203-2

Candido, D.S., Claro, I.M., de Jesus, J.G., Souza, W.M., Moreira, F.R.R., Dellicour, S., Mellan, T.A., du Plessis, L., Pereira, R.H.M., Sales, F.C.S., Manuli, E.R., Thézé, J., Almeida, L., Menezes, M.T., Voloch, C.M., Fumagalli, M.J., Coletti, T.M., da Silva, C.A.M., Ramundo, M.S., Amorim, M.R., Hoeltgebaum, H.H., Mishra, S., Gill, M.S., Carvalho, L.M., Buss, L.F., Prete, C.A., Ashworth, J., Nakaya, H.I., Peixoto, P.S., Brady, O.J., Nicholls, S.M., Tanuri, A., Rossi, Á.D., Braga, C.K.V., Gerber, A.L., de Guimarães, A.P.C., Gaburo, N., Alencar, C.S., Ferreira, A.C.S., Lima, C.X., Levi, J.E., Granato, C., Ferreira, G.M., Francisco, R.S., Granja, F., Garcia, M.T., Moretti, M.L., Perroud, M.W., Castiñeiras, T.M.P.P., Lazari, C.S., Hill, S.C., de Souza Santos, A.A., Simeoni, C.L., Forato, J., Sposito, A.C., Schreiber, A.Z., Santos, M.N.N., de Sá, C.Z., Souza, R.P., Resende-Moreira, L.C., Teixeira, M.M., Hubner, J., Leme, P.A.F., Moreira, R.G., Nogueira, M.L., Ferguson, N.M., Costa, S.F., Proenca-Modena, J.L., Vasconcelos, A.T.R., Bhatt, S., Lemey, P., Wu, C.H., Rambaut, A., Loman, N.J., Aguiar, R.S., Pybus, O.G., Sabino, E.C., Faria, N.R., 2020. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science (80-.). 369, 1255–1260. https://doi.org/10.1126/science.abd2161

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X., 2012a. Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front. Genet. 3, 1–9. https://doi.org/10.3389/fgene.2012.00035

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6, 80–92. https://doi.org/10.4161/fly.19695

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10, 1–4. https://doi.org/10.1093/gigascience/giab008

de Jesus, J.G., Sacchi, C., Candido, Sd. da S., Claro, I.M., Sales, F.C.S., Manuli, E.R., Silva, D.B.B. da, Paiva, T.M. De, Pinho, M.A.B., Santos, K.C. de O., Hill, S.C., Aguiar, R.S., Romero, F., Santos, F.C.P. dos, Gonçalves, C.R., Timenetsky, M. do C., Quick, J., Croda, J.H.R., Oliveira, W. de, Rambaut, A., Pybus, O.G., Loman, N.J., Sabino, E.C., Faria, N.R., 2020. Importation and early local transmission of COVID-19 in Brazil, 2020. Rev. Inst. Med. Trop. Sao Paulo 1–5.

dos Santos, C.A., Bezerra, G.V.B., de Azevedo Marinho, A.R.R.A., Alves, J.C., Tanajura, D.M., Martins-Filho, P.R., 2021. SARS-CoV-2 genomic surveillance in Northeast Brazil: timing of emergence of the Brazilian variant of concern P1. J. Travel Med. 28, 1–3. https://doi.org/10.1093/jtm/taab066

Fang, S., Liu, S., Shen, J., Lu, A.Z., Wang, A.K.Y., Zhang, Y., Li, K., Liu, J., Yang, L., Hu, C.D., Wan, J., 2021. Updated SARS-CoV-2 single nucleotide variants and mortality association. J. Med. Virol. 93, 6525–6534. https://doi.org/10.1002/jmv.27191

Faria, N.R., Claro, I.M., Candido, D., Franco, L.A.M., Andrade, P.S., Coletti, T.M., Silva, C.A.M., Sales, F.C., Manulli, E.R., Aguiar, R.S., Gaburo, N., Camilo, C. da C., Fraiji, N.A., Crispim, M.A.E., Carvalho, M.P.S.S., Rambaut, A., Loman, N., Phybus, O.G., Sabino, E.C., 2021a. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological.Org 1–9.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulswit, Ruben, J.G., Franco, L.A.M., Raimundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., da Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Jr Gaburo, N., Camilo, C. da C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., Souza, L.M. de, Pinho, M.C. de, Araujo, L.J.T., Malta, F.S. V, Lima, A.B. de, Silva, J. do P., Zauli, D.A.., Ferreira, A.C.S., Schnekenberg, R.P., Laydon, D.J., Walker, Patrick, G.T., Schluter, H.M., Santos, A.L.P. dos, Vidal, M.S., Caro, V.S. Del, Filho, R.M.F., Santos, H.M. dos, Aguiar, R.S., Proenca-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Carlos, P.J., Nascimento, V.H., A, S.M., Bowden, T.A., Pond, S.L.K., Wu, C.-H., Ratmann, O., Ferguson, N.M., Christopher, D., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M. do P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021b. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science (80-.). 372, 815–821.

Fung, T.S., Liu, D.X., 2019. Human Coronavirus : Host-Pathogen Interaction. Annu. Rev. Microbiol. 73, 529–557.

Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat. Rev. Microbiol. 19, 409–424. https://doi.org/10.1038/s41579-021-00573-0

Khan, A., Zia, T., Suleman, M., Khan, T., Ali, S.S., Abbasi, A.A., Mohammad, A., Wei, D.Q., 2021. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. J. Cell. Physiol. 236, 7045–7057. https://doi.org/10.1002/jcp.30367

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Angyal, A., Brown, R.L., Carrilero, L., Green, L.R., Groves, D.C., Johnson, K.J., Keeley, A.J., Lindsey, B.B., Parsons, P.J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R.M., Wang, D., Wyles, M.D., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182, 812–827. https://doi.org/10.1016/j.cell.2020.06.043

Laamarti, Meriem, Alouane, T., Kartti, S., Chemao-Elfihri, M.W., Hakmi, M., Essabbar, A., Laamarti, Mohamed, Hlali, H., Bendani, H., Boumajdi, N., Benhrif, O., Allam, L., Hafidi, N. El, Jaoudi, R. El, Allali, I., Marchoudi, N., Fekkak, J., Benrahma, H., Nejjari, C., Amzazi, S., Belyamani, L., Ibrahimi, A., 2020. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geodistribution and a rich genetic variations of hotspots mutations. PLoS One 15, 1–18. https://doi.org/10.1371/journal.pone.0240345

Lamarca, A.P., de Almeida, L.G.P., Francisco, R. da S., Lima, L.F.A., Scortecci, K.C., Perez, V.P., Brustolini, O.J., Sousa, E.S.S., Secco, D.A., Santos, A.M.G., Albuquerque, G.R., Mariano, A.P.M., Maciel, B.M., Gerber, A.L., Guimarães, A.P. de C., Nascimento, P.R., Neto, F.P.F., Gadelha, S.R., Porto, L.C., Campana, E.H., Jeronimo, S.M.B., Vasconcelos, A.T.R., 2021. Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil. PLoS Negl.

Trop. Dis. 15, e0009835. https://doi.org/10.1371/journal.pntd.0009835

Lauring, A.S., Hodcroft, E.B., 2021. Genetic Variants of SARS-CoV-2 - What Do They Mean? JAMA - J. Am. Med. Assoc. 325, 529–531. https://doi.org/10.1001/jama.2020.27124

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J.Y., Liao, C.H., Wang, Q., Tan, Y.J., Luo, R., Qiu, Y., Ge, X.Y., 2020. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Res. 286, 198074. https://doi.org/10.1016/j.virusres.2020.198074

Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407. https://doi.org/10.1038/s41564-020-0770-5

Resende, P.C., Delatorre, E., Gräf, T., Mir, D., Motta, F.C., Appolinario, L.R., Paixão, A.C.D. da, Mendonça, A.C. da F., Ogrzewalska, M., Caetano, B., Wallau, G.L., Docena, C., Santos, M.C. dos, de Almeida Ferreira, J., Sousa Junior, E.C., Silva, S.P. da, Fernandes, S.B., Vianna, L.A., Souza, L. da C., Ferro, J.F.G., Nardy, V.B., Santos, C.A., Riediger, I., do Carmo Debur, M., Croda, J., Oliveira, W.K., Abreu, A., Bello, G., Siqueira, M.M., 2021a. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. Front. Microbiol. 11, 1–14. https://doi.org/10.3389/fmicb.2020.615280

Resende, P.C., Gräf, T., Paixão, A.C.D., Appolinario, L., Lopes, R.S., Mendonça, A.C. da F., da Rocha, A.S.B., Motta, F.C., Neto, L.G.L., Khouri, R., de Oliveira, C.I., Santos-Muccillo, P., Bezerra, J.F., Teixeira, D.L.F., Riediger, I., Debur, M.D.C., Ribeiro-Rodrigues, R., Leite, A.B., Do Santos, C.A., Gregianini, T.S., Fernandes, S.B., Bernardes, A.F.L., Cavalcanti, A.C., Miyajima, F., Sachhi, C., Mattos, T., da Costa, C.F., Delatorre, E., Wallau, G.L., Naveca, F.G., Bello, G., Siqueira, M.M., 2021b. 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 13, 1–7. https://doi.org/10.3390/v13050724

Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., Kosugi, Y., Shirakawa, K., Sadamasu, K., Kimura, I., Ito, J., Wu, J., Iwatsuki-Horimoto, K., Ito, M., Yamayoshi, S., Loeber, S., Tsuda, M., Wang, L., Ozono, S., Butlertanaka, E.P., Tanaka, Y.L., Shimizu, R., Shimizu, K., Yoshimatsu, K., Kawabata, R., Sakaguchi, T., Tokunaga, K., Yoshida, I., Asakura, H., Nagashima, M., Kazuma, Y., Nomura, R., Horisawa, Y., Yoshimura, K., Takaori-Kondo, A., Imai, M., Chiba, M., Furihata, H., Hasebe, H., Kitazato, K., Kubo, H., Misawa, N., Morizako, N., Noda, K., Oide, A., Suganami, M., Takahashi, M., Tsushima, K., Yokoyama, M., Yuan, Y., Tanaka, S., Nakagawa, S., Ikeda, T., Fukuhara, T., Kawaoka, Y., Sato, K., 2021. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. Nature 602. https://doi.org/10.1038/s41586-021-04266-9

Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: Strategic k-mer extension for scrupulous assemblies. Genome Biol. 19, 1–13. https://doi.org/10.1186/s13059-018-1540-z

Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., Kiyotani, K., 2020. SARS-CoV-2

genomic variations associated with mortality rate of COVID-19. J. Hum. Genet. 65, 1075–1082. https://doi.org/10.1038/s10038-020-0808-9

Van der Auwera, G., O'Connor, B., 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.

Voloch, C.M., Jr, R.S.F., Almeida, L.G.P., Cardoso, C.C., Brustolini, O.J., Gerber, A.L., de Guimarães, A.P.C., Mariani, D., da Costa, R.M., 2020. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. medRxiv.

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. Lancet 395, 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9

Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., Graham, B.S., Mascola, J.R., Chang, J.Y., Yin, M.T., Sobieszczyk, M., Kyratsous, C.A., Shapiro, L., Sheng, Z., Huang, Y., Ho, D.D., 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593, 130–135. https://doi.org/10.1038/s41586-021-03398-2

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3

Yang, H.C., Chen, C.H., Wang, J.H., Liao, H.C., Yang, C.T., Chen, C.W., Lin, Y.C., Kao, C.H., Lu, M.Y.J., Liao, J.C., 2020. Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. Proc. Natl. Acad. Sci. U. S. A. 117, 30679–30686. https://doi.org/10.1073/pnas.2007840117

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S.F., Lemieux, J.E., Munro, J.B., Rafique, A., Barve, A., Sabeti, P.C., Kyratsous, C.A., Dudkina, N. V., Shen, K., Luban, J., 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell 183, 739–751. https://doi.org/10.1016/j.cell.2020.09.032

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733. https://doi.org/10.1056/nejmoa2001017

Suppl. Table S1. SARS-CoV-2 genome sequences from Ceará state, Northeast Brazil.

Laboratory ID	Cq value	Collection Month/year	Pangolin Lineage	GISAID
FARMAGEN_A02	13	jul/20	B.1.1.33	EPI_ISL_10909960
FARMAGEN_A74	21	jul/20	B.1.212	EPI_ISL_10909993
FARMAGEN_A07	13	jul/20	B.1.1.33	EPI_ISL_10909961
FARMAGEN_A08	11	jul/20	B.1.1.28	EPI_ISL_10909962
FARMAGEN_A16	12	jul/20	B.1.1	EPI_ISL_10909963
FARMAGEN_A20	14	jul/20	B.1.212	EPI_ISL_10909964
FARMAGEN_A21	19	ago/20	B.1.1.33	EPI_ISL_10909965
FARMAGEN_A27	18	ago/20	B.1	EPI_ISL_10909968
FARMAGEN_A24	14	ago/20	B.1.1	EPI_ISL_10909966
FARMAGEN_A25	24	set/20	B.1.1.33	EPI_ISL_10909967
FARMAGEN_A30	24	set/20	B.1.1	EPI_ISL_10909969
FARMAGEN_A36	13	set/20	B.1.1.33	EPI_ISL_10909970
FARMAGEN_A37	17	set/20	B.1.212	EPI_ISL_10909971
FARMAGEN_A39	15	out/20	B.1.1.33	EPI_ISL_10909972
FARMAGEN_A41	15	out/20	P.1	EPI_ISL_10909973
FARMAGEN_A44	18	out/20	B.1.1.33	EPI_ISL_10909974
FARMAGEN_A45	16	out/20	B.1.1.33	EPI_ISL_10909975
FARMAGEN_A47	15	nov/20	B.1.1.33	EPI_ISL_10909976
FARMAGEN_A49	20	nov/20	B.1.1.33	EPI_ISL_10909977
FARMAGEN_A52	17	nov/20	B.1.1.33	EPI_ISL_10909978
FARMAGEN_A53	20	nov/20	B.1.1.33	EPI_ISL_10909979
FARMAGEN_A55	12	nov/20	B.1.1.33	EPI_ISL_10909980
FARMAGEN_A56	14	nov/20	B.1.1.33	EPI_ISL_10909981

FARMAGEN_A57	18	nov/20	B.1.1.28	EPI_ISL_10909982
FARMAGEN_A58	13	dez/20	P.2	EPI_ISL_10909983
FARMAGEN_A60	14	jan/21	P.2	EPI_ISL_10909984
FARMAGEN_A62	20	fev/21	P.1	EPI_ISL_10909985
FARMAGEN_A63	15	fev/21	P.1	EPI_ISL_10909986
FARMAGEN_A66	19	fev/21	B.1.1.33	EPI_ISL_10909989
FARMAGEN_A64	17	fev/21	P.1	EPI_ISL_10909987
FARMAGEN_A65	16	mar/21	P.1	EPI_ISL_10909988
FARMAGEN_A68	15	mar/21	P.1	EPI_ISL_10909990
FARMAGEN_A69	19	mai/21	P.1	EPI_ISL_10909991
FARMAGEN_A70	14	jun/21	P.1	EPI_ISL_10909992

Suppl. Table S2. Number of SARS-CoV-2 genome sequences used to analysis.

Pangolin Lineage	Genomes from GISAID	Present study	Total	Collection Month
B.1	19	-	27	mar/20
B.1.1.33	1	-		
B.1.212	7	-		
B.1	1	-	14	abr/20
B.1.1.33	9	-		
B.1.1	1	-		
B.1.212	2	-		
P.2	1	-		
B.1.1.33	2	-	9	mai/20
B.1.212	7	-		
B.1.1.33	2	-	2	jun/20
B.1.1	-	1	6	jul/20
B.1.1.33	-	2		
B.1.1.28	-	1		
B.1.212	-	2		
B.1	-	1	4	ago/20
B.1.1	-	1		
B.1.1.33	1	1		
B.1.1	-	1	5	set/20
B.1.1.33	1	2		
B.1.212	-	1		
B.1.1.33	-	3	4	out/20
P.1	-	1		
B.1.1.33	7	6	19	nov/20
B.1.1.28	3	1		
P.2	2	-		
B.1.1	1	-	25	dez/20
B.1.1.28	4	-		
B.1.1.33	8	-		
B.1.1.371	1	-		
P.2	10	1		

Suppl. Table S3. Prevalence of SNVs identified in the 34 genomes.

Gene	Position	SNV	Protein annotation	Mutation	Number of genome sequences with mutation	Prevalence
ORF1ab	285	c.20G>T	p.G7V	Missense variant	1	2,94%
	367	c.102C>T	p.S34S	Synonymous variant	1	2,94%
	617	c.352T>C	p.Y118H	Missense variant	1	2,94%
	733	c.468T>C	p.D156D	Synonymous variant	8	23,53%
	761	c.496A>G	p.S166G	Missense variant	1	2,94%
	1230	c.965A>G	p.K322R	Missense variant	2	5,88%
	1263	c.998C>T	p.T333M	Missense variant	1	2,94%
	1363	c.1098T>C	p.V366V	Synonymous variant	1	2,94%
	1580	c.1315G>A	p.V439I	Missense variant	1	2,94%
	1742	c.1477G>A	p.G493S	Missense variant	1	2,94%
	1915	c.1650C>A	p.R550R	Synonymous variant	1	2,94%
	1929	c.1664C>T	p.A555V	Missense variant	1	2,94%
	2271	c.2006A>G	p.K669R	Missense variant	2	5,88%
	2335	c.2070T>C	p.A690A	Synonymous variant	1	2,94%
	2388	c.2123C>T	p.T708I	Missense variant	2	5,88%
	2469	c.2204C>T	p.A735V	Missense variant	2	5,88%
	2484	c.2219T>C	p.I740T	Missense variant	2	5,88%
	2638	c.2373C>T	p.L791L	Synonymous variant	1	2,94%
	2749	c.2484C>T	p.D828D	Synonymous variant	8	23,53%
	3037	c.2772C>T	p.F924F	Synonymous variant	33	97,06%
	3052	c.2787G>T	p.E929D	Missense variant	2	5,88%
	3096	c.2831C>T	p.S944L	Missense variant	3	8,82%
	3340	c.3075G>T	p.V1025V	Synonymous	1	2,94%

				variant		
	3392	c.3127G>A	p.A1043T	Missense variant	1	2,94%
	3468	c.3203G>A	p.G1068E	Missense variant	1	2,94%
	3777	c.3512C>T	p.T1171I	Missense variant	1	2,94%
	3784	c.3519C>T	p.V1173V	Synonymous variant	1	2,94%
	3828	c.3563C>T	p.S1188L	Missense variant	7	20,59%
	3912	c.3647C>T	p.T1216I	Missense variant	1	2,94%
	3954	c.3689A>G	p.K1230R	Missense variant	3	8,82%
	3966	c.3701C>T	p.A1234V	Missense variant	1	2,94%
	3994	c.3729G>T	p.L1243L	Synonymous variant	1	2,94%
	4084	c.3819C>T	p.D1273D	Synonymous variant	2	5,88%
	4192	c.3927T>C	p.T1309T	Synonymous variant	1	2,94%
	4420	c.4155A>G	p.T1385T	Synonymous variant	1	2,94%
	4683	c.4418C>T	p.A1473V	Missense variant	2	5,88%
	5051	c.4786C>T	p.P1596S	Missense variant	1	2,94%
	5065	c.4800T>A	p.D1600E	Missense variant	1	2,94%
	5192	c.4927C>T	p.L1643L	Synonymous variant	2	5,88%
	5206	c.4941G>T	p.M1647I	Missense variant	1	2,94%
	5628	c.5363C>T	p.T1788M	Missense variant	2	5,88%
	5648	c.5383A>C	p.K1795Q	Missense variant	8	23,53%
	5972	c.5707G>T	p.D1903Y	Missense variant	1	2,94%
	6255	c.5990C>T	p.A1997V	Missense variant	2	5,88%
	6319	c.6054A>G	p.P2018P	Synonymous variant	8	23,53%
	6401	c.6136C>A	p.P2046T	Missense variant	1	2,94%
	6449	c.6184C>T	p.L2062F	Missense variant	1	2,94%

	6541	c.6276C>T	p.H2092H	Synonymous variant	1	2,94%
	6613	c.6348A>G	p.V2116V	Synonymous variant	6	17,65%
	6673	c.6408T>C	p.D2136D	Synonymous variant	1	2,94%
	6723	c.6458C>T	p.T2153I	Missense variant	2	5,88%
	6753	c.6488G>T	p.R2163L	Missense variant	1	2,94%
	6865	c.6600G>T	p.K2200N	Missense variant	1	2,94%
	7042	c.6777G>T	p.M2259I	Missense variant	1	2,94%
	7043	c.6778C>A	p.P2260T	Missense variant	3	8,82%
	7087	c.6822T>C	p.T2274T	Synonymous variant	2	5,88%
	7113	c.6848C>T	p.T2283I	Missense variant	2	5,88%
	7164	c.6899C>T	p.T2300I	Missense variant	2	5,88%
	7761	c.7496G>A	p.G2499D	Missense variant	1	2,94%
	8017	c.7752G>T	p.A2584A	Synonymous variant	1	2,94%
	8156	c.7891T>C	p.S2631P	Missense variant	1	2,94%
	8305	c.8040T>A	p.V2680V	Synonymous variant	1	2,94%
	8344	c.8079C>T	p.D2693D	Synonymous variant	1	2,94%
	8692	c.8427C>T	p.Y2809Y	Synonymous variant	1	2,94%
	8739	c.8474C>A	p.T2825N	Missense variant	1	2,94%
	9203	c.8938G>A	p.D2980N	Missense variant	1	2,94%
	9246	c.8981C>T	p.A2994V	Missense variant	1	2,94%
	9430	c.9165C>T	p.I3055I	Synonymous variant	1	2,94%
	9470	c.9205A>C	p.R3069R	Synonymous variant	1	2,94%
	9661	c.9396C>T	p.F3132F	Synonymous variant	1	2,94%
	9967	c.9702C>T	p.L3234L	Synonymous variant	1	2,94%
	10116	c.9851C>T	p.T3284I	Missense	1	2,94%

				variant		
	10191	c.9926C>A	p.S3309Y	Missense variant	1	2,94%
	10323	c.10058A>G	p.K3353R	Missense variant	3	8,82%
	10369	c.10104C>T	p.R3368R	Synonymous variant	1	2,94%
	10507	c.10242C>T	p.N3414N	Synonymous variant	1	2,94%
	10645	c.10380C>T	p.D3460D	Synonymous variant	2	5,88%
	10667	c.10402T>G	p.L3468V	Missense variant	2	5,88%
	10833	c.10568C>T	p.A3523V	Missense variant	1	2,94%
	10969	c.10704C>T	p.F3568F	Synonymous variant	1	2,94%
	11417	c.11152G>T	p.V3718F	Missense variant	7	20,59%
	11516	c.11251G>A	p.V3751I	Missense variant	1	2,94%
	11595	c.11330A>G	p.Q3777R	Missense variant	1	2,94%
	11824	c.11559C>T	p.I3853I	Synonymous variant	2	5,88%
	12053	c.11788C>T	p.L3930F	Missense variant	4	11,76%
	12067	c.11802G>T	p.M3934I	Missense variant	1	2,94%
	12295	c.12030C>T	p.T4010T	Synonymous variant	1	2,94%
	12778	c.12513C>T	p.Y4171Y	Synonymous variant	8	23,53%
	12964	c.12699A>G	p.G4233G	Synonymous variant	2	5,88%
	13396	c.13131A>C	p.K4377N	Missense variant	3	8,82%
	13459	c.13194G>T	p.S4398S	Synonymous variant	1	2,94%
	13514	c.13250G>A	p.G4417D	Missense variant	1	2,94%
	13620	c.13356C>T	p.D4452D	Synonymous variant	1	2,94%
	13712	c.13448A>G	p.K4483R	Missense variant	1	2,94%
	13860	c.13596C>T	p.D4532D	Synonymous variant	9	26,47%
	13887	c.13623C>T	p.Y4541Y	Synonymous variant	1	2,94%

	14230	c.13966C>T	p.P4656S	Missense variant	1	2,94%
	14408	c.14144C>T	p.P4715L	Missense variant	33	97,06%
	14592	c.14328T>A	p.S4776S	Synonymous variant	1	2,94%
	15324	c.15060C>T	p.N5020N	Synonymous variant	3	8,82%
	15982	c.15718G>A	p.V5240I	Missense variant	1	2,94%
	16329	c.16065C>T	p.Y5355Y	Synonymous variant	2	5,88%
	16353	c.16089T>C	p.H5363H	Synonymous variant	1	2,94%
	16428	c.16164C>T	p.Y5388Y	Synonymous variant	2	5,88%
	16611	c.16164C>T	p.Y5388Y	Synonymous variant	1	2,94%
	16733	c.16469C>T	p.S5490L	Missense variant	1	2,94%
	17155	c.16891A>G	p.T5631A	Missense variant	1	2,94%
	17259	c.16995G>T	p.E5665D	Missense variant	8	23,53%
	17340	c.17076A>G	p.A5692A	Synonymous variant	2	5,88%
	17518	c.17254C>T	p.L5752F	Missense variant	1	2,94%
	18225	c.17961G>T	p.M5987I	Synonymous variant	2	5,88%
	18508	c.18244C>T	p.L6082F	Missense variant	2	5,88%
	18686	c.18422G>T	p.C6141F	Missense variant	1	2,94%
	18803	c.18539G>T	p.S6180I	Missense variant	1	2,94%
	18981	c.18717C>T	p.H6239H	Synonymous variant	1	2,94%
	19185	c.18921C>T	p.C6307C	Synonymous variant	2	5,88%
	19810	c.19546A>G	p.K6516E	Missense variant	2	5,88%
	19983	c.19719C>T	p.V6573V	Missense variant	3	8,82%
	20157	c.19893T>C	p.Y6631Y	Synonymous variant	1	2,94%
	20318	c.20054A>C	p.E6685A	Missense variant	1	2,94%
	21008	c.20744C>T	p.T6915I	Missense	1	2,94%

				variant		
	21024	c.20760T>C	p.N6920N	Synonymous variant	2	5,88%
S	21568	c.6T>A	p.F2L	Missense variant	1	2,94%
	21575	c.13C>T	p.L5F	Missense variant	1	2,94%
	21614	c.52C>T	p.L18F	Missense variant	7	20,59%
	21621	c.59C>A	p.T20N	Missense variant	6	17,65%
	21629	c.67C>A	p.Q23K	Missense variant	3	8,82%
	21638	c.76C>T	p.P26S	Missense variant	7	20,59%
	21707	c.145C>T	p.H49Y	Missense variant	1	2,94%
	21721	c.159C>T	p.D53D	Synonymous variant	1	2,94%
	21805	c.243C>T	p.N81N	Synonymous variant	1	2,94%
	21818	c.256T>A	p.F86I	Missense variant	1	2,94%
	21841	c.279T>G	p.A93A	Synonymous variant	1	2,94%
	21974	c.412G>T	p.D138Y	Missense variant	7	20,59%
	21987	c.425G>A	p.G142D	Missense variant	1	2,94%
	22132	c.570G>T	p.R190S	Missense variant	6	17,65%
	22222	c.660T>G	p.F220L	Missense variant	1	2,94%
	22812	c.1250A>C	p.K417T	Missense variant	7	20,59%
	22987	c.1425C>T	p.A475A	Synonymous variant	1	2,94%
	23009	c.1447G>A	p.V483I	Missense variant	1	2,94%
	23012	c.1450G>A	p.E484K	Missense variant	9	26,47%
	23063	c.1501A>T	p.N501Y	Missense variant	7	20,59%
	23216	c.1654C>A	p.L552I	Missense variant	1	2,94%
	23248	c.1686C>T	p.F562F	Synonymous variant	1	2,94%
	23255	c.1693T>C	p.F565L	Missense variant	1	2,94%

	23403	c.1841A>G	p.D614G	Missense variant	34	100,00%
	23525	c.1963C>T	p.H655Y	Missense variant	7	20,59%
	24130	c.2568C>T	p.N856N	Synonymous variant	3	8,82%
	24193	c.2631G>T	p.L877L	Synonymous variant	1	2,94%
	24642	c.3080C>T	p.T1027I	Missense variant	7	20,59%
	24933	c.3371G>T	p.G1124V	Missense variant	2	5,88%
	25088	c.3526G>T	p.V1176F	Missense variant	11	32,35%
ORF3a	25577	c.185T>C	p.I62T	Missense variant	1	2,94%
	25618	c.226G>A	p.G76S	Missense variant	1	2,94%
	25793	c.401G>T	p.R134L	Missense variant	2	5,88%
	25904	c.512C>T	p.S171L	Missense variant	2	5,88%
	26028	c.636C>T	p.Y212Y	Synonymous variant	1	2,94%
	26061	c.669T>A	p.T223T	Synonymous variant	3	8,82%
	26149	c.757T>C	p.S253P	Missense variant	8	23,53%
	26152	c.760G>A	p.G254R	Missense variant	2	5,88%
	26158	c.766G>T	p.V256F	Missense variant	3	8,82%
	26162	c.770A>T	p.N257I	Missense variant	3	8,82%
E	26461	c.217C>T	p.L73F	Missense variant	1	2,94%
M	26681	c.159C>T	p.F53F	Synonymous variant	1	2,94%
ORF6	27297	c.96C>T	p.I32I	Synonymous variant	1	2,94%
	27299	c.98T>C	p.I33T	Missense variant	15	44,12%
ORF7a	27603	c.210C>T	p.G70G	Synonymous variant	1	2,94%
	27673	c.280C>T	p.Q94*	stop_gained	1	2,94%
	27737	c.344C>T	p.T115I	Missense variant	1	2,94%
	27741	c.348C>T	p.L116L	Synonymous variant	2	5,88%

	27747	c.354A>T	p.R118S	Missense variant	1	2,94%
	27750	c.357G>T	p.K119N	Missense variant	1	2,94%
ORF7b	27808	c.53T>C	p.L18P	Missense variant	1	2,94%
	23153	c.260C>T	p.T87I	Missense variant	1	2,94%
	27916	c.23G>T	p.G8V	Missense variant	1	2,94%
	27917	c.24A>G	p.G8G	Synonymous variant	1	2,94%
	27999	c.106C>T	p.P36S	Missense variant	1	2,94%
	28167	c.274G>A	p.E92K	Missense variant	7	20,59%
ORF8	28253	c.360C>T	p.F120F	Synonymous variant	2	5,88%
	28447	c.174A>G	p.Q58Q	Synonymous variant	3	8,82%
	28512	c.239C>G	p.P80R	Missense variant	8	23,53%
	28618	c.345T>A	p.T115T	Synonymous variant	1	2,94%
	28628	c.355G>T	p.A119S	Missense variant	2	5,88%
	28867	c.594T>G	p.T198T	Synonymous variant	3	8,82%
	28681	c.408G>A	p.E136E	Synonymous variant	2	5,88%
	28868	c.595C>T	p.P199S	Missense variant	1	2,94%
N	28877	c.604A>T	p.S202C	Missense variant	8	23,53%
	28878	c.605G>C	p.S202T	Missense variant	8	23,53%
	28881	c.608G>A	p.R203K	Missense variant	31	91,18%
	28882	c.609G>A	p.R203R	Synonymous variant	31	91,18%
	28883	c.610G>C	p.G204R	Missense variant	31	91,18%
	28892	c.619C>T	p.P207S	Missense variant	1	2,94%
	28899	c.626G>T	p.R209I	Missense variant	1	2,94%
	28905	c.632C>T	p.A211V	Missense variant	1	2,94%
	28975	c.702G>T	p.M234I	Missense	2	5,88%

				variant		
	28979	c.706G>T	p.G236C	Missense variant	3	8,82%
	29032	c.759G>C	p.E253D	Missense variant	1	2,94%
	29140	c.867G>T	p.Q289H	Missense variant	3	8,82%
	29148	c.875T>C	p.I292T	Missense variant	14	41,18%
	29167	c.894C>T	p.Y298Y	Synonymous variant	1	2,94%
	29179	c.906G>T	p.P302P	Synonymous variant	1	2,94%
	29666	c.109C>T	p.L37F	Missense variant	1	2,94%

Suppl. File S1. SARS-CoV-2 genome sequences available from GISAID (números de acesso)

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_722139	DB Diagnosticos do Brasil	Laboratório de Parasitologia Médica - Instituto de Medicina Tropical - Universidade de São Paulo	Brazil-UK Centre for Arbovirus Discovery Diagnosis Genomics and Epidemiology (CADDE) Genomic Network - Instituto de Medicina Tropical
EPI_ISL_476182, EPI_ISL_476183, EPI_ISL_476184, EPI_ISL_476185, EPI_ISL_476186, EPI_ISL_476198, EPI_ISL_476199, EPI_ISL_476200, EPI_ISL_476212, EPI_ISL_476213, EPI_ISL_476214, EPI_ISL_476218, EPI_ISL_476219, EPI_ISL_476225, EPI_ISL_476226, EPI_ISL_476227, EPI_ISL_476228, EPI_ISL_476229, EPI_ISL_476232, EPI_ISL_476281, EPI_ISL_476283, EPI_ISL_476288, EPI_ISL_476289, EPI_ISL_476306, EPI_ISL_476307, EPI_ISL_476308, EPI_ISL_476331, EPI_ISL_476368, EPI_ISL_476369			
see above	DB Diagnósticos do Brasil	Instituto de Medicina Tropical da Univesidade de São Paulo	Camila Alves Maia da Silva; Darlan da Silva Candido; Erika Regina Manuli; Ester Sabino; Flavia Cristina da Silva Sales; Giula Magalhaes Ferreira; Jaqueline Goes de Jesus; Julien Theze; Mariana Severo Ramundo; Nuno Faria; Samples: Nelson Gaburo Jr; Sequencing: Ingra Morales Claro; Thais de Moura Coletti
EPI_ISL_470573	Hermes Pardini	Bioinformatics Laboratory / LNCC	Alexandra Gerber; Amilcar Tanuri; Ana Paula Guimarães; CADDE-group; Carolina Voloch; Ester Cerdeira Sabino; Filipe Romero; Ingra Morales Claro; Jaqueline Goes de Jesus; Laboratorio Hermes Pardini; Laboratorio Simile; Luiz Gonzaga Paula de Almeida; Mariane Talon; Nuno Rodrigues Faria; Renato Santana Aguiar e Ana Tereza Vasconcelos; Ronaldo da Silva Francisco Junior; Terezinha Marta Pereira; working group UFMG; working group UFRJ; Átila Duque Rossi
EPI_ISL_476217	Hospital da Clínicas da Faculdade de Medicina da Universidade de São Paulo	Instituto de Medicina Tropical da Univesidade de São Paulo	Camila Alves Maia da Silva; Carolina S. Lazar; Cecilia Salete Alencar; Darlan da Silva Candido; Erika Regina Manuli; Ester Sabino; Flavia Cristina da Silva Sales; Giula Magalhaes Ferreira; Jaqueline Goes de Jesus; Julien Theze; Mariana Severo Ramundo; Nuno Faria; Samples: Ingra Morales Claro; Sequencing: Ingra Morales Claro; Silvia F. Costa; Thais de Moura Coletti
EPI_ISL_918536, EPI_ISL_918540	LACEN - Laboratório Central de Saúde Pública do Ceara	Evandro Chagas Institute	A.M.; Barbagelata; E.C.; E.M.A.; Ferreira; J.A.; Junior; K.C.; L.C.; L.S.; M.C.; P.S.; Pinheiro; Santos; Silva; Sousa; Sousa Junior; W.D.C.; da Silva
EPI_ISL_1164995	LACEN - Laboratório Central de Saúde Pública do Ceará	Evandro Chagas Institute	A.M.; Barbagelata; E.C.; E.M.A.; Ferreira; J.A.; Junior; K.C.; L.C.; L.S.; M.C.; P.S.; Pinheiro; Santos; Silva; Sousa; Sousa Junior; W.D.C.; da Silva
EPI_ISL_2298745	Laboratório Central de Saúde Pública do Ceará	Coordenação Geral de Laboratórios de Saúde Pública (CGLAB/DAEVS/SVS/MS)	Vagner Fonseca; et al.
EPI_ISL_1181356, EPI_ISL_1181520, EPI_ISL_1181522, EPI_ISL_1181575, EPI_ISL_1181577, EPI_ISL_1181586, EPI_ISL_1181588, EPI_ISL_1181591, EPI_ISL_1181593, EPI_ISL_1181596, EPI_ISL_1181598, EPI_ISL_1181604, EPI_ISL_1181621, EPI_ISL_1181622			
see above	Oswaldo Cruz Foundation, FIOCRUZ - Ceara (Fiocruz-CE)	Laboratory of Respiratory Viruses and Measles, Oswaldo Cruz Institute, FIOCRUZ	Alice Sampaio Rocha; Ana Carolina Mendonca; Anna Carolina Paixao; Fabio Miyajima; Fernando Motta; Joaquim César do Nascimento Sousa Júnior; Luciana Appolinario; Marilda Siqueira on behalf of the Fiocruz COVID-19 Genomic Surveillance Network; Paola Resende; Renata Serrano Lopes; Thais de Oliveira Costa
EPI_ISL_2661874, EPI_ISL_2661875, EPI_ISL_2661876, EPI_ISL_2661877, EPI_ISL_2661878, EPI_ISL_2661879, EPI_ISL_2661880, EPI_ISL_2661881, EPI_ISL_2661882, EPI_ISL_2661883, EPI_ISL_2661884, EPI_ISL_2661885, EPI_ISL_2661886, EPI_ISL_2661887, EPI_ISL_2661888, EPI_ISL_2661889, EPI_ISL_2661890, EPI_ISL_2661891, EPI_ISL_2661892, EPI_ISL_2661893, EPI_ISL_2661894, EPI_ISL_2661895, EPI_ISL_2661896, EPI_ISL_2661897, EPI_ISL_2661898, EPI_ISL_2661899, EPI_ISL_2661900, EPI_ISL_2661901, EPI_ISL_2661902, EPI_ISL_2661903, EPI_ISL_2661904, EPI_ISL_2661905, EPI_ISL_2661906, EPI_ISL_2731444, EPI_ISL_2731445, EPI_ISL_2731446, EPI_ISL_2731447, EPI_ISL_2731448, EPI_ISL_2731449, EPI_ISL_2731450			
see above	Oswaldo Cruz Institute, FIOCRUZ/CE	Analytical Competence Molecular Epidemiology Lab/ACME, Oswaldo Cruz Foundation, Ceara (FIOCRUZ CE)	Alice Sampaio Rocha; Ana Carolina Mendonca; Anna Carolina Paixao; Elisa Cavalcante Pereira; Fabio Miyajima; Fernando Motta; Luciana Appolinario; Marilda Siqueira on behalf of the Fiocruz COVID-19 Genomic Surveillance Network; Paola Resende; Renata Serrano Lopes; Taina Venas

References

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., Schatz, M.C., 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20, 1–17. https://doi.org/10.1186/s13059-019-1829-6

Andrews, S., 2010. A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Botelho-Souza, L.F., Nogueira-Lima, F.S., Roca, T.P., Naveca, F.G., de Oliveria dos Santos, A., Maia, A.C.S., da Silva, C.C., de Melo Mendonça, A.L.F., Lugtenburg, C.A.B., Azzi, C.F.G., Fontes, J.L.F., Cavalcante, S., de Cássia Pontello Rampazzo, R., Santos, C.H.N., Di Sabatino Guimarães, A.P., Máximo, F.R., Villalobos-Salcedo, J.M., Vieira, D.S., 2021. SARS-CoV-2 genomic surveillance in Rondônia, Brazilian Western Amazon. Sci. Rep. 11, 1–12. https://doi.org/10.1038/s41598-021-83203-2

Candido, D.S., Claro, I.M., de Jesus, J.G., Souza, W.M., Moreira, F.R.R., Dellicour, S., Mellan, T.A., du Plessis, L., Pereira, R.H.M., Sales, F.C.S., Manuli, E.R., Thézé, J., Almeida, L., Menezes, M.T., Voloch, C.M., Fumagalli, M.J., Coletti, T.M., da Silva, C.A.M., Ramundo, M.S., Amorim, M.R., Hoeltgebaum, H.H., Mishra, S., Gill, M.S., Carvalho, L.M., Buss, L.F., Prete, C.A., Ashworth, J., Nakaya, H.I., Peixoto, P.S., Brady, O.J., Nicholls, S.M., Tanuri, A., Rossi, Á.D., Braga, C.K.V., Gerber, A.L., de Guimarães, A.P.C., Gaburo, N., Alencar, C.S., Ferreira, A.C.S., Lima, C.X., Levi, J.E., Granato, C., Ferreira, G.M., Francisco, R.S., Granja, F., Garcia, M.T., Moretti, M.L., Perroud, M.W., Castiñeiras, T.M.P.P., Lazari, C.S., Hill, S.C., de Souza Santos, A.A., Simeoni, C.L., Forato, J., Sposito, A.C., Schreiber, A.Z., Santos, M.N.N., de Sá, C.Z., Souza, R.P., Resende-Moreira, L.C., Teixeira, M.M., Hubner, J., Leme, P.A.F., Moreira, R.G., Nogueira, M.L., Ferguson, N.M., Costa, S.F., Proenca-Modena, J.L., Vasconcelos, A.T.R., Bhatt, S., Lemey, P., Wu, C.H., Rambaut, A., Loman, N.J., Aguiar, R.S., Pybus, O.G., Sabino, E.C., Faria, N.R., 2020. Evolution and epidemic spread of SARS-CoV-2 in Brazil. Science (80-.). 369, 1255–1260. https://doi.org/10.1126/science.abd2161

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X., 2012a. Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front. Genet. 3, 1–9. https://doi.org/10.3389/fgene.2012.00035

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6, 80–92. https://doi.org/10.4161/fly.19695

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10, 1–4. https://doi.org/10.1093/gigascience/giab008

de Jesus, J.G., Sacchi, C., Candido, Sd. da S., Claro, I.M., Sales, F.C.S., Manuli, E.R., Silva, D.B.B. da, Paiva, T.M. De, Pinho, M.A.B., Santos, K.C. de O., Hill, S.C., Aguiar, R.S., Romero, F., Santos, F.C.P. dos, Gonçalves, C.R., Timenetsky, M. do C., Quick, J., Croda, J.H.R., Oliveira, W. de, Rambaut, A., Pybus, O.G., Loman, N.J., Sabino, E.C., Faria, N.R., 2020. Importation and early local transmission of COVID-19 in Brazil, 2020. Rev. Inst. Med. Trop. Sao Paulo 1–5.

dos Santos, C.A., Bezerra, G.V.B., de Azevedo Marinho, A.R.R.A., Alves, J.C., Tanajura,

D.M., Martins-Filho, P.R., 2021. SARS-CoV-2 genomic surveillance in Northeast Brazil: timing of emergence of the Brazilian variant of concern P1. J. Travel Med. 28, 1–3. https://doi.org/10.1093/jtm/taab066

Emam, M., Oweda, M., Antunes, A., El-Hadidi, M., 2021. Positive selection as a key player for SARS-CoV-2 pathogenicity: Insights into ORF1ab, S and E genes. Virus Res. 302. https://doi.org/10.1016/j.virusres.2021.198472

Fang, S., Liu, S., Shen, J., Lu, A.Z., Wang, A.K.Y., Zhang, Y., Li, K., Liu, J., Yang, L., Hu, C.D., Wan, J., 2021. Updated SARS-CoV-2 single nucleotide variants and mortality association. J. Med. Virol. 93, 6525–6534. https://doi.org/10.1002/jmv.27191

Faria, N.R., Claro, I.M., Candido, D., Franco, L.A.M., Andrade, P.S., Coletti, T.M., Silva, C.A.M., Sales, F.C., Manulli, E.R., Aguiar, R.S., Gaburo, N., Camilo, C. da C., Fraiji, N.A., Crispim, M.A.E., Carvalho, M.P.S.S., Rambaut, A., Loman, N., Phybus, O.G., Sabino, E.C., 2021a. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological.Org 1–9.

Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.S., Mishra, S., Crispim, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulswit, Ruben, J.G., Franco, L.A.M., Raimundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., da Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Jr Gaburo, N., Camilo, C. da C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., Souza, L.M. de, Pinho, M.C. de, Araujo, L.J.T., Malta, F.S. V, Lima, A.B. de, Silva, J. do P., Zauli, D.A.., Ferreira, A.C.S., Schnekenberg, R.P., Laydon, D.J., Walker, Patrick, G.T., Schluter, H.M., Santos, A.L.P. dos, Vidal, M.S., Caro, V.S. Del, Filho, R.M.F., Santos, H.M. dos, Aguiar, R.S., Proenca-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscouridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Carlos, P.J., Nascimento, V.H., A, S.M., Bowden, T.A., Pond, S.L.K., Wu, C.-H., Ratmann, O., Ferguson, N.M., Christopher, D., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M. do P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021b. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science (80-.). 372, 815–821.

Fung, T.S., Liu, D.X., 2019. Human Coronavirus : Host-Pathogen Interaction. Annu. Rev. Microbiol. 73, 529–557.

Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat. Rev. Microbiol. 19, 409–424. https://doi.org/10.1038/s41579-021-00573-0

Khan, A., Zia, T., Suleman, M., Khan, T., Ali, S.S., Abbasi, A.A., Mohammad, A., Wei, D.Q., 2021. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. J. Cell. Physiol. 236, 7045–7057. https://doi.org/10.1002/jcp.30367

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Angyal, A., Brown, R.L., Carrilero, L., Green, L.R., Groves, D.C., Johnson, K.J., Keeley, A.J., Lindsey, B.B., Parsons, P.J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R.M., Wang, D., Wyles, M.D., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori,

D.C., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182, 812–827. https://doi.org/10.1016/j.cell.2020.06.043

Laamarti, Meriem, Alouane, T., Kartti, S., Chemao-Elfihri, M.W., Hakmi, M., Essabbar, A., Laamarti, Mohamed, Hlali, H., Bendani, H., Boumajdi, N., Benhrif, O., Allam, L., Hafidi, N. El, Jaoudi, R. El, Allali, I., Marchoudi, N., Fekkak, J., Benrahma, H., Nejjari, C., Amzazi, S., Belyamani, L., Ibrahimi, A., 2020. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geodistribution and a rich genetic variations of hotspots mutations. PLoS One 15, 1–18. https://doi.org/10.1371/journal.pone.0240345

Lamarca, A.P., de Almeida, L.G.P., Francisco, R. da S., Lima, L.F.A., Scortecci, K.C., Perez, V.P., Brustolini, O.J., Sousa, E.S.S., Secco, D.A., Santos, A.M.G., Albuquerque, G.R., Mariano, A.P.M., Maciel, B.M., Gerber, A.L., Guimarães, A.P. de C., Nascimento, P.R., Neto, F.P.F., Gadelha, S.R., Porto, L.C., Campana, E.H., Jeronimo, S.M.B., Vasconcelos, A.T.R., 2021. Genomic surveillance of SARS-CoV-2 tracks early interstate transmission of P.1 lineage and diversification within P.2 clade in Brazil. PLoS Negl. Trop. Dis. 15, e0009835. https://doi.org/10.1371/journal.pntd.0009835

Lauring, A.S., Hodcroft, E.B., 2021. Genetic Variants of SARS-CoV-2 - What Do They Mean? JAMA - J. Am. Med. Assoc. 325, 529–531. https://doi.org/10.1001/jama.2020.27124

Lei, J., Kusov, Y., Hilgenfeld, R., 2018. Nsp3 of coronaviruses : Structures and functions of a large multi-domain protein. Antiviral Res. 149, 58–74. https://doi.org/10.1016/j.antiviral.2017.11.001

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Mariano, G., Farthing, R.J., Lale-Farjat, S.L.M., Bergeron, J.R.C., 2020. Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. Front. Mol. Biosci. 7. https://doi.org/10.3389/fmolb.2020.605236

Nagy, Á., Pongor, S., Győrffy, B., 2021. Different mutations in SARS-CoV-2 associate with severe and mild outcome. Int. J. Antimicrob. Agents 57, 1–5. https://doi.org/10.1016/j.ijantimicag.2020.106272

Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5, 1403–1407. https://doi.org/10.1038/s41564-020-0770-5

Resende, P.C., Delatorre, E., Gräf, T., Mir, D., Motta, F.C., Appolinario, L.R., Paixão, A.C.D. da, Mendonça, A.C. da F., Ogrzewalska, M., Caetano, B., Wallau, G.L., Docena, C., Santos, M.C. dos, de Almeida Ferreira, J., Sousa Junior, E.C., Silva, S.P. da, Fernandes, S.B., Vianna, L.A., Souza, L. da C., Ferro, J.F.G., Nardy, V.B., Santos, C.A., Riediger, I., do Carmo Debur, M., Croda, J., Oliveira, W.K., Abreu, A., Bello, G., Siqueira, M.M., 2021a. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. Front. Microbiol. 11, 1–14. https://doi.org/10.3389/fmicb.2020.615280

Resende, P.C., Gräf, T., Paixão, A.C.D., Appolinario, L., Lopes, R.S., Mendonça, A.C. da F., da Rocha, A.S.B., Motta, F.C., Neto, L.G.L., Khouri, R., de Oliveira, C.I., Santos-Muccillo,

P., Bezerra, J.F., Teixeira, D.L.F., Riediger, I., Debur, M.D.C., Ribeiro-Rodrigues, R., Leite, A.B., Do Santos, C.A., Gregianini, T.S., Fernandes, S.B., Bernardes, A.F.L., Cavalcanti, A.C., Miyajima, F., Sachhi, C., Mattos, T., da Costa, C.F., Delatorre, E., Wallau, G.L., Naveca, F.G., Bello, G., Siqueira, M.M., 2021b. 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 13, 1–7. https://doi.org/10.3390/v13050724

Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., Kosugi, Y., Shirakawa, K., Sadamasu, K., Kimura, I., Ito, J., Wu, J., Iwatsuki-Horimoto, K., Ito, M., Yamayoshi, S., Loeber, S., Tsuda, M., Wang, L., Ozono, S., Butlertanaka, E.P., Tanaka, Y.L., Shimizu, R., Shimizu, K., Yoshimatsu, K., Kawabata, R., Sakaguchi, T., Tokunaga, K., Yoshida, I., Asakura, H., Nagashima, M., Kazuma, Y., Nomura, R., Horisawa, Y., Yoshimura, K., Takaori-Kondo, A., Imai, M., Chiba, M., Furihata, H., Hasebe, H., Kitazato, K., Kubo, H., Misawa, N., Morizako, N., Noda, K., Oide, A., Suganami, M., Takahashi, M., Tsushima, K., Yokoyama, M., Yuan, Y., Tanaka, S., Nakagawa, S., Ikeda, T., Fukuhara, T., Kawaoka, Y., Sato, K., 2021. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. Nature 602. https://doi.org/10.1038/s41586-021-04266-9

Souvorov, A., Agarwala, R., Lipman, D.J., 2018. SKESA: Strategic k-mer extension for scrupulous assemblies. Genome Biol. 19, 1–13. https://doi.org/10.1186/s13059-018-1540-z

Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., Kiyotani, K., 2020. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. J. Hum. Genet. 65, 1075–1082. https://doi.org/10.1038/s10038-020-0808-9

Van der Auwera, G., O'Connor, B., 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.

Voloch, C.M., Jr, R.S.F., Almeida, L.G.P., Cardoso, C.C., Brustolini, O.J., Gerber, A.L., de Guimarães, A.P.C., Mariani, D., da Costa, R.M., 2020. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. medRxiv.

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. Lancet 395, 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9

Wang, P., Casner, R.G., Nair, M.S., Wang, M., Yu, J., Cerutti, G., Liu, L., Kwong, P.D., Huang, Y., Shapiro, L., Ho, D.D., 2021a. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. Cell Host Microbe 29, 747–751.e4. https://doi.org/10.1016/j.chom.2021.04.007

Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., Graham, B.S., Mascola, J.R., Chang, J.Y., Yin, M.T., Sobieszczyk, M., Kyratsous, C.A., Shapiro, L., Sheng, Z., Huang, Y., Ho, D.D., 2021b. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593, 130–135. https://doi.org/10.1038/s41586-021-03398-2

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269. https://doi.org/10.1038/s41586-020-2008-3

Yang, H.C., Chen, C.H., Wang, J.H., Liao, H.C., Yang, C.T., Chen, C.W., Lin, Y.C., Kao, C.H., Lu, M.Y.J., Liao, J.C., 2020. Analysis of genomic distributions of SARS-CoV-2 reveals

a dominant strain type with strong allelic associations. Proc. Natl. Acad. Sci. U. S. A. 117, 30679–30686. https://doi.org/10.1073/pnas.2007840117

Yin, C., 2020. Genotyping coronavirus SARS-CoV-2: methods and implications. Genomics 112, 3588–3596. https://doi.org/10.1016/j.ygeno.2020.04.016

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S.F., Lemieux, J.E., Munro, J.B., Rafique, A., Barve, A., Sabeti, P.C., Kyratsous, C.A., Dudkina, N. V., Shen, K., Luban, J., 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell 183, 739–751. https://doi.org/10.1016/j.cell.2020.09.032

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733. https://doi.org/10.1056/nejmoa2001017

CONSIDERAÇÕES FINAIS

O vírus SARS-CoV-2 está circulando mundialmente há aproximadamente dois anos, e ainda não há perspectiva de quando a pandemia chegará ao fim, sendo difícil precisar quanto tempo ainda permaneceremos tendo que adotar medidas como o uso de máscaras ou o distanciamento social. Contudo, o que está ao nosso alcance para melhor compreendermos o cenário que estamos vivendo e nos protegendo conforme necessário, é continuarmos a realizar a vigilância genômica do vírus, monitorando novas mutações e também acompanhando o surgimento e dispersão de novas linhagens.

Neste trabalho, abordamos duas frentes de combate/monitoramento da pandemia: aspectos do diagnóstico molecular e monitoramento das linhagens circulantes através do sequenciamento genético. Nossa abordagem sobre os métodos de diagnósticos elucidou os fatores que podem contribuir para a obtenção de resultados falso-positivos ou falto-negativos. Essas informações são essenciais para o melhor entendimento da sociedade sobre o qRT-QPCR, método considerado o padrão ouro de detecção do vírus SARS-CoV-2. Quanto à vigilância genômica, trouxemos novas informações sobre as linhagens circulantes no estado do Ceará, detalhando as mutações encontradas nos genomas de 34 amostras de pacientes positivos para SARS-Cov-2, no período de março a dezembro de 2020. Avaliando esses genomas juntamente com os depositados no banco de dados GISAID, também provenientes do estado do Ceará, conseguimos determinar as linhagens mais dominantes durante o ano de 2020 e identificar as mutações presentes, comparando-os ao primeiro genoma sequenciado.

Como conclusão, reforçamos a importância de estudos semelhantes a esse, que contribuam para o monitoramento da pandemia do SARS-CoV-2 e também para futuros surtos de outros vírus que podem vir à tona. Estudos como esse colaboram para o melhor entendimento da situação mundial frente ao vírus circulante e para a tomada de decisões relacionadas às medidas de combate e proteção adequadas por parte dos governantes.