



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS



ZILDENE DE SOUSA SILVEIRA

**DIVERSIDADE GENÔMICA DE CEPAS DE SARS-CoV-2 DURANTE A PRIMEIRA
ONDA DA COVID-19 NO ESTADO DE PERNAMBUCO, BRASIL**

Recife
2022

ZILDENE DE SOUSA SILVEIRA

**DIVERSIDADE GENÔMICA DE CEPAS DE SARS-CoV-2 DURANTE A PRIMEIRA
ONDA DA COVID-19 NO ESTADO DE PERNAMBUCO, BRASIL**

Dissertação apresentada ao
Programa de Pós-Graduação em
Ciências Biológicas da Universidade
Federal de Pernambuco, como
requisito parcial para obtenção do
título de mestre em Ciências
Biológicas

Área de concentração:
Bioinformática

Orientador: Prof. Dr. Valdir de Queiroz Balbino

Coorientador: Prof. Dr. Francisco Assis Bezerra da Cunha

Recife

2022

Catálogo na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

Silveira, Zildene de Sousa

Diversidade genômica de cepas de SARS-CoV-2 durante a primeira onda da Covid-19 no Estado de Pernambuco, Brasil / Zildene de Sousa Silveira. - 2022.

107 f. : il.

Orientador: Prof. Dr. Valdir de Queiroz Balbino.

Coorientador: Prof. Dr. Francisco Assis Bezerra da Cunha.

Dissertação (mestrado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-Graduação em Ciências Biológicas, Recife, 2022.

Inclui referências e apêndices.

1. Virologia. 2. Covid-19. 3. Epidemiologia. I. Balbino, Valdir de Queiroz (orientador). II. Cunha, Francisco Assis Bezerra da (coorientador). III. Título.

616.91

CDD (22.ed.)

UFPE/CB-050-2022

ZILDENE DE SOUSA SILVEIRA

DIVERSIDADE GENÔMICA DE CEPAS DE SARS-CoV-2 DURANTE A PRIMEIRA ONDA DA COVID-19 NO ESTADO DE PERNAMBUCO, BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestre em Ciências Biológicas.

Aprovada em: 04/02/2022

BANCA EXAMINADORA

Prof^o. Dr. Valdir de Queiroz Balbino (Orientador)
Universidade Federal de Pernambuco

Prof^a. Dr^a. Rosane Silva (Examinadora externa)
Universidade Federal do Rio de Janeiro

Prof^o. Dr. Reginaldo Gonçalves de Lima Neto (Examinador externo)
Universidade Federal de Pernambuco

Dedico a minha família por todo apoio e amor compartilhado!

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus pela força e determinação a mim dada ao longo deste período. Pelos momentos de desânimo em que me manteve de pé.

Agradeço a minha mãe, Irene de Sousa Silveira, pelos ensinamentos, confiança, amparo e força. A meu Pai e irmãos pelos conselhos e apoio nas horas que necessitei. Ao meu namorado e companheiro, Rodrigo Vaz, pela paciência, conselhos e contribuições ao meu trabalho e principalmente por todo seu amor.

Agradeço aos meus amigos de longa data que estiveram ao meu lado ajudando-me a enfrentar mais esse desafio. Em especial, agradeço a minha amiga Nair pelos desabafos e parcerias.

Também estendo meus agradecimentos aos novos amigos e colaboradores que ganhei ao longo desses 2 anos, não só em ambiente acadêmico, mas também na vida cotidiana que construí em Recife. Cito aqui, a minha querida amiga Ana Cláudia, Jussara e a Marcos Regueira por todo apoio a realização do meu trabalho e conselhos quando precisei.

Ao Prof. Dr. Valdir Balbino, agradeço a orientação, espaço, conselhos e críticas, ensinamentos e oportunidades.

Agradeço também ao meu coorientador Prof. Dr. Francisco Assis Bezerra da Cunha, pelo apoio, ensinamentos, conselhos e principalmente pela parceria ao longo do mestrado.

Ao Laboratório de Bioinformática e Biologia Evolutiva - LABBE pelo espaço e infraestrutura e a sua equipe pela parceria e colaboração. Cito aqui: Dr. Marcos Regueira, Dr. Wilson da Silva Júnior, Prof. Dr. Sérgio Paiva, Dra. Heidi Lacerda, Dr. Moisés Freitas, MSc. Bruno Sampaio, Jussara, José Bandeira e Dayana. Obrigada por todos os ensinamentos durante a pesquisa e pelos momentos de descontração. Sem a colaboração de vocês, a conclusão deste trabalho seria impossível.

À Universidade Federal de Pernambuco, pelo espaço. Minha gratidão aos professores e funcionários que integram o Programa de Pós-Graduação em Ciências Biológicas - PPGCB. Agradeço o suporte financeiro da agência de fomento à pesquisa Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES.

RESUMO

Síndrome Respiratória Aguda Grave coronavírus 2 (SARS-CoV-2) é responsável pela pandemia da doença do coronavírus-2019 (COVID-19), a qual levou à notificação de mais de 200 milhões de casos em todo o mundo. Desde então, a vigilância genômica tem sido uma ferramenta fundamental para rastrear a disseminação viral. Neste estudo fornecemos uma análise detalhada da dinâmica temporal de mutações e linhagens de SARS-CoV-2 durante a primeira onda da pandemia no estado de Pernambuco, Brasil. Para isso, realizamos a caracterização de 67 genomas obtidos a partir de *swabs* nasofaríngeos e orofaríngeos de pacientes positivos para COVID-19 de 22 municípios, os quais foram combinados com 293 sequências de Pernambuco disponíveis no GISAID obtidas entre fevereiro e dezembro de 2020, período adotado como marco temporal da primeira onda pandêmica no estado de Pernambuco. Descrevemos a disseminação temporal e a evolução do SARS-CoV-2 durante a primeira onda pandêmica em Pernambuco, sendo detectados 962 SNPs em regiões codificantes, a maioria deles localizados na ORF1ab. Verificamos que as mutações C241T, C3037T (F924F), C14408T (P4715L/P323L), A23403G (D614G), G28881A (R203K), G28882A (R203R) e G28883C (G204R), mantiveram suas frequências relativas acima de 94 % durante todo o período de amostragem. Foi detectada a circulação simultânea de 18 linhagens em Pernambuco, sendo as mais prevalentes B.1.1, B.1.1.28, B.1.1.33 e P.2, respectivamente. A rede de haplótipos e a análise filogenômica mostraram que as sequências se agruparam em 4 clados principais de acordo com a dinâmica evolutiva de SARS-CoV-2, tendo havido importações iniciais de linhagens com provável origem da Europa, seguida por expansão e transmissão local de linhagens nacionais, contribuindo para o crescimento acelerado da epidemia. Esses resultados evidenciam elevados níveis de diversidade genômica em Pernambuco e destacam a importância de investigações contínuas em vigilância genômica para rastreamento de variantes virais.

Palavras-chave: SARS-CoV-2; Pernambuco; Vigilância genômica; Filogenômica.

ABSTRACT

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is responsible for the coronavirus disease pandemic-2019 (COVID-19), which led to the reporting of more than 200 million cases worldwide. Since then, genomic surveillance has been a key tool to track viral spread. In this study we provide a detailed analysis of the temporal dynamics of SARS-CoV-2 mutations and strains during the first wave of the pandemic in the state of Pernambuco, Brazil. To this end, we performed the characterization of 67 genomes obtained from nasopharyngeal and oropharyngeal swabs of COVID-19 positive patients from 22 municipalities, which were combined with 293 Pernambuco sequences available in GISAID obtained between February and December 2020, the period adopted as the temporal marker of the first pandemic wave in the state of Pernambuco. It describes the temporal dissemination and evolution of SARS-CoV-2 during the first pandemic wave in Pernambuco, being detected 962 SNPs in coding regions, most of them located in ORF1ab. We found that mutations C241T, C3037T (F924F), C14408T (P4715L/P323L), A23403G (D614G), G28881A (R203K), G28882A (R203R) and G28883C (G204R), maintained their relative frequencies above 94% during the entire sampling period. Simultaneous circulation of 18 strains was detected in Pernambuco, the most prevalent being B.1.1, B.1.1.28, B.1.1.33 and P.2, respectively. The haplotype network and phylogenomic analysis showed that the sequences grouped into 4 main clusters according to the evolutionary dynamics of SARS-CoV-2, with initial importation of strains with probable origin from Europe, followed by expansion and local transmission of national strains, contributing to the rapid growth of the epidemic. These results highlight high levels of genomic diversity in Pernambuco and highlight the importance of continued investigations in genomic surveillance for tracking viral variants.

Keywords: SARS-CoV-2; Pernambuco; Genomic surveillance; Phylogenomics.

LISTA DE ILUSTRAÇÕES

Referencial Teórico

Figura 1 –	Esquema de taxonomia da família <i>Coronaviridae</i> de acordo com a classificação do Comitê Internacional de Taxonomia de Vírus (ICTV)	17
Figura 2 –	Representação esquemática da organização genômica e características estruturais de SARS-CoV-2	18
Figura 3 –	Distribuição de sequências genômicas disponíveis por continente de amostragem entre janeiro de 2020 e dezembro de 2021	22
Figura 4 –	Principais aplicações das ferramentas de bioinformática no estudo de SARS-CoV-2	24

Artigo

Figura 1 –	Indicadores epidemiológicos da COVID-19 no estado de Pernambuco durante a primeira onda da pandemia	33
Figura 2–	Períodos de coleta dos 360 genomas coletados em Pernambuco	34
Figura 3 –	Alteração temporal na prevalência de mutações durante a primeira onda da pandemia em Pernambuco	36
Figura 4 –	Correlação entre a frequência temporal de linhagens e principais mutações de SARS-CoV-2 a cada mês de amostragem	38
Figura 5 –	Árvore filogenética de máxima verossimilhança resolvida no tempo dos genomas SARS-CoV-2 do estado de Pernambuco, coletados entre fevereiro e dezembro de 2020	39
Figura 6 –	Rede de haplótipos de genomas completos de SARS-CoV-2 usando o método <i>median joining network</i>	42

LISTA DE TABELAS

Tabela 1 –	Sítios variáveis e composição nucleotídica dos genomas sequenciados para Pernambuco, Brasil	35
------------	--	----

LISTA DE ABREVIATURAS E SIGLAS

ECA2	Enzima Conversora de Angiotensina 2Malondialdeído
COVID-19	Coronavirus Disease 2019 – Doença do Coronavírus 2019
GISAID	<i>Global Initiative on Sharing Avian Influenza Data</i> – Iniciativa Global de Compartilhamento de Dados da Gripe Aviária
PANGOLIN	<i>Phylogenetic Assignment of Named Global Outbreak Lineages</i> – Atribuição filogenética de linhagens de surto globais nomeadas
RBD	<i>Receptor Binding Domain</i> – Domínio de ligação ao receptor
SNP	<i>Single Nucleotide Polymorphism</i> – Polimorfismo de Nucleotídeo Único
VOC	<i>Variant of Concern</i> – Variante de Preocupação
VOI	<i>Variants of Interest</i> – Variante de Interesse
WHO	<i>World Health Organization</i> – Organização Mundial de Saúde

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS	14
1.1.1	Objetivo geral	14
1.1.2	Objetivos específicos	14
2	REFERENCIAL TEÓRICO	15
2.1	ORIGEM ZOONÓTICA DE SARS-CoV-2	15
2.2	CLASSIFICAÇÃO E ESTRUTURA TAXONÔMICA DE SARS-CoV-2	16
2.3	ORGANIZAÇÃO ESTRUTURAL E GENÔMICA DE SARS-CoV-2	18
2.4	VIGILÂNCIA GENÔMICA DE SARS-CoV-2	21
2.5	SEQUENCIAMENTO DE PRÓXIMA GERAÇÃO (NGS) E FERRAMENTAS DE BIOINFORMÁTICA PARA PESQUISA E VIGILÂNCIA GENÔMICA DE SARS-COV-2	22
3	RESULTADOS E DISCUSSÃO	25
3.1	ARTIGO 1: Altos níveis de diversidade genômica em cepas SARS-CoV-2 circulando no Estado de Pernambuco (Nordeste do Brasil) durante a “primeira onda” do COVID-19	25
4	CONCLUSÕES	54
5	SÚMULA CURRICULAR	55
	REFERÊNCIAS	56
	APÊNDICE A - DISTRIBUIÇÃO DAS AMOSTRAS E DADOS DO RELATÓRIO QUAST DOS GENOMAS SEQUENCIADOS DE SARS-COV-2 ENTRE OS MUNICÍPIOS DO ESTADO DE PERNAMBUCO, BRASIL	64
	APÊNDICE B - MUTAÇÕES COM FREQUÊNCIAS MAIORES QUE 5 % ENTRE OS 67 GENOMAS DE PERNAMBUCO OBTIDOS NESTE ESTUDO E AS 293 RECUPERADAS DO GISAID	71
	APÊNDICE C - MUTAÇÕES DEFINIDORAS DE CADA CLADO	74
	APÊNDICE D - DEFINIÇÕES DE HAPLÓTIPOS DE SARS-COV-2 ANALISADOS E SUAS FREQUÊNCIAS	83

1 INTRODUÇÃO

O Coronavírus 2 da Síndrome Respiratória Aguda Grave (SARS-CoV-2) é um membro da família *Coronaviridae* que recentemente emergiu na província de Wuhan, China, tendo se espalhado rapidamente pelo mundo, causando a atual pandemia da Doença do Coronavírus 2019 (COVID-19) (WU *et al.*, 2020; GORBALENYA *et al.*, 2020). Trata-se de um vírus envelopado de genoma de RNA de fita simples e sentido positivo, com cerca de 29 kbs de comprimento (CUI *et al.*, 2019). Seu RNA genômico compreende até 14 quadros de leitura aberta (ORFs), sendo que os dois ORFs principais, ORF1a e ORF1b, sobrepostos a um *frameshift* ribossomal, abrangem dois terços do genoma. O restante do genoma tem ORFs sobrepostos, codificando quatro proteínas estruturais principais: pico (S, do inglês *Spike*), Membrana (M), Envelope (E) e Nucleocapsídeo (N) (LU *et al.*, 2020; CHAN *et al.*, 2020; MOUSAVIZADEH; GHASEMI, 2021). Além disso, a anotação atual do SARS-CoV-2 inclui previsões de seis proteínas acessórias (3a, 6, 7a, 7b, 8 e 10, NC_045512.2) (FINKEL *et al.*, 2021).

A emergência de saúde pública representada pela COVID-19 levou a um esforço colaborativo global para compreensão epidemiológica de SARS-CoV-2, culminando em milhares de sequências genômicas disponibilizadas e disseminação de ferramentas de visualização e análise de dados (DELLICOUR *et al.*, 2021). Atualmente estão disponíveis mais de 3 milhões de sequências completas e com alta cobertura no serviço de compartilhamento de dados da *Global Initiative on Sharing All Influenza Data* (GISAID) (SHU; MCCAULEY, 2017). Além disso, as evidências recentes do surgimento de mutações que proporcionam escape ao sistema imunológico hospedeiro e de múltiplas variantes de rápida disseminação, destacam a importância da vigilância genômica em todo o mundo para rastreamento e verificação da prevalência local dessas variantes, bem como a emergência de outras (WANG *et al.*, 2021).

Em 29 de dezembro de 2021, haviam sido notificados 281.808.270 de casos da COVID-19 e 5.411.759 mortes em todo o mundo (WHO, 2021a). Na América do Sul, considerada um dos epicentros da pandemia, foi registrada a marca de 39,59 milhões de casos acumulados, sendo cerca de 22,25 milhões apenas no Brasil (RITCHIE *et al.*, 2020).

O primeiro caso de COVID-19 confirmado no Brasil ocorreu em 26 de

fevereiro de 2020, no estado de São Paulo (OPAS, 2020). A epidemia no país foi caracterizada por várias introduções independentes, seguida de transmissão comunitária, o que culminou em registros contínuos de casos e mortes, os quais atingiram um pico no início de maio de 2020, seguido de queda nos indicadores epidemiológicos, os quais permaneceram mais ou menos estáveis de junho a novembro de 2020, caracterizando a primeira onda (CANDIDO *et al.*, 2020; NAVECA *et al.*, 2021). No entanto, entre o final de novembro e dezembro o número de casos começou a crescer exponencialmente, estabelecendo a segunda onda da epidemia no país. Até a data da escrita desse trabalho, o Brasil já havia perdido mais de 600 mil vidas e notificado mais de 22 milhões de casos de infecções (WHO, 2021a).

O estado de Pernambuco, localizado no nordeste do Brasil, tem uma população estimada em 9.674.793 pessoas e extensão territorial de 98.067.880 km² (IBGE, 2021). No estado, a notificação do primeiro caso de COVID-19 se deu em 12 de março de 2020 e atualmente já foram registrados 644.852 casos e 20.431 óbitos (MINISTÉRIO DA SAÚDE, 2021). Assim como em outras regiões do país a pandemia em Pernambuco, foi caracterizada por maior concentração dos casos na capital do estado, seguida por intensa disseminação por seu interior (CARMO *et al.*, 2021). No entanto, apesar das proporções que a epidemia tomou no estado, a amostragem de dados genômicos ainda é bastante limitada, o que impede uma investigação genômica abrangente da dinâmica da primeira onda da pandemia a nível local.

1.1 OBJETIVOS

1.1.1 objetivo geral

O objetivo deste estudo foi caracterizar a disseminação temporal e a evolução do SARS-CoV-2 durante a primeira onda pandêmica em Pernambuco, através de uma análise de predominância de variantes e linhagens em circulação no estado a partir de genomas de SARS-CoV-2 obtidas entre fevereiro e dezembro de 2020.

1.1.2 Objetivos específicos

- Sequenciar e identificar cepas de SARS-CoV-2 circulantes no estado de Pernambuco, a partir de amostras biológicas obtidas durante a primeira onda da pandemia no estado
- Avaliar a diversidade genética das cepas de SARS-CoV-2 obtidas em Pernambuco
- Caracterizar a dinâmica temporal de variantes e linhagens em Pernambuco durante a primeira onda da pandemia
- Traçar relações filogenéticas entre genomas de SARS-CoV-2 obtidos em Pernambuco e sequências disponíveis para todo o Brasil no GISAID coletados até o final da primeira onda pandêmica

2 REFERENCIAL TEÓRICO

2.1 ORIGEM ZOONÓTICA DE SARS-CoV-2

O surgimento de surtos de vírus zoonóticos são responsáveis por uma complexa ameaça aos sistemas de saúde, economia e segurança globais (PÉREZ-LOSADA *et al.*, 2015). A disseminação viral através da transferência zoonótica varia com base em vários fatores, como fisiologia, história de vida, ecologia e a tendência de vários animais em transmitir vírus aos seres humanos (MOLLENTZE; STREICKER, 2020). Essas infecções surgem principalmente devido à semelhança filogenética entre hospedeiros reservatórios e humanos e a sobreposição dos espaços com populações humanas (OLIVAL *et al.*, 2017). Além disso, fatores como mudanças climáticas, distribuição e prevalência de reservatórios hospedeiros e seus parasitas oferecem grande contribuição para o surgimento de infecções (ALLEN *et al.*, 2017; VANWAMBEKE; LINARD; GILBERT, 2019).

Os coronavírus (CoVs) são uma das principais ameaças a saúde dos humanos e outras espécies de vertebrados, sendo patógenos de rápida evolução, o que os tornam altamente suscetíveis a mutações adaptativas e consequentes saltos entre hospedeiros (FORNI *et al.*, 2017). Desse modo, esses vírus podem infectar uma variedade de animais domésticos e selvagens como gado, pássaros, morcegos, ratos, entre outros, sendo responsáveis por infecções respiratórias, gastrointestinais, hepáticas e do sistema nervoso central (WANG *et al.*, 2006; CUI; LI; SHI, 2019). Até hoje, sete espécies de coronavírus são conhecidas por causar doenças em humanos. Quatro dessas (229E, OC43, NL63 e HKU1) são as mais prevalentes e geralmente causam sintomas de resfriado comum em indivíduos imunocompetentes (SU *et al.*, 2016). Além disso, estão incluídos nesse grupo, o Coronavírus da Síndrome Respiratória Aguda Grave (SARS-CoV) (DROSTEN *et al.*, 2003) e o Coronavírus da Síndrome Respiratória do Oriente Médio (MERS-CoV) (ZAKI *et al.*, 2012) que são de origem zoonótica e foram associadas a doenças respiratórias graves nos anos de 2002-2003 e 2012 respectivamente (CUI; LI; SHI, 2019).

Mais recentemente, no final de dezembro de 2019, os primeiros casos da doença do coronavírus 2019 (COVID-19), causada pela Síndrome Respiratória Aguda Grave Coronavírus 2 (SARS-CoV-2), foram descritas na cidade de Wuhan, na

província de Hubei, China (ZHU *et al.*, 2020; ZHOU *et al.*, 2020). O vírus se espalhou rapidamente sendo declarado uma emergência de saúde pública de interesse mundial (SHEREEN *et al.*, 2020; WHO, 2020). No entanto, as circunstâncias que envolveram o surgimento do SARS-CoV-2 na província de Hubei ainda permanecem ocultas (CALLAWAY, LEDFORD; MALLAPATY, 2020; DELAUNE *et al.*, 2021).

A origem exata de SARS-CoV-2 é desconhecida, no entanto, um relatório anterior de Zhou *et al.* (2020) identificou uma sequência de genoma de SARSr-CoV (*RaTG13*) que compartilhava uma identidade de sequência de 96% do genoma inteiro com SARS-CoV-2, indicando uma provável origem de morcego. Desde então, mais sequências do genoma viral de morcegos relacionadas ao SARS-CoV-2 foram relatadas no leste da China, no Japão, Tailândia e Camboja, e de pangolins na China (ZHOU H *et al.*, 2020; MURAKAMI *et al.*, 2020; WACHARAPLUESADEE *et al.*, 2021; DELAUNE *et al.*, 2021; LAM *et al.*, 2020). Porém, ainda não se sabe o ancestral imediato, nem se o vírus emergiu diretamente de seu reservatório animal ou se circulava inicialmente em um hospedeiro intermediário (PEKAR *et al.*, 2021).

A identificação da origem e a compreensão de como SARS-CoV-2 ultrapassou a barreira entre espécies para infectar humanos de forma tão eficaz, não é apenas importante em termos científicos, mas também se constitui um componente crítico na prevenção de novos surtos causados por SARS-CoV-2 e de futuras pandemias com características semelhantes às da pandemia de COVID-19 (RELMAN, 2020; ANDERSEN *et al.*, 2020; WHO, 2021b).

2.2. CLASSIFICAÇÃO E ESTRUTURA TAXONÔMICA DE SARS-CoV-2

Os coronavírus (CoVs) integram a família *Coronaviridae*, ordem *Nidovirales*, subordem *Cornidovirineae* e reino *Riboviria* (Figura 1) (ICTV, 2020; WASSENAAR; ZOU, 2020). Atualmente são reconhecidas 46 espécies em 27 subgêneros, cinco gêneros e duas subfamílias (GORBALENYA *et al.*, 2020). Dentre as subfamílias, a *Orthocoronavirinae* é a maior, abrigando 45 das 46 espécies de CoVs, as quais estão alocadas em quatro gêneros, nomeadamente: *Alphacoronavírus* (α -CoV), *Betacoronavírus* (β -CoV), *Gammacoronavírus* (γ -CoV) e *Deltacoronavírus* (δ -CoV). Os patógenos dos gêneros *Deltacoronavirus* e *Gammacoronavirus* circulam principalmente em aves, porém alguns deles também podem infectar mamíferos

(WOO *et al.*, 2012).

Figura 1: Esquema de taxonomia da família *Coronaviridae* de acordo com a classificação do Comitê Internacional de Taxonomia de Vírus (ICTV)



Fonte: Silveira (2021).

Os *Alphacoronavirus* e *Betacoronavirus* são capazes de infectar uma ampla variedade de mamíferos (LEAO *et al.*, 2020). O primeiro compreende 19 espécies que infectam um grupo diversificado de mamíferos, incluindo duas (229E e NL63) das sete espécies conhecidas de coronavírus humanos (CUI; LI; SHI, 2019). Já os *Betacoronavirus* compreendem 14 espécies distribuídas em cinco subgêneros (GORBALENYA *et al.*, 2020). Os subgêneros incluem *Embecovirus*, *Nobecovirus*, *Hibecovirus*, *Merbecovirus* e *Sarbecovirus*, estabelecidos de acordo com análises filogenéticas de domínios proteicos conservados (ICTV, 2020).

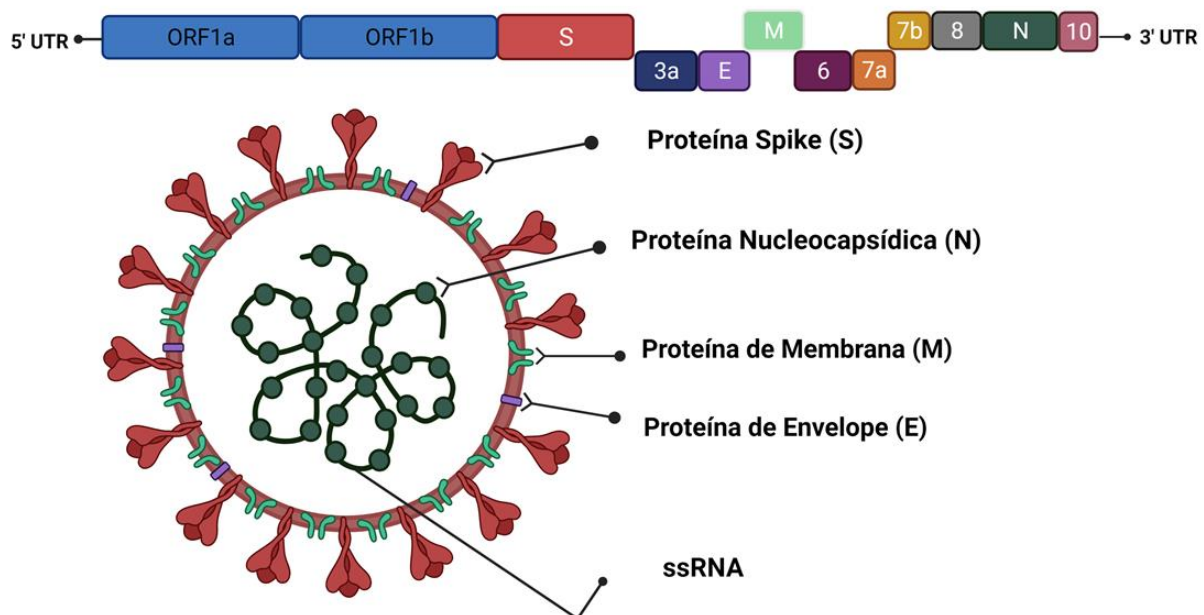
Os *Embecovirus* incluem dois coronavírus humanos (HKU1 e OC43), bem como vários coronavírus de relevância veterinária. Já *Nobecovirus* e *Hibecovirus* compreendem apenas coronavírus de morcego, isolados principalmente de *Rousettus* e *Hipposideros*, respectivamente (LLANES *et al.*, 2020). Em contraste, os CoVs altamente patogênicos SARS-CoV e SARS-CoV-2 integram o grupo dos *Sarbecovirus*, enquanto *MERS-CoV* compõe o subgênero *Merbecovirus* (GORBALENYA *et al.*, 2020).

2.3. ORGANIZAÇÃO ESTRUTURAL E GENÔMICA DE SARS-CoV-2

SARS-CoV-2, assim como outros coronavírus (CoVs), é caracterizado como um vírus envelopado de formato esférico, simetria helicoidal e com a presença de peplômeros compostos de glicoproteína sobre sua superfície, dando-lhe a forma de coroa (WASSENAAR; ZOU, 2020). Possuem de 60 a 140 nm de diâmetro e material nucleico com RNA de fita simples e sentido positivo (+ssRNA) de cerca de 29 kbs de comprimento (SHEREEN *et al.*, 2020; PENG *et al.*, 2020; LAI *et al.*, 2020).

A análise da estrutura genômica de SARS-CoV-2, sugere que são codificadas até 14 quadros de leitura aberta (ORFs), sendo 12 funcionais. Eles incluem a ORF1a, ORF1b, S, ORF3a, E, M, ORF6, ORF7a, ORF7b, ORF8, N e ORF10, os quais desempenham papéis importantes na entrada, fusão e sobrevivência viral nas células hospedeiras (CHEN *et al.*, 2021, ALEXANDERSEN; CHAMINGS; BHATTA, 2020) (Figura 2).

Figura 2: Representação esquemática da organização genômica e características estruturais de SARS-CoV-2



Fonte: Silveira (2021), elaborada com a ferramenta BioRender.

Duas ORFs sobrepostas em um *frameshift* ribossomal, ORF1a e ORF1b, são traduzidas do RNA genômico de fita positiva e cobrem dois terços do genoma, sendo

a tradução de ORF1b mediada por um deslocamento do quadro 1, o qual permite que a tradução continue além do códon de parada de ORF1a. ORF1a e ORF1b são traduzidos em duas poliproteínas (pp1a e pp1ab), as quais são clivadas proteoliticamente em 16 proteínas não-estruturais (Nsps) (CUI; LI; SHI, 2019). Nsp1 a nsp11 são codificados por ORF1a e, portanto, estão presentes em pp1a e pp1ab, enquanto nsp12 a nsp16 são codificados por ORF1b e estão presentes apenas em pp1ab (LLANES *et al.*, 2020). Os intermediários de RNA de fita negativa são produzidos a partir do genoma viral e servem como modelos para a síntese de mRNA genômico de fita positiva e mRNAs subgenômicos (CHAN *et al.*, 2020; FINKEL *et al.*, 2021).

Os mRNAs subgenômicos contêm uma sequência líder 5' comum fundido a diferentes segmentos da extremidade 3' do genoma viral, uma estrutura de capa 5' e uma cauda 3' poliadenilada (A) (NAQVI *et al.*, 2020; DAVIDSON *et al.*, 2020). Os diferentes mRNAs subgenômicos codificam quatro proteínas estruturais conservadas, incluindo a proteína de pico (*Spike*), envelope (E), membrana (M) e nucleocapsídeo (N) e várias proteínas acessórias previstas na anotação atual de SARS-CoV-2 que incluem a ORF3a, ORF 6, ORF7a, ORF7b, ORF8 e ORF10 (FINKEL *et al.*, 2021).

A proteína *Spike* possui 1273 resíduos de aminoácidos (aa), sendo composta por duas subunidades: S1, que contém o domínio de ligação ao receptor (RBD) responsável pela interação com receptores da Enzima Conversora de Angiotensina 2 (ECA2) nas células hospedeiras, e S2, que medeia a fusão da membrana e a entrada viral (WALLS *et al.*, 2020; ROGERS *et al.*, 2020). Portanto, a *Spike* desempenha um papel fundamental na patogênese viral a partir da entrada do vírus nas células hospedeiras, além de apresentar potencial plasticidade (WALLS *et al.*, 2020; ARYA *et al.*, 2021).

A proteína N é a única proteína estrutural presente no nucleocapsídeo do genoma do SARS-CoV-2 e possui 419 resíduos de aminoácidos. Essa proteína contribui de forma significativa para o empacotamento do RNA viral no ribonucleocapsídeo helicoidal (RNP), também interagindo com as outras proteínas estruturais durante a montagem dos vírions e consequente encapsulação do genoma (KUO *et al.*, 2016). Além disso, a proteína N possui propriedades imunológicas únicas, o que a torna uma candidata potencial para o desenvolvimento

de vacinas (KHAN *et al.*, 2020). tem 419 aminoácidos (aa) de comprimento

A proteína E é a menor dentre as proteínas estruturais de SARS-CoV-2, apresentando apenas 75 resíduos de aminoácidos (WU *et al.*, 2020). Essa proteína está associada à montagem viral, brotamento, formação de envelope, no entanto os mecanismos moleculares que envolvem sua participação na patogênese viral ainda não são bem compreendidos (HOQUE *et al.*, 2020; RAHMAN *et al.*, 2021). Já a proteína M, compreende 22 resíduos de aminoácidos, sendo caracterizada como uma glicoproteína transmembrana tipo III, constituindo-se na proteína estrutural mais abundante, sendo importante para definir a forma do envelope viral e organização central da montagem do CoV, interagindo com todas as outras proteínas estruturais (NEUMAN *et al.*, 2011; THOMAS, 2020; ASTUTI, 2020; BIANCHI *et al.*, 2020).

O restante do genoma é constituído principalmente pelas proteínas acessórias. A anotação atual do SARS-CoV-2 inclui previsões de seis proteínas acessórias (ORF3a, ORF6, ORF7a, ORF7b, ORF8 e ORF10), no entanto, nem todas foram confirmados experimentalmente (NAQVI *et al.*, 2020, FINKEL *et al.*, 2021). Essas proteínas também desempenham importantes papéis no ciclo de vida viral e embora não afetem diretamente a liberação e estabilidade do vírus, podem contribuir para a patogênese e virulência ao contribuírem para a infecção em células hospedeiras (XU *et al.*, 2009). Estudos recentes demonstraram que as proteínas ORF8 e ORF10 de sequências correspondentes a SARS-CoV-2 são únicas em relação a proteínas codificadas pelo SARS-CoV. A análise estrutural da ORF8 mostrou que essa proteína pode formar conjuntos únicos em grande escala que não estão presentes no SARS-CoV (FLOWER *et al.*, 2021). Em contraste, estudos apontam que a ORF10 é a única proteína que não está presente em outros coronavírus humanos, sendo pouco explorada (DHAKAL *et al.*, 2020). Dessa forma, a pouca homologia de sequência das proteínas acessórias de SARS-CoV-2 com as de outros coronavírus, torna necessária a investigação de mutações nessas proteínas e seus impactos no vírus (HASSAN; CHOUDHURY; ROY, 2020).

2.4. VIGILÂNCIA GENÔMICA DE SARS-CoV-2

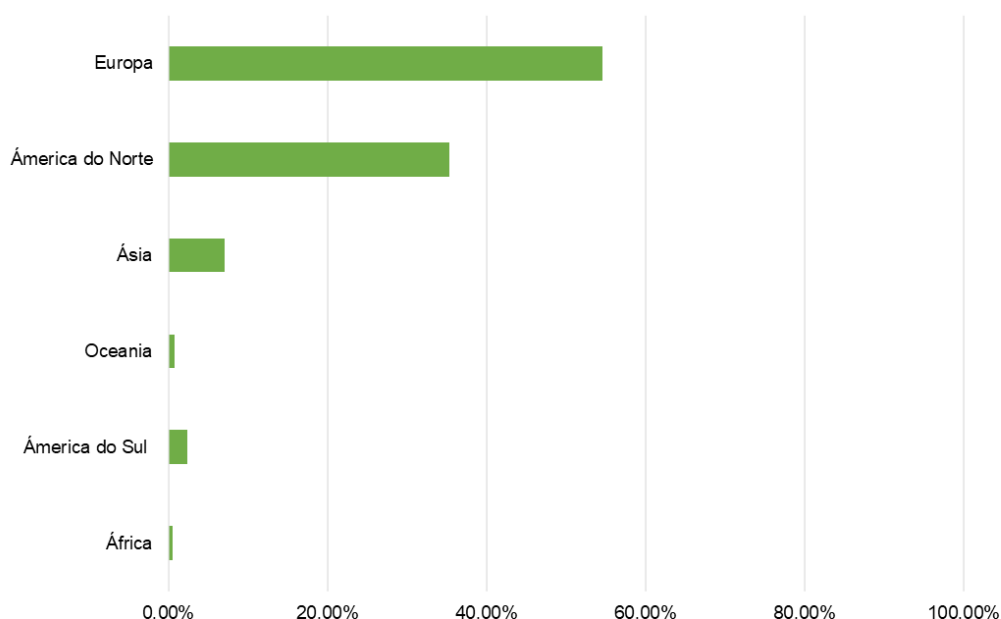
A evolução contínua de SARS-CoV-2 destaca a necessidade de fortalecimento dos sistemas de vigilância genômica para o controle da propagação da COVID-19 (BAKER *et al.*, 2021). Nessa perspectiva, desde sua identificação como agente causador da atual pandemia da COVID-19 (ZHU *et al.*, 2020), amostras de genoma viral de todo o mundo foram sequenciadas, permitindo reconstruções filogenéticas da propagação e evolução viral em diferentes países e continentes, bem como a identificação de novas variantes virais incluindo as variantes de preocupação (VOC, do inglês *Variant of Concern*) como a *alfa* (RAMBAUT *et al.*, 2020), *beta* (TEGALLY *et al.*, 2021), *gamma* (FARIA *et al.*, 2021), *delta* (SINGH; RAHMAN, 2021) e a recentemente surgida, *Omicron* (CALLAWAY, 2021), as quais carregam mutações que podem afetar a patogenicidade e infectividade do vírus, determinando, portanto, mudanças no fenótipo viral e aumentando facilidade do vírus se espalhar, além de impactar no desempenho dos testes de diagnóstico molecular e na eficácia de vacinas atualmente disponíveis (LAURING; HODCROFT, 2021; ZIEGLER *et al.*, 2020). A designação dessas variantes por letras do alfabeto grego, segue a recomendação do Grupo Técnico Consultivo sobre Evolução de SARS-CoV-2 da OMS, por apresentarem maior facilidade de serem discutidas por públicos não científicos (WHO, 2021c).

O principal repositório para sequências genômicas de SARS-CoV-2 é o *Global Initiative on Sharing All Influenza Data* (GISAID), o qual contava, até 29 de dezembro de 2021, com mais de 4.600.000 sequências genômicas completas e de alta qualidade (SHU; MCCAULEY, 2017). Esses dados genômicos têm sido importantes para fornecer informações sobre uma variedade de parâmetros epidemiológicos associados ao surgimento e origens geográficas das epidemias em escalas locais (FARIA *et al.*, 2021; PAIVA *et al.*, 2020; FRANCESCHI *et al.*, 2021), continentais (WOROBAY *et al.*, 2020; MUÑOZ *et al.*, 2021) e até mesmo em termos globais (VAN DORP *et al.*, 2020).

No entanto, a maioria das sequências disponíveis é oriunda de países da Europa (54,66%) e América do Norte (35,38%), conforme descrito na Figura 3, destacando-se as massivas contribuições do Reino Unido e dos Estados Unidos (FURUSE, 2021). Em contraste, países da América do Sul ainda apresentam níveis de amostragem pouco representativos (BOTELHO-SOUZA *et al.*, 2021). Além disso,

as informações genômicas obtidas através de sequenciamento estão mal distribuídas, o que pode levar a interpretações errôneas em estudos de epidemiologia genômica, dificultando a compreensão da disseminação e evolução do SARS-CoV-2 nesta região (WOROBEY *et al.*, 2020).

Figura 3: Distribuição de sequências genômicas disponíveis por continente de amostragem entre janeiro de 2020 e dezembro de 2021. Dados obtidos do GISAID (<https://www.gisaid.org>).



Fonte: Silveira (2021).

2.5 SEQUENCIAMENTO DE PRÓXIMA GERAÇÃO (NGS) E FERRAMENTAS DE BIOINFORMÁTICA PARA PESQUISA E VIGILÂNCIA GENÔMICA DE SARS-COV-2

O avanço das tecnologias de sequenciamento de próxima geração (NGS, do inglês *next generation sequencing*), também denominado de sequenciamento de alto rendimento, permitiu o uso dessa abordagem na pesquisa de vírus, especialmente nas áreas de diagnóstico de doenças infecciosas e evolução viral. Desse modo, observou-se um notável aumento do volume de dados de sequências genômicas disponíveis, os quais possibilitaram, por conseguinte uma melhor compreensão da composição genética desses patógenos e interpretação de sua epidemiologia

evolutiva (HU *et al.*, 2021; BHOYAR *et al.*, 2021).

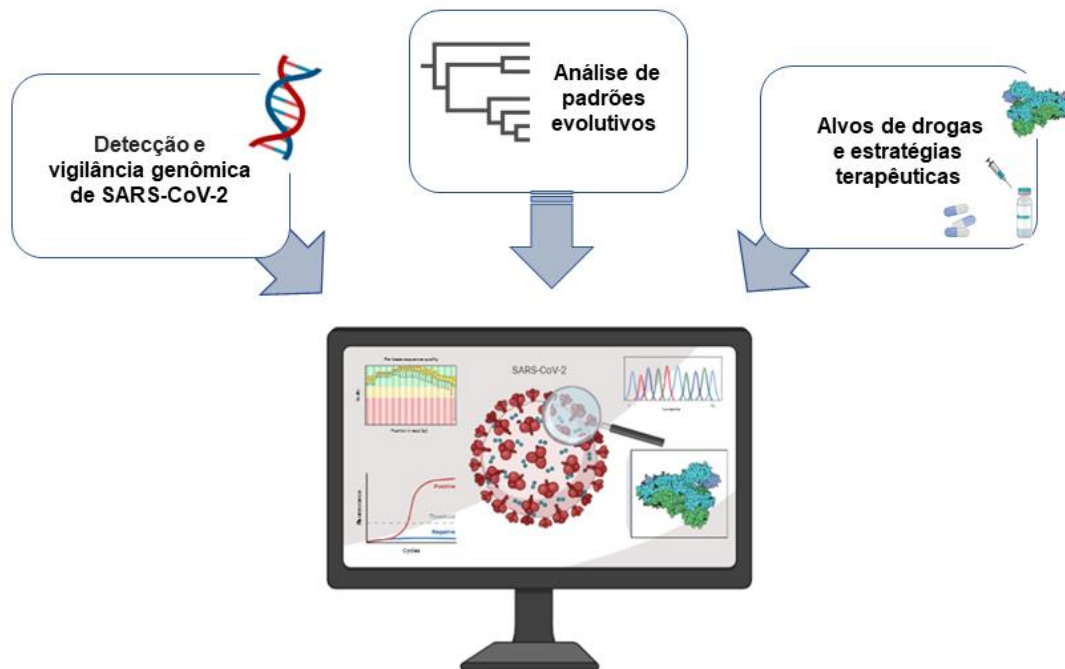
Desde o surgimento de SARS-CoV-2, várias estratégias de sequenciamento foram aplicadas para descobrir e monitorar o agente causador da pandemia da COVID-19 (RAY *et al.*, 2020). Durante os estágios iniciais do surto, muitos genomas de SARS-CoV-2 foram obtidos usando sequenciamento metagenômico (ZHOU *et al.*, 2020). Posteriormente, passou-se a utilizar a tecnologia de amplificação em cadeia da polimerase multiplex, onde o RNA total é transcrito reversamente para sintetizar cDNA e uma PCR é executada usando múltiplos pares de iniciadores de amplificação, seguido por reação de ligação para adicionar os indexadores (HU *et al.*, 2021). As bibliotecas podem ser subsequentemente processadas utilizando as plataformas de sequenciamento, sendo as mais usuais as da Illumina e da Oxford Nanopore MinION (XIAO *et al.*, 2020; PADEN, *et al.*, 2020).

Como sabemos, o sequenciamento genômico gera uma grande quantidade de dados brutos que requerem quantidades substanciais de processamento computacional para resultar em relatórios científicos (JOHN *et al.*, 2021). Dessa forma, precisam ser aplicados fluxos de trabalho com uso de ferramentas de bioinformática, as quais se referem a uma série de algoritmos pelos quais os dados brutos são processados, gerando dados passíveis de interpretação (VANDERBILT; MIDDHA, 2020). Esses fluxos de trabalho incluem recursos para geração de sequências, montagem e alinhamento, metodologia de controle de qualidade rigorosa, identificação e anotação de variantes, assim como análises de genômicas comparativa e evolutivas (ROY *et al.*, 2018; HU *et al.*, 2021).

Além das abordagens já existentes, a pandemia da COVID-19 impulsionou o desenvolvimento de ferramentas específicas para o estudo de SARS-CoV-2, as quais têm sido aplicadas em diversas áreas de pesquisa relacionadas a esse patógeno (HUFSKY *et al.*, 2021). Elas incluem estratégias de detecção, vigilância genômica e estudos epidemiológicos, com foco no rastreamento da propagação e análise de padrões evolutivos em escala global (BAUER *et al.*, 2020; CANNATARO; HARRISON, 2021). Dentre estes podemos citar várias plataformas interativas como o Coronapp, Nextclad, CoV-Spectrum, COVID-19 CG, que se constituem em ferramentas para a anotação de sequências genômicas virais fornecidas pelo usuário, verificações de qualidade, atribuição de clados e chamada e rastreamento de variantes (MERCATELLI *et al.*, 2021; AKSAMENTOV *et al.*, 2021; CHEN C. *et al.*, 2021; CHEN A. *et al.*, 2021). Além disso, diversos estudos estão sendo

desenvolvidos visando a descoberta de potenciais alvos de drogas e no desenvolvimento de estratégias terapêuticas (WAMAN *et al.*, 2021) (Figura 4).

Figura 4: Principais aplicações das ferramentas de bioinformática no estudo de SARS-CoV-2



Fonte: Silveira (2021), elaborada com a ferramenta BioRender.

3 RESULTADOS E DISCUSSÃO

3.1 ARTIGO

Altos níveis de diversidade genômica em cepas SARS-CoV-2 circulantes no Estado de Pernambuco (Região Nordeste do Brasil) durante a “primeira onda” da COVID-19

Zildene de Sousa Silveira¹, Marcos da Silveira Regueira Neto², Wilson José da Silva Junior², Heidi Lacerda Alves da Cruz², Bruno Sampaio², Jussara Moura de Lima², Moisés Thiago de Souza Freitas³; Sérgio de Sá Paiva Leitão Junior⁴, Francisco Assis Bezerra da Cunha⁵, Valdir de Queiroz Balbino^{1,2}

¹Programa de Pós-graduação em Ciências Biológicas - PPGCB; Universidade Federal de Pernambuco -UFPE, Recife, Brasil

²Departamento de Genética, Universidade Federal de Pernambuco -UFPE, Recife, Brasil

³ Núcleo de Saúde, Centro Universitário do Vale do Ipojuca - UNIFAVIP, Caruaru, Pernambuco, Brasil

⁴Unidade Acadêmica Serra Talhada, Universidade Federal Rural de Pernambuco, Serra Talhada, Pernambuco, Brasil

⁵Universidade Regional do Cariri, Crato, Ceará, Brasil

RESUMO/ABSTRACT

Síndrome Respiratória Aguda Grave coronavírus 2 (SARS-CoV-2) é responsável pela pandemia da doença do coronavírus-2019 (COVID-19), a qual levou à notificação de mais de 200 milhões de casos em todo o mundo. Desde então, a vigilância genômica tem sido uma ferramenta fundamental para rastrear a disseminação viral. Neste estudo fornecemos uma análise detalhada da dinâmica temporal de mutações e linhagens de SARS-CoV-2 durante a primeira onda da pandemia no estado de Pernambuco, Brasil. Para isso, realizamos a caracterização de 67 genomas obtidos a partir de swabs nasofaríngeos e orofaríngeos de pacientes positivos para COVID-19 de 22 municípios, os quais foram combinados com 293 sequências de Pernambuco disponíveis no GISAID

obtidas entre fevereiro e dezembro de 2020, período adotado como marco temporal da primeira onda pandêmica no estado de Pernambuco. Descreve a disseminação temporal e a evolução do SARS-CoV-2 durante a primeira onda pandêmica em Pernambuco, sendo detectados 962 SNPs em regiões codificantes, a maioria deles localizados na ORF1ab. Verificamos que as mutações C241T, C3037T (F924F), C14408T (P4715L/P323L), A23403G (D614G), G28881A (R203K), G28882A (R203R) e G28883C (G204R), mantiveram suas frequências relativas acima de 94% durante todo o período de amostragem. Foi detectada a circulação simultânea de 18 linhagens em Pernambuco, sendo as mais prevalentes B.1.1, B.1.1.28, B.1.1.33 e P.2, respectivamente. A rede de haplótipos e a análise filogenômica mostraram que as sequências se agruparam em 4 clados principais de acordo com a dinâmica evolutiva de SARS-CoV-2, tendo havido importações iniciais de linhagens com provável origem da Europa, seguida por expansão e transmissão local de linhagens nacionais, contribuindo para o crescimento acelerado da epidemia. Esses resultados evidenciam elevados níveis de diversidade genômica em Pernambuco e destacam a importância de investigações contínuas em vigilância genômica para rastreamento de variantes virais.

Palavras-chave/Key words: SARS-CoV-2. Pernambuco. Vigilância genômica. Filogenômica.

INTRODUÇÃO

O surgimento da doença do coronavírus-2019 (COVID-19) e sua rápida disseminação por mais de 190 países, levou a notificação de mais 281 milhões de casos e 5 milhões de mortes em todo mundo (WHO, 2021a), impondo declínio aos sistemas econômicos e de saúde mundiais (ONU, 2020a, 2020b). No Brasil, a primeira infecção por SARS-CoV-2 foi relatada em 26 de fevereiro de 2020 (OPAS, 2020). Desde então, observou-se aumento exponencial do número de casos de COVID-19, sendo notificados até 29 de dezembro de 2021, 22.222.928 casos e 618.091 mortes, constituindo-se o terceiro país em número de casos confirmados no mundo (MINISTÉRIO DA SAÚDE, 2021).

O agente etiológico da COVID-19 foi identificado como sendo um Betacoronavirus (família *Coronaviridae*), designado de Síndrome Respiratória Aguda

Grave Coronavírus 2 – SARS-CoV-2 (ZHANG & HOLMES 2020; WU *et al.*, 2020). Trata-se de um vírus envelopado com um genoma de RNA de fita simples de sentido positivo, com cerca de 30 kb (CHEN, LIU & GUO., 2020). Seu genoma consiste em 11 regiões codificantes, das quais cerca de dois terços abrangem a ORF1ab que dá origem a 16 proteínas não estruturais (nsp). As ORFs restantes codificam para quatro proteínas estruturais, nomeadamente: glicoproteína de pico (S), proteína da matriz (M), proteína do envelope (E) e proteína do nucleocapsídeo (N). Além disso, a anotação atual do SARS-CoV-2 inclui previsões de seis proteínas acessórias (3a, 6, 7a, 7b, 8 e 10, NC_045512.2) (ZHU *et al.* 2020; FINKEL *et al.*, 2021).

A disponibilidade de milhares de sequências genômicas de SARS-CoV-2 e de ferramentas de visualização e análise de dados permitiu traçar inferências sobre a dinâmica evolutiva do vírus, levando à descrição de mais de 1.500 linhagens desde o início da pandemia (RAMBAUT *et al.* 2020). O principal serviço de compartilhamento de dados de sequências de SARS-CoV-2 é o *Global Initiative on Sharing All Influenza Data* (GISAID), o qual contava com mais de quatro milhões de sequências completas até 29 de dezembro de 2021 (SHU & MCCAULEY, 2017). De modo geral, Estados Unidos, Reino Unido e Dinamarca representam os países com maiores esforços de sequenciamento para SARS-CoV-2 (FURUSE, 2021; CHEN *et al.*, 2021). No Brasil, em particular, a vigilância genômica está muito inferior a ideal, com apenas 74.696 genomas completos e de alta cobertura, os quais apresentam uma má distribuição entre seus estados.

Em Pernambuco, estado do nordeste brasileiro, desde a notificação do primeiro caso de COVID-19 em 12 de março de 2020, foram registrados 644.852 casos e 20.431 óbitos, constituindo-se o décimo estado brasileiro em número de infectados e o nono em mortes (MINISTÉRIO DA SAÚDE, 2021; 29 de dezembro de 2021). Até o momento, a pandemia da COVID-19 em Pernambuco foi caracterizada por maior concentração dos casos na capital do estado, seguida por intensa disseminação por seu interior (CARMO *et al.*, 2021). Os dados gerados sugeriram introduções da Europa, bem como a prevalência da linhagem B.1.1 dentre os genomas disponíveis (PAIVA *et al.*, 2020). No entanto, a amostragem de dados genômicos no estado ainda é limitada, o que impede uma investigação genômica abrangente da dinâmica da primeira onda da pandemia a nível local. A forma como as mutações se distribuíram e o seu impacto na diversidade genética do SARS-CoV-2 ao longo da primeira onda são questões que ainda precisam ser respondidas.

Com o objetivo de preencher lacunas em relação à primeira onda de COVID-19 em Pernambuco, realizamos uma análise temporal da variação de mutações e linhagens em circulação no estado. Também analisamos a dinâmica evolutiva de SARS-CoV-2 a partir de amostras obtidas entre fevereiro e dezembro de 2020, período adotado como marco temporal da primeira onda pandêmica no estado de Pernambuco.

MATERIAIS E MÉTODOS

Amostras de pacientes e detecção molecular de SARS-CoV-2

Para este estudo, foram sequenciadas 86 amostras obtidas a partir de *swabs* nasofaríngeos e orofaríngeos de pacientes positivos para COVID-19 de diferentes municípios do Estado de Pernambuco, Brasil. As amostras de *swab* foram encaminhadas ao Núcleo de Pesquisa em Inovação Terapêutica - Suely Galdino (NUPIT-SG), na Universidade Federal de Pernambuco (UFPE), pelo Laboratório Central de Saúde Pública do Estado de Pernambuco (LACEN-PE) para o diagnóstico molecular da COVID-19.

As extrações de RNA foram realizadas utilizando o kit Maxwell® 48 Viral Total *Nucleic Acid Purification Kit* (Promega, Wisconsin, USA), seguindo as instruções do fabricante. O diagnóstico molecular foi realizado utilizando o kit GoTaq Probe 1 Step RT-qPCR (Promega, Wisconsin, USA) para a detecção do gene do nucleocapsídeo (N) do vírus e do gene controle RNase P humano, de acordo com o *Center for Disease Control and Prevention* (CDC, Atlanta, GA).

Sequenciamento Genômico

As amostras de RNA total de pacientes positivos para COVID-19 foram cedidas ao Laboratório de Bioinformática e Biologia Evolutiva da UFPE (LABBE-UFPE) para o sequenciamento dos genomas completos; 100 ng de RNA total foram usados para a produção de cDNA utilizando o kit *Ampliseq™ cDNA Synthesis for Illumina* (Illumina, California, USA), seguindo as instruções do fabricante. O cDNA obtido foi submetido a uma PCR multiplex usando o kit *AmpliSeq for Illumina SARS-CoV-2 Research Panel* (Illumina, Califórnia, USA), o qual disponibiliza 247 pares de primers que cobrem 237 regiões alvo do SARS-CoV-2, seguindo as instruções do fabricante. As bibliotecas contendo as sequências de SARS-CoV-2 foram

preparadas com o *AmpliSeq Library PLUS* (Illumina, California, USA). As bibliotecas foram purificadas usando *AMPure XP Beads* (Beckman Coulter) seguindo as orientações do *kit AmpliSeq da Illumina* e foram quantificadas usando *Qubit® dsDNA HS Assay Kits* (Invitrogen), seguindo as instruções do fabricante. O sequenciamento *paired-end* foi feito na plataforma MiSeq (Illumina).

Montagem dos genomas de SARS-CoV-2

Foi utilizado o FastQC (ANDREWS, 2010) nos dados brutos (*.fastq*) para avaliar a qualidade, o comprimento do fragmento e a presença de adaptadores. Em seguida, os adaptadores foram removidos usando CUTADAPT (MARTIN, 2011). As leituras obtidas foram filtradas com o software Sickle (JOSHI; FASS, 2011), utilizando como padrão comprimento e qualidade do fragmento >20. Posteriormente, os genomas foram montados com o software BWA (LI & DURBIN, 2010), usando a sequência do genoma NC_045512.2 como referência. O software Samtools (LI *et al.*, 2009) foi utilizado para converter os dados, remover duplicatas, indexar e, juntamente com BCFTools (DANECEK & MCCARTHY, 2017) e VCFtools (DANECEK *et al.*, 2011), geraram o genoma consenso para cada amostra. Foi realizada a avaliação da qualidade da montagem do genoma pelo QUAST (GUREVICH *et al.*, 2013) e genomas com cobertura <95% e >1% de Ns (quando comparados com a sequência NC_045512.2) foram removidos.

Aquisição dos dados

Para se obter um melhor entendimento da dinâmica de disseminação de SARS-CoV-2 em Pernambuco, combinamos as sequências genômicas obtidos neste estudo com 293 sequências amostradas em Pernambuco disponíveis no GISAID, nos concentrando nos genomas coletados entre 26 de fevereiro a 31 de dezembro de 2020 (recuperados em 04 de agosto de 2021) e aplicando-se os filtros (*complete* e *high coverage*). Os genomas foram então alinhados ao genoma de referência SARS-CoV-2 (NC_045512.2) usando MAFFT v.7 com os parâmetros default (KATO; STANDLEY, 2013).

Os dados epidemiológicos referentes ao tamanho da população, número de casos e óbitos por semana epidemiológica para Pernambuco foram recuperados do COVID-19 Painel Coronavírus (MINISTÉRIO DA SAÚDE BRASIL., 2021;

<https://covid.saude.gov.br/>). Os indicadores epidemiológicos foram obtidos de acordo com a Organização Mundial de Saúde, onde a incidência foi calculada como o número de casos por 100.000 habitantes, e a mortalidade, como o número de mortes por 100.000 habitantes (BONITA; BEAGLEHOLE; KJELLSTRÖM, 2006).

Análises populacionais

A diversidade genética intrapopulacional, as taxas de transições e transversões, o número de sítios polimórficos e singletons foram obtidos usando o software DNAsp v. 6 (ROZAS *et al.*, 2017). Os testes de neutralidade de D de Tajima's e Fs de Fu foram estimados para os genomas obtidos neste estudo, bem como para os recuperados do GISAID, a partir do software Arlequin v.3.5 (EXCOFFIER & LISCHER, 2010).

Chamada de variantes e classificação de linhagens

As sequências foram alinhadas com o genoma de referência (NC_045512.2) usando Bowtie2 (LANGMEAD & SALZBERG, 2012). Os arquivos mapeados foram convertidos de *sam* para *bam* usando o Samtools (LI *et al.*, 2009). A chamada de variantes foi realizada usando Freebayes versão 1.0.2 (GARRISON & MARTH, 2012). Os VCFs foram gerados e mesclados usando o VCFtools (DANECEK *et al.*, 2011) e anotados usando SnpEff (v4.5covid19) disponível na plataforma web Galaxy (CINGOLANI *et al.*, 2012). Os SNPs foram classificados como de: baixo impacto, para mutações sinônimas na região codificadora, incluindo os códons de início e parada; impacto moderado, para missense; alto impacto, para mutações não sinônimas que afetam locais de emendas, incluindo os códons de início e parada; e modificador, para mutações na região não codificadora (*upstream*, *downstream*, intergênicas e UTR).

A atribuição de linhagens foi realizada pelo método de classificação proposto por Rambaut *et al.* (2020), por meio do pacote de software *Phylogenetic Assignment of Named Global Outbreak Lineages* (PANGOLIN v3.1.14, 2021-10-13) (O'TOOLE *et al.*, 2021).

Análise filogenômica

Para análise filogenômica, além dos 293 genomas completos de SARS-CoV-2 de Pernambuco obtidos do banco de dados GISAID entre 26 de fevereiro e 31 de dezembro de 2020 e os genomas adicionais para o estado de Pernambuco obtidos neste estudo, foram recuperadas do GISAID todas as sequências brasileiras e metadados associados cuja amostragem se deu entre 26 de fevereiro e 31 de dezembro. Também foram recuperadas sequências genômicas completas e metadados para as linhagens B.1, B.1.1 da América do Sul e as linhagens B.1.1.28 e B.1.1.33 de outros países, além do Brasil que apresentaram ocorrências dessas linhagens. Esse conjunto de dados foi submetido a uma filtragem, onde as sequências com mais de 300 Ns (<10% de N) foram removidas do conjunto de dados.

Em seguida, devido ao grande número de sequências disponíveis, um *script* em Python foi escrito para conduzir uma subamostragem de um conjunto representativo de genomas por linhagem, exceto para as sequências geradas neste estudo. Para o conjunto de dados final, as sequências foram obtidas através de um processo de reamostragem (-r = 1000) onde foram selecionadas as sequências mais frequentes que apareceram em cada rodada de amostragem, gerando um conjunto de dados final com 10% das sequências por linhagem. Após a conclusão do processo de subamostragem incluímos no conjunto de dados, as sequências obtidas neste estudo (n = 67).

Por fim, o conjunto final de dados genômicos (n= 630) foi submetido a uma análise filogenômica de máxima verossimilhança (ML) usando o pipeline do Nextstrain (versão 3.0.1) (<https://nextstrain.org/>) (HADFIELD *et al.*, 2018). Resumidamente, esse pipeline usa o kit de ferramentas *augur* para processar os dados em várias etapas. Inicialmente, o MAFFT v.7 foi usado para o alinhamento de múltiplas sequências (KATO; STANDLEY, 2013). Na reconstrução filogenética, o *augur tree* utiliza o software IQ-Tree (V. 2.1.1 COVID-edition) com o modelo de substituição *General Time Reversible Model* (GTR) (NGUYEN *et al.*, 2015). Em seguida, o TreeTime (V.0.8.2) (SAGULENKO; PULLER & NEHER, 2018) foi utilizado para inferir árvores evolutivas resolvidas no tempo e reconstruir sequências ancestrais e mutações. Por fim, os resultados foram exportados para o formato

JSON para permitir a visualização interativa por meio do Auspice (<https://auspice.us/>).

Rede de Haplótipos

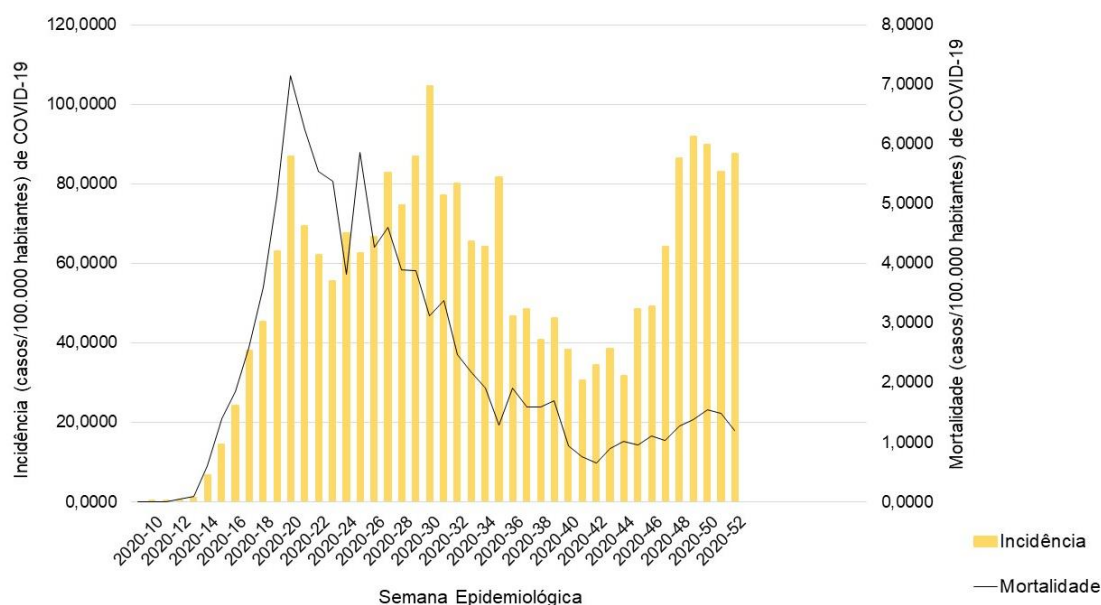
As 630 sequências alinhadas foram importadas para o SequenceMatrix (VAIDYA; LOHMAN & MEIER, 2011) para gerar o arquivo “.nexus”, sendo em seguida construída uma rede de haplótipos pelo método *median joining network*, usando o software *Population Analysis with Reticulate Trees* (POPART v. 1.7), a fim de determinar o número e abundância de haplótipos na população e avaliar suas relações evolutivas (LEIGH & BRYANT, 2015).

RESULTADOS

Caracterização epidemiológica da COVID-19 no estado de Pernambuco

Em Pernambuco foram observados picos na incidência da COVID-19 nas semanas epidemiológicas 21^a (17 a 23 de maio de 2020) e 31^a (26 de julho a 1 de agosto). Também foram registrados picos na mortalidade durante a 21^a (17 a 23 de maio de 2020) e 22^a (24 a 30 de maio de 2020) semanas epidemiológicas, onde foram registradas taxas de mortalidade de aproximadamente 7,15 e 6,24, respectivamente. A partir da 32^a semana epidemiológica, houve acentuada queda nas curvas de incidências e mortalidade da COVID-19 em Pernambuco. No entanto, foi verificada nova elevação nos indicadores a partir da 46^a semana epidemiológica, entre 08 e 14 de novembro de 2020 (Figura 1).

Figura 1: Indicadores epidemiológicos da COVID-19 no estado de Pernambuco durante a primeira onda da pandemia. Incidência (colunas amarelas) e mortalidade (preto) por 100.000 habitantes por semana epidemiológica no estado de Pernambuco. A semana epidemiológica 10 refere-se à primeira semana de março de 2020, semana epidemiológica 11 inclui o primeiro caso oficial notificado em Pernambuco em 12 de março e a semana epidemiológica 53 refere-se à última semana



epidemiológica de 2020.

Fonte: Silveira (2021)

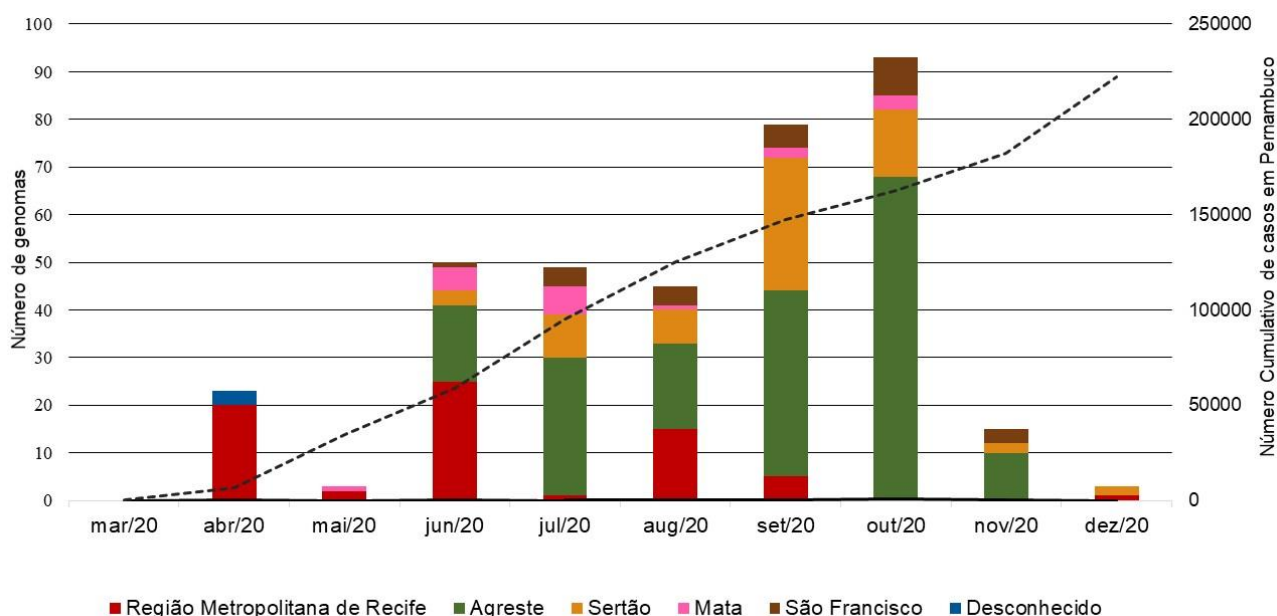
Amostragem e aquisição de dados

Nesse estudo, sequenciamos genomas completos de 86 amostras de SARS-CoV-2 coletadas em diferentes regiões do estado de Pernambuco, Brasil, entre 07 de julho e 07 de novembro de 2020 (Apêndice A). A montagem desses genomas foi realizada utilizando a sequência de referência de Wuhan-Hu-1 (NC_045512.2); após a aplicação dos critérios de qualidade (comprimento dos *reads*, cobertura acima de 95% e profundidade de cobertura dos genomas e quantidade de Ns), 67 sequências genômicas foram selecionadas para o seguimento do estudo. São sequências provenientes de 22 municípios de quatro mesorregiões do estado de Pernambuco, sendo distribuídas: Agreste (n=25), São Francisco (n=9), Sertão (n=27), Mata Pernambucana (n=6). Os genomas selecionados apresentaram frações de genomas sequenciados indo de 99,58 até 99,96% e comprimento total indo de 29779 a 29891 pb.

Em seguida, obedecendo os parâmetros de qualidade, combinamos este conjunto de dados com 293 sequências genômicas adicionais de Pernambuco

disponíveis no GISAID, referentes às amostras submetidas até 04 de agosto de 2021 e coletadas entre 26 de fevereiro e 31 de dezembro de 2020. O conjunto de dados resultante incluiu 360 sequências de cinco mesorregiões do estado, sendo 180 do Agreste (50%), 69 (19,17%) da Região Metropolitana de Recife, 65 (18,05%) do Sertão pernambucano, 25 (6,95%) da Região do São Francisco, 18 (5%) da Mata Pernambucana e três (0,83%) de origem não identificada (Figura 2). A maioria das sequências correspondentes ao período da primeira onda da COVID-19 foram coletadas nos meses de outubro, setembro e junho de 2020, respectivamente. Portanto, embora a cobertura seja desigual entre as mesorregiões, este conjunto de dados é representativo da pandemia em Pernambuco em termos de distribuição geográfica.

Figura 2: Períodos de coleta dos 360 genomas coletados em Pernambuco. O número de genomas é colorido por mesorregião de amostragem em cada mês de coleta (eixo esquerdo). A linha tracejada mostra o número cumulativo de casos confirmados em Pernambuco por mês (eixo direito).



Fonte: Silveira (2021)

Diversidade genômica e análise de variantes

As 360 sequências genômicas de SARS-CoV-2 isoladas de cepas pernambucanas (este estudo + GISAID) foram analisadas no DNAsp com o objetivo de acessar métricas básicas de diversidade nucleotídica (Tabela 1). A diversidade

nucleotídica (*average over loci*) foi de 0,00048, enquanto os valores dos testes de D de Tajima e Fs de Fu foram de -2,80229 e -23,64854, respectivamente.

Tabela 1: Sítios variáveis e composição nucleotídica dos genomas sequenciados para Pernambuco, Brasil

	SV	PI	S	TS	TV	CN			
Pernambuco	1002	396	605	715	287	A:	29.88%	C:	18.36%
						T:	32.14%	G:	19.62%

SV: Sítios variantes; PI: Sítios Parcimônio informativos; S: Singletons; TS: singletons; TV: transversões; CN: Composição nucleotídica

Fonte: Silveira (2021)

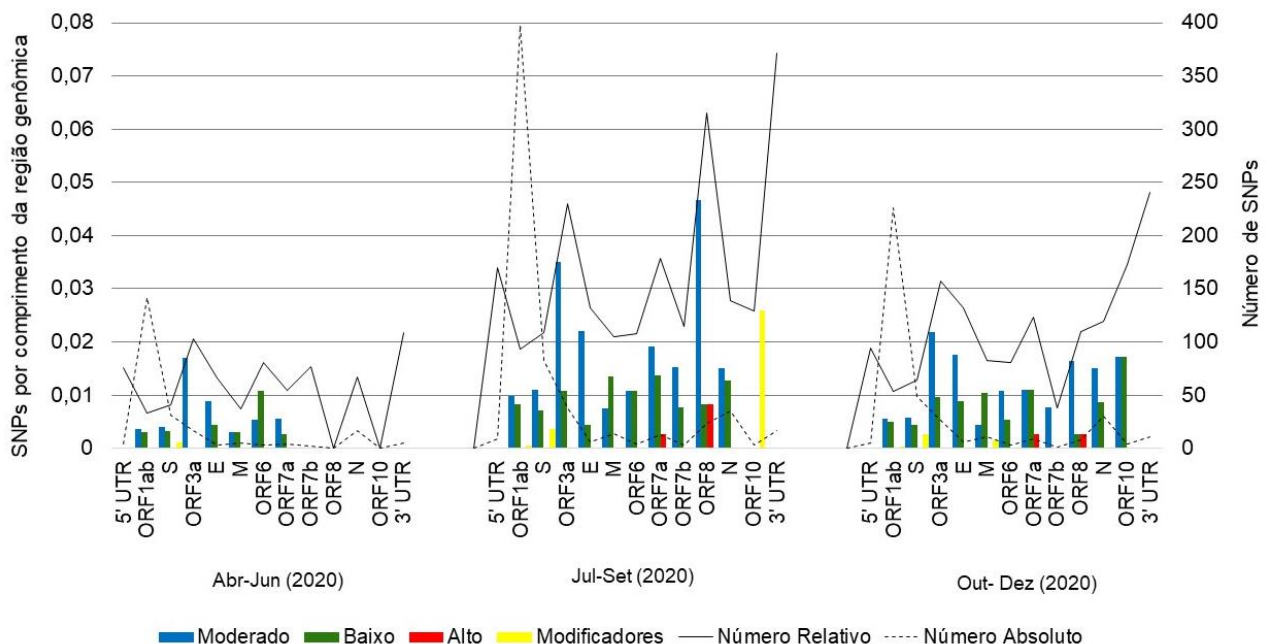
A partir da análise de chamadas variantes do conjunto de dados total, foi possível observar um total de 1002 SNPs, sendo que 962 (96,00%) das substituições nucleotídicas estão localizadas em regiões codificantes do genoma viral e 40 (4,00%) mutações presentes em regiões não-codificantes (UTR). Do total de SNPs mapeados em regiões codificantes, 552 (57,40%) consistem em mutações missense, 405 (42,1%) em mutações sinônimas, 4 (0,4%) e 1 (0,1%) para mutações que configuraram em ganho e perda de códon de parada, respectivamente.

Devido à disseminação de infecções por SARS-CoV-2 no Estado de Pernambuco, padrões dinâmicos de diversidade genômica foram observados ao longo do tempo (Figura 3). Para melhor caracterização desses padrões, dividimos os dados em três períodos de amostragem durante a primeira onda da pandemia (abril-junho; julho-setembro e outubro-dezembro). A partir disso, observamos um aumento do número de SNPs a partir do período de julho a setembro de 2020, com um total de 620 SNPs, em contraste com 224 e 373 SNPs nos períodos de abril a junho e outubro a dezembro, respectivamente. Essa variação também foi acompanhada de um aumento gradual do número de SNPs em todas as regiões genômicas a partir de julho de 2020. Dentre as regiões codificantes, a ORF1ab apresentou o maior número de mutações nos três períodos de amostragem (63,4%; 64,2% e 60,6%). Devido ao seu comprimento, a densidade de mutações se torna relativamente baixa. Desse modo, os padrões temporais de distribuição das mutações foram diferentes em relação ao número relativo de SNPs quando se considera o comprimento do gene,

com maior acúmulo de mutações na ORF3a, ORF6 e ORF7b, no período de abril a junho; na ORF8, ORF3a e ORF7a de julho a setembro; e na ORF10, ORF3a e gene E de outubro a dezembro.

De modo similar, a proporção de mutações por categoria de impacto também apresentou variações ao longo do tempo. No período de abril a junho, maior número de mutações de baixo impacto foram encontradas na ORF6, moderadas na ORF3a e modificadoras no gene S, não sendo observadas mutações de alto impacto nesse período. Para o período de julho a setembro, a ORF7a incorporou o maior número de mutações de baixo impacto, a ORF8, de mutações de moderado e alto impacto e a ORF10, maior número de mutações modificadoras. Já no período final da primeira fase da pandemia (outubro a dezembro), foi observado maior número de mutações de baixo impacto para a ORF10, de impacto moderado para a ORF3a, de alto impacto para a ORF7a e ORF8 e modificadoras para o gene S.

Figura 3: Alteração temporal na prevalência de mutações durante a primeira onda da pandemia em Pernambuco. O tracejado mostra o número absoluto de SNPs por região no genoma do SARS-CoV-2 (eixo direito). A linha contínua mostra o número relativo de SNPs em relação ao comprimento de cada região genômica (eixo esquerdo). As colunas representam a classificação dos SNPs quanto ao impacto no genoma de SARS-CoV-2 e sua distribuição de acordo com o comprimento em número de bases da região genômica.



Fonte: Silveira (2021)

Quando se estabeleceu uma comparação entre os três conjuntos de dados, observamos que 24 SNPs foram compartilhados entre as sequências de

Pernambuco obtidas nos três períodos de amostragem, tendo as mutações C241T, C3037T (F924F), C14408T (P4715L/P323L), A23403G (D614G), G28881A (R203K), G28882A (R203R) e G28883C (G204R), mantido frequências relativas constantes > 94% (APÊNDICE B). Em contraste, as alterações nas posições nucleotídicas C12053T (L3930F) e G25088T (V1176F) apresentaram aumentos significativos de suas frequências a partir do período de julho a setembro de 2020. Também verificamos a emergência em Pernambuco das mutações: C100T, T10667G (L3468V), C11824T (I3853I), A12964G (G4233G), G23012A (E484K), G28628T (A119S), G28975T (M234I) e C29754T, a partir do período decorrente entre outubro e dezembro de 2020.

Classificação de linhagens de SARS-CoV-2

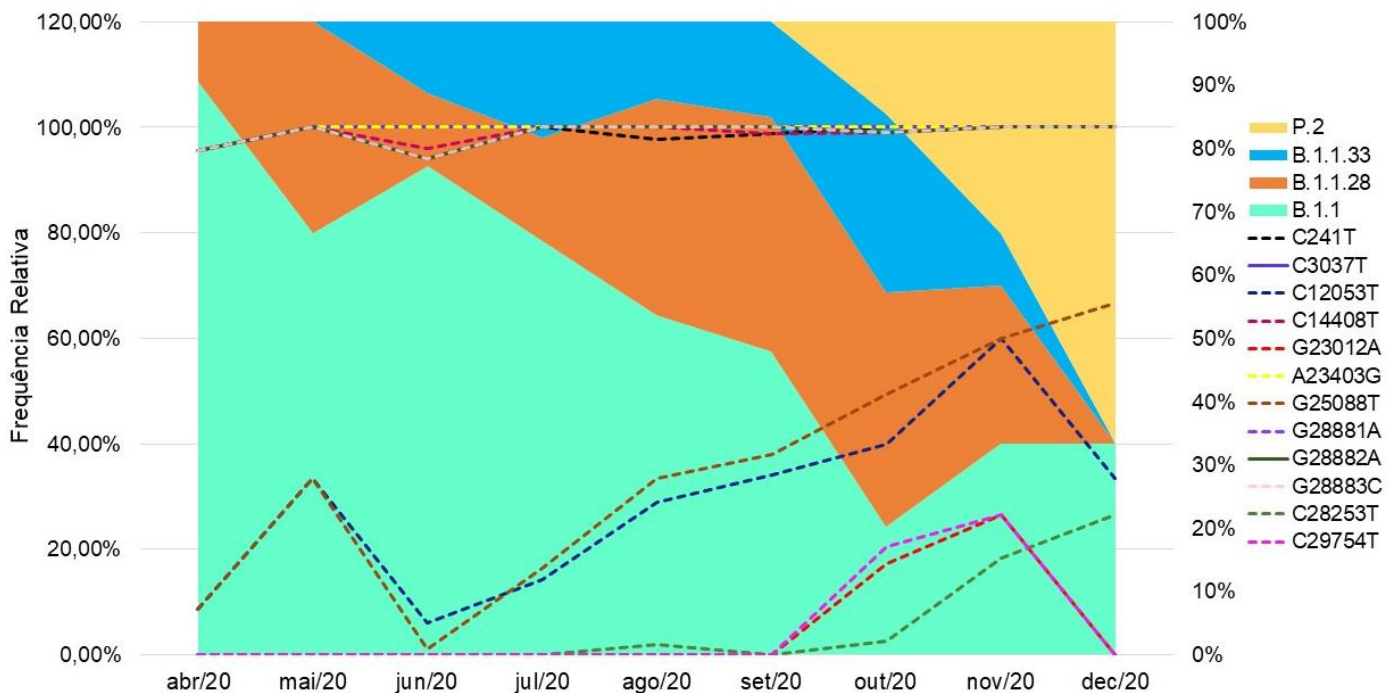
As 360 sequências obtidas neste estudo, mais as recuperadas do GISAID para Pernambuco, foram classificadas de acordo com as linhagens determinadas pelo software PANGOLIN. Foi verificado que os genomas de Pernambuco obtidos neste estudo agrupam sete linhagens, sendo B.1.1.28 a mais frequente com 30 sequências, seguido por B.1.1, B.1.1.33 e P.2, com 17, 8 e 7 sequências, respectivamente. As linhagens B.1.173 (n= 2), B.1.1.192 (n= 1) e B.1.1.371 (n= 2) também foram observadas entre nossas amostras. Quando avaliamos a representatividade de linhagens a partir de todas as sequências de Pernambuco, foi observada a circulação simultânea 18 linhagens, sendo a mais prevalente B.1.1 com 166 sequências, seguido por B.1.1.28, B.1.1.33 e P.2 com 93, 56 e 19 sequências, respectivamente. Além disso, o maior número de linhagens se agrupou nas mesorregiões do Agreste e Região Metropolitana, ambos com 11 linhagens circulantes.

As frequências de cada uma das linhagens verificadas em circulação em Pernambuco diferiram ao longo do tempo (Figura 4). Observamos que a linhagem B.1.1, contendo 166 genomas, representou consistentemente > 65% dos genomas amostrados entre abril e julho de 2020, tendo essa proporção caído a partir de agosto. A partir desse ponto, notamos um aumento significativo na frequência da linhagem B.1.1.28, próximo ao surgimento da linhagem P.2 em outubro de 2020 em nosso conjunto de dados.

Estabelecendo uma correlação entre os padrões dinâmicos de variantes genômicas e linhagens de SARS-CoV-2 ao longo do tempo em Pernambuco,

notamos que as variantes nas posições C241T, C3037T, C14408T, A23403G, G28881A, G28882A e G28883C mantiveram altas frequências durante todo o período de amostragem, o que se justifica pelo aparecimento dessas mutações na fase inicial da pandemia e por serem compartilhadas por linhagens que possuem B.1 e B.1.1 como linhagens parentais (Figura 4). Em contraste, observamos aumentos significativos da frequência das mutações C12053T e G25088T a partir de agosto de 2020, concomitantemente com o aumento da frequência de genomas classificados para linhagem B.1.1.28. Nossos dados também mostram o aumento da frequência das mutações G23012A, C28253T e C29754T a partir de outubro de 2020, estando esses dados em consonância com a detecção da linhagem P.2 em genomas obtidos em Pernambuco.

Figura 4: Correlação entre a frequência temporal de linhagens e principais mutações de SARS-CoV-2 a cada mês de amostragem. As cores das quatro maiores linhagens são representadas na figura (eixo direito). O tracejado mostra a frequência relativa das principais mutações encontradas neste estudo (eixo esquerdo).

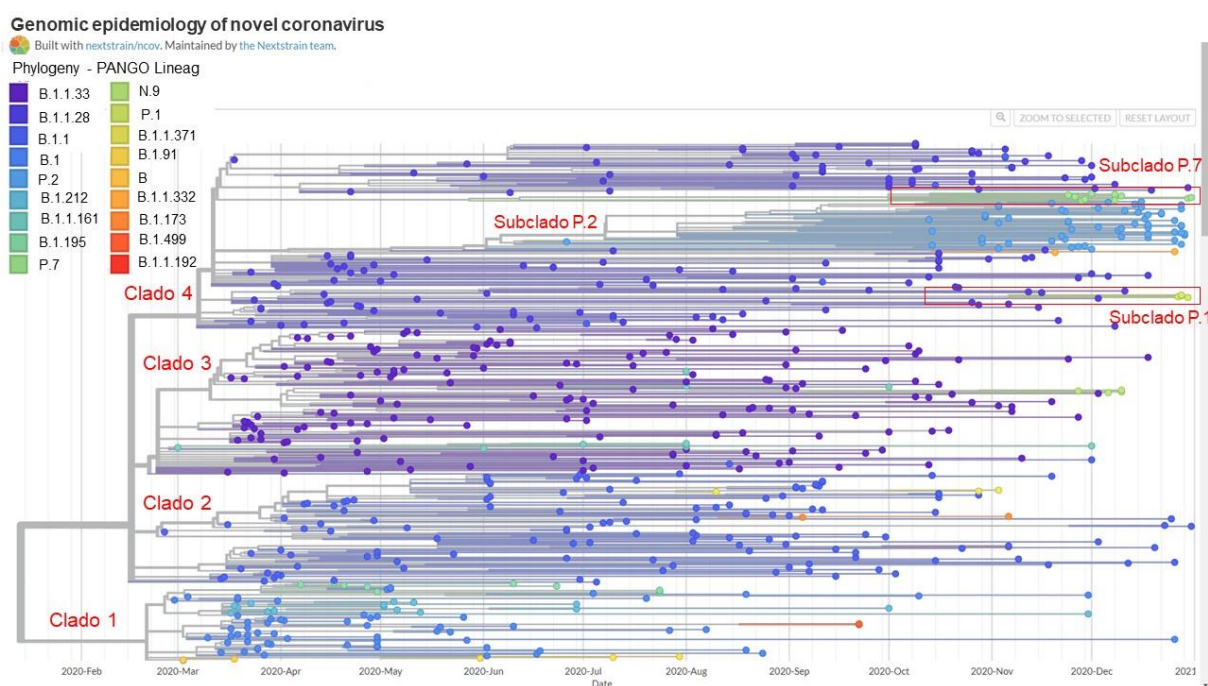


Fonte: Silveira (2021).

Análise filogenômica de SARS-CoV-2 em Pernambuco

A partir da análise filogenômica de máxima verossimilhança (ML) resolvida no tempo (arquivo Auspice disponível em <https://figshare.com/s/3c059cfe3c7f7cb07f61>), observa-se o enraizamento a partir da sequência mais ancestral Wuhan-Hu-1 e a formação de quatro clados principais, de acordo com as linhagens B.1; B.1.1; B.1.1.33; B.1.1.28, também havendo a presença de outros subclados formados por linhagens de menor representação ou em expansão inicial durante o período de amostragem (Figura 5).

Figura 5: Árvore filogenética de máxima verossimilhança resolvida no tempo dos genomas SARS-CoV-2 do estado de Pernambuco, coletados entre fevereiro e dezembro de 2020. Cada genoma é denotado com linhagem Pangolin, rotulados com cores. A árvore foi enraizada com genoma de referência (NC_045512.2)



Fonte: Silveira (2021)

O clado 1, composto principalmente por B.1, inclui amostras brasileiras coletadas entre 29 de fevereiro e abril de 2020, também incluindo cepas de países sul-americanos como Argentina, Chile, Colômbia, Equador e Peru, obtidas principalmente nos meses de março e abril. Esses genomas carregam as substituições C241T; C3037T; C14408T (P4715L); A23403G (D614G). Neste agrupamento também identificamos sequências de B.1.195 apresentando a mutação sinônima A6466G na ORF4b e B.1.91 apresentando duas alterações de

aminoácidos na proteína *Spike* (A23403G: D614G e G24077T: D839Y). Além disso, os membros dessa clado também compartilham mutações com sequências das linhagens B.1.212 (C15324T); B.1.499 (C1059T: T265I; G25563T: Q57H; C28863T: S197L e G29540A), o que justifica esse agrupamento (Apêndice C).

As sequências do clado 2 (B.1.1) são predominantes em Pernambuco, sendo 17 delas referentes a cepas obtidas neste estudo, as quais compartilham as mutações G28881A (R203K); G28882A (R203R) e G28883C (G204R). Também observamos a coocorrência das substituições C19895T (A2143V) em ORF1b e C23604A (P681H) em *Spike* em alguns dos genomas obtidos neste trabalho, coletados em outubro de 2020 (UFPE110; UFPE111; UFPE119).

O clado 3 agrupa principalmente sequências de B.1.1.33 obtidas especialmente entre o final de março e julho, estando inclusos neste clado oito genomas obtidos neste estudo representativas de três diferentes mesorregiões do estado. Observou-se que a maioria das sequências desse grupo são caracterizadas pela presença das mutações A23403G (D614G) em *Spike*; T27299C (I33T) na ORF6 e A27853C (E33A) na ORF7b compartilhadas também por cepas da linhagem N.9 e a substituição T29148C (I292T) também comum em genomas de B.1.1.161 e N.9. A mutação missense C7851T (A2529V) em ORF1a e as sinônimas C7600T e G1264T foram comuns para o subgrupo de sequências N.9 obtidas nos estados do Maranhão, Bahia e Paraíba e São Paulo, entre os meses de novembro e dezembro de 2020.

O clado 4 é composto majoritariamente por genomas estreitamente relacionados da linhagem B.1.1.28. A maioria dos genomas desse clado apresenta as mutações C12053T (L3930F) G25088T (V1176F). Além disso, dentro desse clado, seis genomas de nossas amostras apresentavam a substituição de aminoácido M902V (A2969G) e a mutação sinônima C16308T em ORF1ab. Também observamos a presença do subclado formado por 59 genomas da linhagem P.2, incluindo sete sequências dentre as obtidas neste estudo, abrangendo o período de coleta entre 30 de outubro e 06 de novembro. Esse agrupamento é caracterizado pelas mutações de assinatura C100T; T10667G (L3468V); C12053T (L3930F) na ORF1ab; G23012A (E484K) e G25088T (V1176F) em *Spike*; G28628T (A119S) e G28975T (M234I) na proteína N e C29754T na região não codificante. Essas cepas apresentaram associação com cepas da linhagem B.1.1.28 portadoras das mutações em *Spike*: E484K e V1176F, obtidas entre julho e novembro de 2020.

Por sua vez, um outro subclado dentro de B.1.1.28, composto por genomas da linhagem P.1, é constituído por cepas coletadas em dezembro nos estados de Amazonas e Bahia. Essas cepas compartilham as mutações ORF1ab: C3828T (S1188L), A5648C (K1795Q); G17259T (E1264D); em *Spike*: C21614T (L18F), C21621A (T20N), C21638T (P26S), G21974T (D138Y), G22132T (R190S), A22812C (K417T), G23012A (E484K), A23063T (N501Y), C23525T (H655Y), C24642T (T1027I), G25088T (V1176F); ORF3a: T26149C (S253P); E: A41S; ORF8: G28167A (E92K) e na proteína N: C28512G (P80R). O subclado P.7, inclui 10 genomas que carregam as alterações de aminoácidos definidoras P2287S (C7124T); L3027F (C9344T); V2588F (G8027T); Q3777H (G11596T); T151I (C25844T); P13L (C28311T), além de compartilhar as V1176F (G25088T); L3930F (C12053T) com outras linhagens desse clado.

Rede de Haplótipos

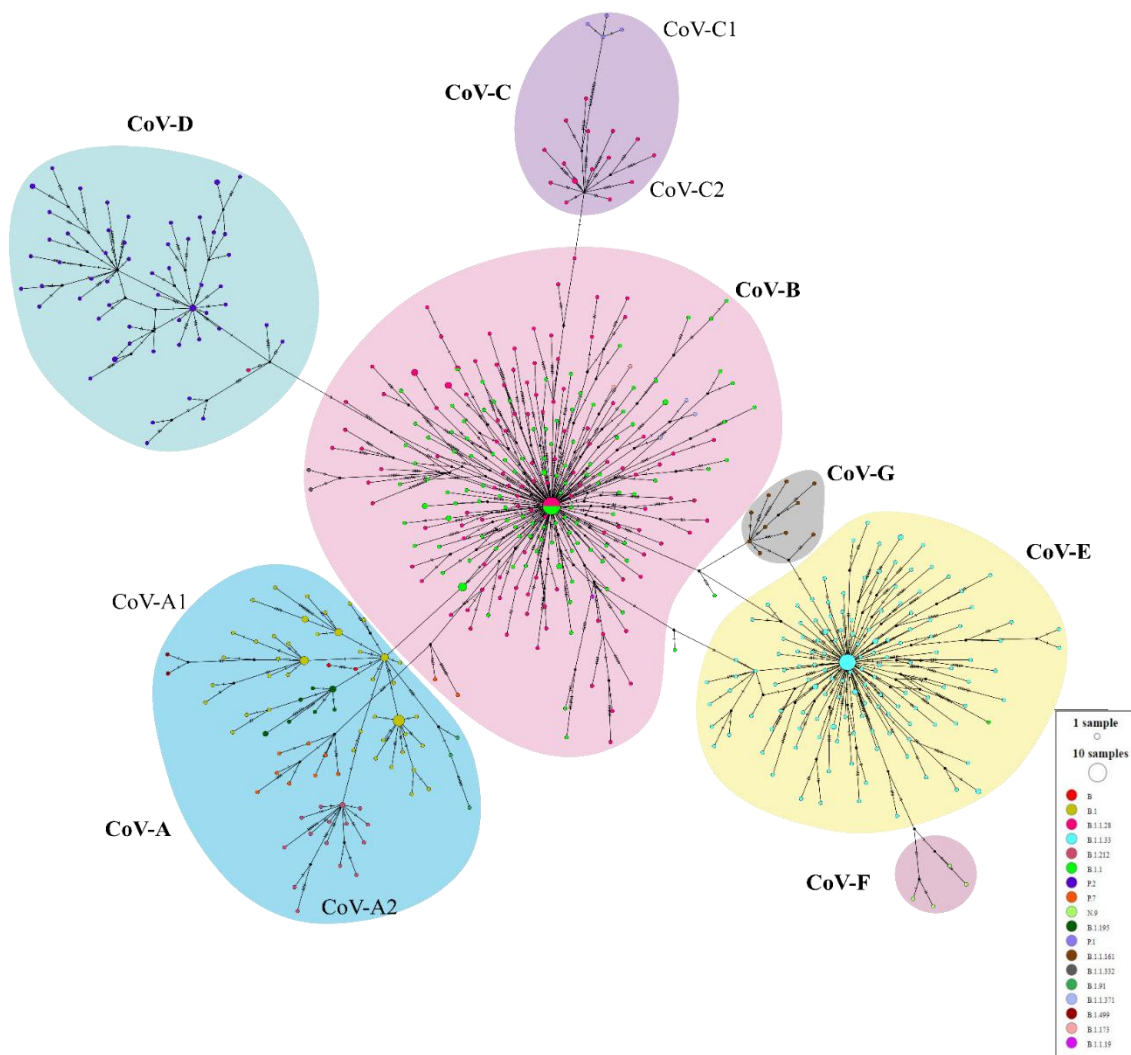
Dentre as 630 sequências genômicas de SARS-CoV-2 consideradas neste estudo, foram identificados 541 haplótipos e 31 apresentaram frequência superior a 1% (Apêndice D). Além disso, 82 haplótipos eram exclusivos de Pernambuco, sendo 49 deles representados por amostras obtidas neste estudo, marcados pelo surgimento de novas mutações aleatórias. A rede de haplótipos revelou resultados semelhantes aos da análise filogenômica com a formação de quatro grupos principais (Figura 6).

O primeiro cluster, denominado CoV-A, inclui haplótipos associados a cepas mais ancestrais principalmente das linhagens B e B.1. Nesse cluster observamos o compartilhamento das mutações G25563T (Q57H) na ORF3a entre cepas da Argentina, Chile, Colômbia e dos estados brasileiros do Paraná, Rondônia, Santa Catarina e Sergipe e cepas da linhagem B.1.499. Além disso, observamos outras linhagens agrupadas a B.1 como B.1.91, B.1.195; B.1.212, o que se deu pela presença de mutações específicas.

O cluster 2 (CoV-B) é composto B.1.1 e B.1.1.28, seu haplótipo mais frequente Hap_35 (22 sequências), inclui sequências de Pernambuco (n= 5), Alagoas, Bahia, São Paulo, Paraná e Pará e de países como Chile, Peru e Equador, caracterizados por carregarem os SNPs C241T, C3037T, C14408T, A23403G, G28881A, G28882A, G28883C. Desse agrupamento se deriva o cluster CoV-C formado por genomas coletados entre abril e início de dezembro (CoV-C1) portadores das mutações

A6319G, G25088T (V1176F) e T26149C (S253P), as quais são compartilhadas entre cepas da linhagem P.1 (CoV-C2). A partir do Cluster (CoV-D), composto por cepas da linhagem P.2 relacionados a cepas B.1.1.28 (novembro-2020) que apresentam as alterações de aminoácidos A119S e M234I na proteína N. O cluster CoV-E é formado por genomas da linhagem B.1.1.33, seu maior haplótipo Hap_25, inclui 19 sequências majoritariamente do sul e sudeste do Brasil, além de amostras do Chile e Estados Unidos, apresentando as mutações definidoras T27299C (I33T); T29148C (I292T). Desse cluster se derivam os agrupamentos CoV-F e CoV-G, formados por genomas das linhagens N.9 e B.1.1.161, evidenciando proximidade entre essas linhagens.

Figura 6: Rede de haplótipos de genomas completos de SARS-CoV-2 usando o método *median joining network*. Os haplótipos são classificados por linhagem Pangolin. O tamanho dos círculos é indicativo do número de sequências em cada haplótipo com cores diferentes indicando as diferentes linhagens do Pangolin e os números de traços na linha de conexão representam o número de mutações.



Fonte: Silveira (2021)

DISCUSSÃO

Durante a primeira onda da pandemia, o Brasil registrou ampla disseminação de SARS-CoV-2 nas suas 27 unidades federativas (CARMO *et al.*, 2020). Em Pernambuco, sétimo estado mais populoso do Brasil (IBGE, 2020), o primeiro caso confirmado de COVID-19 foi notificado em 12 de março de 2020, seguido por intensa notificação de casos e rápida expansão da COVID-19 em direção ao interior do estado (MINISTÉRIO DA SAÚDE, 2021; MAGALHÃES *et al.*, 2020). A curva epidêmica indicou uma primeira onda de casos da COVID-19 que atingiu picos em sua incidência entre 17 a 23 de maio de 2020 e entre 26 de julho a 1 de agosto. A partir desse ponto, houve diminuição nas taxas de incidência e mortalidade por COVID-19, provavelmente devido às medidas de intervenção implementadas em nível estadual, o que impediu um maior crescimento da epidemia no estado. No entanto, embora a epidemia de COVID-19 em Pernambuco durante a primeira onda tenha tido grandes proporções, a amostragem genômica do vírus não é representativa para esse período.

Para adquirir uma compreensão mais aprofundada da diversidade genética das variantes do SARS-CoV-2 que circulavam no estado do Pernambuco durante essa fase da pandemia, realizamos o sequenciamento genômico de 67 cepas de SARS-CoV-2 circulantes no estado a partir de amostras de 22 municípios de quatro mesorregiões do estado, coletadas entre o início de julho e novembro de 2020. Também recuperamos genomas disponíveis no GISAID e coletados em Pernambuco até 31 de dezembro de 2020, a fim de realizar uma caracterização dos perfis de propagação temporal e evolução do SARS-CoV-2 na primeira fase da pandemia no estado.

Nossas análises forneceram um resumo informativo da diversidade genética em nossas amostras, com diversidade nucleotídica de 0,00048 e valores dos testes de neutralidade D de Tajima e F_s de Fu de -2,80229 e -23,64854, respectivamente. Estes resultados estão em concordância com relatórios anteriores, onde um D de Tajima negativo foi estimado a partir dos genomas do SARS-CoV-2 indicando expansão populacional (LIU *et al.*, 2020; YU *et al.*, 2020; PANDIT *et al.*, 2021).

A partir da análise de chamada de variantes, foi possível observar um total de 1.002 SNPs, sendo 962 (96,00%) localizados em regiões codificantes do genoma viral, havendo predominância de mutações na ORF1ab. A identificação de um

grande número de mutações também foi observada em nível nacional durante esse período, com cerca de 5.837 mutações em regiões codificantes, a maioria delas também localizadas na ORF1ab. Esses dados obedecem a uma tendência relatada em estudos anteriores, por se tratar da maior ORF de SARS-CoV-2, que ocupa cerca de dois terços de todo o genoma (YANG *et al.*, 2020; KOYAMA *et al.*, 2020; ARMERO *et al.*, 2021). Também relatamos evidências de variações temporais na prevalência de mutações em genomas de SARS-CoV-2 pernambucanos, com pronunciado aumento no número total de SNPs a partir de julho de 2020, o que também está relacionado a um crescimento do número de amostras sequenciadas a partir desse período, evidenciando um aumento da diversidade ao longo do tempo embora o número de sequências obtidas para o estado não seja homogêneo por mês de amostragem.

Os dados de propagação temporal de mutações em genomas Pernambucanos mostraram que sete delas mantiveram suas frequências constantes superiores a 94% ao longo da primeira onda da pandemia, sendo encontradas na região 5' (C241T), em ORF1ab (C3037T: F924F; C14408T: P4715L/P323L), na proteína *Spike* (A23403G: D614G) e na proteína nucleocapsídica (G28881A: R203K, G28882A: R203R, e G28883C: G204R). Resultados semelhantes foram observados em genomas de todo o Brasil, cuja amostragem se deu no mesmo período. Análises de perfis mutacionais anteriores também evidenciaram a coexistência dessas mutações no período inicial da pandemia (PACHETTI *et al.*, 2020). Dentre elas, a mutação D614G em *Spike*, foi associada à maior liberdade conformacional na estrutura da proteína, sendo também relacionada a maior transmissibilidade e infectividade de SARS-CoV-2 (KORBER *et al.*, 2020; YURKOVETSKIY *et al.*, 2020). Já alterações no gene N: G28881A, G28882A e G28883C (R203K, R203R, G204R), foram associadas a mudanças consideráveis na ligação prevista com alguns miRNAs na região de codificação do nucleocapsídeo, podendo influenciar no progresso da infecção por SARS-CoV-2 (RAHMAN *et al.*, 2020). Em comparação às mutações que apresentaram altas frequências durante todo o curso inicial da pandemia, outras mutações emergiram e apresentaram aumentos significativos de suas frequências impulsionadas pelo processo de diversificação viral.

Foi identificada a circulação simultânea de 18 linhagens de SARS-CoV-2 em Pernambuco durante essa fase da pandemia, sendo as mais prevalentes: B.1.1, seguida por B.1.1.28, B.1.1.33 e P.2. Dentre essas, a linhagem B.1.1, apresentou

maior predominância em genomas sequenciados no início da pandemia entre os meses de abril e julho de 2020, tendo sua prevalência em nível de circulação diminuído rapidamente após esse ponto. Essa linhagem havia sido anteriormente relatada como prevalente em isolados brasileiros na fase pandêmica inicial e associada à transmissão comunitária no Brasil (NASCIMENTO *et al.*, 2020).

Para a linhagem B.1.1.28, segunda mais prevalente durante a primeira onda da pandemia no estado de Pernambuco, foi registrado um aumento significativo em sua frequência entre agosto e outubro de 2020, seguido de redução significativa de sua circulação, período no qual foi detectada a expansão inicial da linhagem P.2. Essa linhagem foi documentada em amostras sequenciadas no Rio de Janeiro em abril de 2020, segundo os dados do GISAID, no entanto sua atribuição só foi realizada após sua detecção em amostras do mesmo estado em outubro de 2020 (VOLOCH *et al.*, 2021). Em seguida, essa variante foi designada pela OMS, durante o período de março ao início de julho de 2021 como uma Variante de Interesse (VOI: Zeta; Clado GISAID: GR/484K.V2 e Nextstrain clado: 20B/ S.484K); no entanto, sua diminuição significativa em circulação levou a OMS a reclassificá-la como uma das variantes anteriormente monitorada (WHO, 2021b).

Em nossos dados, a linhagem P.2 também foi detectada a partir de outubro de 2020, incluindo genomas caracterizados por conjuntos específicos de substituições de aminoácidos: L3468V e L3930F na ORF1ab; E484K e V1176F em *Spike*; A119S e M234I na proteína N.

Nossos dados também evidenciaram a detecção da linhagem B.1.1.33 a partir de março e um rápido aumento de sua frequência durante junho em Pernambuco. Essa dinâmica de circulação também foi relatada em outros estudos em nível regional e nacional do Brasil, onde foi demonstrada a disseminação dessa linhagem a partir da Região Sudeste do Brasil para todo o país (SANTOS *et al.*, 2021; RESENDE *et al.*, 2020, FRANCESCHI *et al.*, 2021) e também países da América do Sul como Argentina e Uruguai (MIR *et al.*, 2021).

Os dados das análises filogenômica e haplotípica forneceram uma contextualização das relações evolutivas de SARS-CoV-2 de Pernambuco no contexto nacional e global. Os resultados estabeleceram uma relação positiva entre divergência genética e o tempo de amostragem das cepas, evidenciando a emergência e diversificação das linhagens ao longo do tempo. O agrupamento de

haplótipos esteve fortemente associado à ocorrência de mutações, onde cada cluster resulta de uma combinação de algumas das mutações mais frequentes e outras específicas de cada cluster, estando também associadas a linhagens específicas da história evolutiva do SARS-CoV-2.

Foi demonstrada a introdução e dispersão de quatro clados principais de SARS-CoV-2 de março a abril de 2020, formados pelas linhagens B.1; B.1.1; B.1.1.33 e B.1.1.28. O surto inicial da pandemia da COVID-19 foi impulsionado principalmente pela linhagem B.1, tendo a epidemia local em Pernambuco sido possivelmente semeada por introduções independentes de SARS-CoV-2. Esse contexto é justificado pelo compartilhamento de mutações como C15324T entre sequências de Pernambuco e de outras regiões do Brasil, bem como com genomas europeus e de países da América do Sul como Argentina, Chile (STANGE *et al.*, 2020) e da substituição G25563T (ORF3a: Q57H) a qual apresenta alta frequência dos casos de COVID-19 relacionados a linhagem B.1 dos EUA e também em Hong Kong (WANG P. *et al.*, 2021; CHU *et al.*, 2021).

A maioria das sequências de cepas de SARS-CoV-2 em Pernambuco foram agrupadas no clado da linhagem B.1.1, as quais tiveram sucesso em se espalhar por todas as mesorregiões do Estado. Embora a amostragem incompleta torne difícil localizar as fontes geográficas exatas dos eventos de introdução e dispersão em Pernambuco, a diversidade genômica das cepas em circulação no estado durante a primeira onda epidêmica corrobora pelo menos parte da diversidade global na população de SARS-CoV-2, tendo a origem de tais cepas sido provavelmente a Europa, cujas sequências dominaram a base do clado B.1.1 (MAGALHÃES *et al.*, 2020; PAIVA *et al.*, 2020; GIOVANETTI *et al.*, 2020). Também foram identificadas instâncias de transmissão comunitária, incluindo a aquisição de mutações específicas nessa linhagem encontradas apenas em genomas Pernambucanos, bem como a circulação exclusiva da linhagem a B.1.1.192 no estado.

De forma posterior à introdução e rápida disseminação de cepas internacionais em Pernambuco, também houve, entre o final de março e início de abril de 2020, a expansão de linhagens nacionais, tais como o Clado B.1.1.33 e B.1.1.28, os quais foram amplamente disseminados, atingindo as localidades mais isoladas do interior, o que destaca o sucesso epidemiológico de SARS-CoV-2. Mais tarde, a partir de B.1.1.28, houve detecção de 2 subclados emergentes (P.1 e P.2) (FARIA *et al.*, 2021a; VOLOCH *et al.*, 2021). A detecção da linhagem P.2 em genomas

pernambucanos se deu em outubro de 2020, no entanto essa linhagem já circulava em genomas brasileiros desde abril de 2020, estando associadas a cepas obtidas principalmente no Rio de Janeiro e Paraná, conforme já relatado (GIOVANETTI *et al.*, 2021).

Já o subclado da linhagem P.1 se segregou com as sequências B.1.1.28 de Manaus no estado de Amazonas. Os primeiros casos conhecidos de P.1 foram detectados no Japão e estavam relacionados a viajantes procedentes de Manaus (FUJINO *et al.*, 2021; FARIA *et al.*, 2021b). Essa linhagem impulsionou o aumento do número casos no Amazonas (Brasil) a partir de dezembro de 2020 e posteriormente se espalhou amplamente por todo o país, o que configurou o início da segunda onda da epidemia. Esta linhagem é considerada mais agressiva por exibir 17 mutações, incluindo três (K417T, E484K e N501Y) na proteína *Spike* (NAVECA *et al.*, 2021). Dentre elas, a alteração localizada no domínio de ligação ao receptor (RBD) da proteína *Spike* (G23012A: E484K) está presente não só em P.1, mas também em outras duas linhagens relatadas entre o final de 2020 e início de 2021, como a linhagem da África do Sul B.1.351 e a brasileira P.2, sendo associada ao escape de anticorpos neutralizantes contra SARS-CoV-2 e a uma atividade neutralizante *in vitro* consideravelmente menor do plasma convalescente (BAUM *et al.*, 2020; WEISBLUM *et al.*, 2020), anticorpos monoclonais (mAbs) (WANG R. *et al.*, 2021), além de estar associada a um aumento da afinidade entre RBD-ACE2 (NELSON *et al.*, 2021). Apesar desses efeitos funcionais justificarem a elevação da frequência dessa variante e o correlacionarem com o aumento do número de casos em diversos contextos epidemiológicos no Brasil, até o final de dezembro de 2020 não houve detecção da ocorrência da linhagem P.1 em Pernambuco.

CONCLUSÃO

A análise de 360 genomas de SARS-CoV-2 disponíveis para o estado de Pernambuco, revelou características dinâmicas em relação a prevalência de mutações e linhagens virais, levando ao estabelecimento e cocirculação de 18 linhagens no estado, as quais contribuíram conjuntamente para o crescimento acelerado da epidemia local. Foram evidenciados eventos de importações internacionais de linhagens, as quais se disseminaram efetivamente e possibilitaram a transmissão comunitária de B.1.1, B.1.1.28 e B.1.1.33 no estado entre março e abril de 2020, impulsionada pela incorporação de novas mutações. A partir do final

de outubro de 2020 foi detectado o surgimento da linhagem P.2 em nosso conjunto de dados. Desse modo, nosso estudo destaca a importância do monitoramento contínuo das linhagens mais frequentes de SARS-CoV-2, bem como seus processos específicos de evolução dinâmica que podem favorecer o surgimento de linhagens locais e a disseminação na comunidade. Essas informações são importantes para impulsionar respostas de saúde pública, rastrear variantes emergentes e assim limitar a disseminação viral (CÂNDIDO *et al.*, 2020; PLESSIS *et al.*, 2021).

AGRADECIMENTOS

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão de bolsa de pós-graduação vinculada ao enfrentamento da COVID-19; ao Ministério Público do Trabalho de Pernambuco (MPT-PE), ao Ministério da Educação (MEC) e à Universidade Federal de Pernambuco, pelo financiamento deste trabalho.

REFERÊNCIAS

- ANDREWS, S. A quality control tool for high throughput sequence data. 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Acesso em: 10 Ago. 2021.
- ARMERO, A.; BERTHET, N.; AVARRE, J. C. Intra-host diversity of SARS-CoV-2 should not be neglected: case of the State of Victoria, Australia. **Viruses**, v. 13, n. 1, p. 133, 2021.
- BAUM, A. *et al.* Antibody cocktail to SARS-CoV-2 *Spike* protein prevents rapid mutational escape seen with individual antibodies. **Science**, v. 369, n. 6506, p. 1014-1018, 2020.
- BONITA, R.B R.; KJELLSTRÖM, T. **Basic epidemiology, 2nd ed. World Health Organization**. 2006. Disponível em: <https://apps.who.int/iris/handle/10665/43541>. Acesso em: 28 out. 2021.
- CANDIDO, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil. **Science**, v. 369, n. 6508, p. 1255-1260, 2020.
- CARMO, R. F. *et al.* Expansion of COVID-19 within Brazil: the importance of highways. **Journal of travel medicine**, v. 27, n. 5, p. taaa106, 2020.
- CHEN, A. T. *et al.* COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. **Elife**, v. 10, p. e63409, 2021.

CHEN, Y.; LIU, Q.; GUO, D. Emerging coronaviruses: genome structure, replication, and pathogenesis. **Journal of medical virology**, v. 92, n. 4, p. 418-423, 2020.

CHU, D. K. W. *et al.* Introduction of ORF3a-Q57H SARS-CoV-2 variant causing fourth epidemic wave of COVID-19, Hong Kong, China. **Emerging infectious diseases**, v. 27, n. 5, p. 1492, 2021.

CINGOLANI, P. *et al.* 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 vol.** V.6, n.2, p.80-92, 2012.

DANECEK, P. *et al.* The variant call format and VCFtools. **Bioinformatics**, v.27, p. 2156-2158, 2011.

DANECEK, P.; MCCARTHY, S. A. BCFtools/csq: haplotype-aware variant consequences. **Bioinformatics**, v. 33, n. 13, p. 2037-2039, 2017.

EXCOFFIER, L.; LISCHER, H. EL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular ecology resources**, v. 10, n. 3, p. 564-567, 2010.

FARIA, N. R. *et al.* Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. **Science**, v. 372, n. 6544, p. 815-821, 2021.

FARIA, N. R. *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. **Virological**, 2021a. Disponível em: <https://virological.org/t/genomic-characterisation-of-an-emergent-SARS-CoV-2-lineage-in-manaus-preliminary-findings/586>. Acesso em: 29 ago. 2021.

FINKEL, Y. *et al.* The coding capacity of SARS-CoV-2. **Nature**, v. 589, p. 125–130, 2021.

FRANCESCHI, V. B. *et al.* Mutation hotspots and spatiotemporal distribution of SARS-CoV-2 lineages in Brazil, February 2020-2021. **Virus Research**, v. 304, p. 198532, 2021.

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

FURUSE, Y. Genomic sequencing effort for SARS-CoV-2 by country during the pandemic. **International Journal of Infectious Diseases**, v. 103, p. 305-307, 2021.

GARRISON, E.; MARTH, G. Haplotype-based variant detection from short-read sequencing. **arXiv preprint arXiv:1207.3907**, 2012.

GIOVANETTI, M. *et al.* A doubt of multiple introduction of SARS-CoV-2 in Italy: a preliminary overview. **Journal of medical virology**, v. 92, n. 9, p. 1634-1636, 2020.

GIOVANETTI, M. *et al.* Genomic epidemiology reveals how restriction measures shaped the SARS-CoV-2 epidemic in Brazil. **medRxiv**, 2021.

GUREVICH, A. *et al.* QUAST: quality assessment tool for genome assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072-1075, 2013.

HADFIELD, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. **Bioinformatics**, v. 34, n. 23, p. 4121-4123, 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Portal das cidades**. Disponível em: <https://cidades.ibge.gov.br/brasil/pe/pesquisa/48/48986?tipo=ranking&indicador=48985>. Acesso em: 1 mar. 2021.

JOSHI, N.; FASS, J. SICKLE- A windowed adaptive trimming tool for FASTQ files using quality. (Version 1.33), 2011.

KATOH, K.; TOH, H. Parallelization of the MAFFT multiple sequence alignment program. **Bioinformatics**, v. 26, n. 15, p. 1899-1900, 2010.

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

KOYAMA, T. *et al.* Variant analysis of SARS-CoV-2 genomes. **Bulletin of the World Health Organization**, v. 98, n. 7, p. 495, 2020.

LANGMEAD, B.; SALZBERG, S. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v.9, p. 357–359, 2012.

LEIGH, J. W.; BRYANT, D. POPART: Full-feature software for haplotype network construction. **Methods in Ecology and Evolution**, v.6, n.9, p. 1110-1116, 2015.

LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. **Bioinformatics**, v. 26, n. 5, p. 589-595, 2010.

LI, H. *et al.* The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009.

LIU, Q. *et al.* Population Genetics of SARS-CoV-2: Disentangling Effects of Sampling Bias and Infection Clusters. **Genomics, proteomics & bioinformatics**, 2020.

MAGALHÃES, J. J. F. *et al.* Epidemiological and clinical characteristics of the first 557 successive patients with COVID-19 in Pernambuco state, Northeast Brazil. **Travel medicine and infectious disease**, v. 38, p. 101884, 2020.

MARTIN, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. **EMBnet. journal**, v. 17, n. 1, p. 10-12, 2011.

MINISTÉRIO DA SAÚDE BRASIL (2021). **COVID-19 Painel Coronavírus**. Disponível em: <https://covid.saude.gov.br/> Acesso em: 21 set. 2021.

MIR, D. *et al.* Recurrent Dissemination of SARS-CoV-2 Through the Uruguayan–Brazilian Border. **Frontiers in microbiology**, v. 12, 2021.

NASCIMENTO, V. A. *et al.* Genomic and phylogenetic characterisation of an imported case of SARS-CoV-2 in Amazonas State, Brazil. **Memórias do Instituto Oswaldo Cruz**, v. 115, 2020.

NAVECA, F. G. *et al.* COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P. 1 emergence. **Nature Medicine**, p. 1-9, 2021.

NELSON, G. Molecular dynamic simulation reveals E484K mutation enhances *Spike* RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y. V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. **BioRxiv**, 2021.

NGUYEN, L. T. *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Molecular biology and evolution**, v. 32, n. 1, p. 268-274, 2015.

O'TOOLE, Á. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. **Virus Evolution**, v. 7, n. 2, p. veab064, 2021.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE (OPAS). **Brasil confirma primeiro caso de infecção pelo novo coronavírus**. 2020 Disponível em: [https://www.paho.org/bra/index.php?option=com_content&view=article&id=6113:brasil-confirma-primeiro-caso-de-infeccao-pelo-novo-coronavirus&Itemid=812-~:text=26%20de%20fevereiro%20de%202020,21%20de%20fevereiro%20deste%20ano](https://www.paho.org/bra/index.php?option=com_content&view=article&id=6113:brasil-confirma-primeiro-caso-de-infeccao-pelo-novo-coronavirus&Itemid=812-~:text=26%20de%20fevereiro%20de%202020,21%20de%20fevereiro%20deste%20ano.Acesso em: 21 set. 2020). Acesso em: 21 set. 2020.

PACHETTI, M. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. **Journal of translational medicine**, v. 18, n. 1, p. 1-9, 2020.

PAIVA, M. H. S. *et al.* Multiple introductions followed by ongoing community spread of SARS-CoV-2 at one of the largest metropolitan areas of Northeast Brazil. **Viruses**, v. 12, n. 12, p. 1414, 2020.

PANDIT, B. *et al.* Association of clade-G SARS-CoV-2 viruses and age with increased mortality rates across 57 countries and India. **Infection, Genetics and Evolution**, v. 90, p. 104734, 2021.

PLESSIS, L. *et al.* Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. **Science**, v. 371, n. 6530, p. 708-712, 2021.

RAHMAN, M. S. *et al.* Evolutionary dynamics of SARS-CoV-2 nucleocapsid protein and its consequences. **Journal of medical virology**, v. 93, n. 4, p. 2177-2195, 2021.

RAMBAUT, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. **Nature microbiology**, v. 5, n. 11, p. 1403-1407, 2020.

RESENDE, P. C. *et al.* Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B. 1.1. 33 during the early pandemic phase in Brazil. **Frontiers in microbiology**, v. 11, 2020.

ROZAS, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. **Molecular biology and evolution**, v. 34, n. 12, p. 3299-3302, 2017.

SAGULENKO, P.; PULLER, V.; NEHER, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. **Virus evolution**, v. 4, n. 1, p. vex042, 2018.

SANTOS, C. A. *et al.* SARS-CoV-2 Genomic surveillance in Northeast Brazil: timing of emergence of the Brazilian variant of concern P1. **Journal of Travel Medicine**, 2021.

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

STANGE, M. *et al.* SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B. 1 lineage variant linked to a mass gathering event. **PLoS pathogens**, v. 17, n. 3, p. e1009374, 2021.

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION. **Coronavirus: the economic impact. Retrieved.** 2020a. Disponível em: <https://www.unido.org/stories/coronavirus-economic-impact-10-july-2020>. Acesso em: 30 ago. 2021.

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION. **The Social Impact of COVID-19.** 2020b. Disponível em: <https://www.un.org/development/desa/dspd/everyone-included-covid-19.html>. Acesso em: 30 ago. 2021.

VAIDYA, G.; LOHMAN, D. J.; MEIER, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. **Cladistics**, v. 27, n. 2, p. 171-180, 2011.

VOLOCH, C. M. *et al.* Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. **Journal of virology**, v. 95, n. 10, p. e00119-21, 2021.

WANG, P. *et al.* Antibody resistance of SARS-CoV-2 variants B. 1.351 and B. 1.1. 7. **Nature**, v. 593, n. 7857, p. 130-135, 2021.

WANG, R. *et al.* Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. **Communications biology**, v. 4, n. 1, p. 1-14, 2021.

WEISBLUM, Y. *et al.* Escape from neutralizing antibodies by SARS-CoV-2 *Spike* protein variants. **Elife**, v. 9, p. e61312, 2020.

WORLD HEALTH ORGANIZATION (WHO). **Coronavirus Disease (COVID-19): Dashboard.** 2021a. Disponível em: <https://covid19.who.int>. Acesso em: 19 set. 2021.

WORLD HEALTH ORGANIZATION (WHO). **Tracking SARS-CoV-2 Variants** 2021b). Disponível em: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Acesso em: 30 nov. de 2021.

WU, F. *et al.* A new coronavirus associated with human respiratory disease in China. **Nature**, v. 579, n. 7798, p. 265-269, 2020.

YANG, H. C. *et al.* Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. **Proceedings of the National Academy of Sciences**, v. 117, n. 48, p. 30679-30686, 2020.

YU, W. B. *et al.* Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2/HCoV-19) using whole genomic data. **Zoological research**, v. 41, n. 3, p. 247, 2020.

YURKOVETSKIY, L. *et al.* Structural and functional analysis of the D614G SARS-CoV-2 *Spike* protein variant. **Cell**, v. 183, n. 3, p. 739-751. e8, 2020.

ZHANG, Y. Z.; HOLMES, E. C. A genomic perspective on the origin and emergence of SARS-CoV-2. **Cell**, v. 181, n. 2, p. 223-227, 2020.

ZHU, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. **New England journal of medicine**, 2020.

4 CONCLUSÕES

A análise de 360 genomas de SARS-CoV-2 disponíveis para o estado de Pernambuco, revelou características dinâmicas em relação a prevalência de mutações e linhagens virais, levando ao estabelecimento e cocirculação de 18 linhagens no estado, as quais contribuíram conjuntamente para o crescimento acelerado da epidemia local. Foram evidenciados eventos de importações internacionais de linhagens, as quais se disseminaram efetivamente e possibilitaram a transmissão comunitária de B.1.1, B.1.1.28 e B.1.1.33 no estado entre março e abril de 2020, impulsionada pela incorporação de novas mutações. A partir do final de outubro de 2020 foi detectado o surgimento da linhagem P.2 em nosso conjunto de dados. Desse modo, nosso estudo destaca a importância do monitoramento contínuo das linhagens mais frequentes de SARS-CoV-2, bem como seus processos específicos de evolução dinâmica que podem favorecer o surgimento de linhagens locais e a disseminação na comunidade. Essas informações são importantes para impulsionar respostas de saúde pública, rastrear variantes emergentes e assim limitar a disseminação viral (CÂNDIDO *et al.*, 2020; PLESSIS *et al.*, 2021).

5 SÚMULA CURRICULAR

- Capítulo de livro publicado: Estudo das Propriedades Biológicas de *Jatropha mollissima* (POHL BAILL). Link: <https://www.atenaeditora.com.br/post-ebook/4487>
- Artigo publicado: Phytochemistry and Biological Activities of *Amburana cearensis* (Allemão) ACSm. Link: <https://doi.org/10.3390/molecules27020505>
- Evaluation of the antibacterial activity and efflux pump reversal of thymol and carvacrol against *Staphylococcus aureus* and their toxicity in *Drosophila melanogaster*. Link: <https://doi.org/10.3390/molecules25092103>
- Artigo publicado: Toxicity of methyl eugenol against *Drosophila melanogaster* and its myorelaxant activity in bronchioles isolated from *Sus scrofa domesticus*. Link: <https://doi.org/10.2478/s11756-020-00672-y>
- Artigo publicado: *Caesalpinia ferrea* C. Mart. (Fabaceae) phytochemistry, ethnobotany, and bioactivities: a review. Link: <https://doi.org/10.3390/molecules25173831>
- Artigo publicado: Evaluation of phytochemical composition, toxicity in *Drosophila melanogaster* and effects on antibiotics modulation of *Plathymenia reticulata* Benth extract. Link: <https://doi.org/10.1016/j.toxrep.2021.03.020>
- Artigo publicado: Evaluation of isoeugenol in inhibition of *Staphylococcus aureus* efflux pumps and their toxicity using *Drosophila melanogaster* model.. Link: <https://doi.org/10.1016/j.lfs.2021.119940>
- Artigo publicado: Enhancement of the antibiotic activity by quercetin against *Staphylococcus aureus* efflux pumps. Link: <https://doi.org/10.1007/s10863-021-09886-4>
- Artigo publicado: Effect of Carvacrol and Thymol on NorA efflux pump inhibition in multidrug-resistant (MDR) *Staphylococcus aureus* strains. Link: <https://doi.org/10.1007/s10863-021-09906-3>

REFERÊNCIAS

- AKSAMENOV, I. *et al.* Nextclade: clade assignment, mutation calling and quality control for viral genomes. **Journal of Open Source Software**, v.6, n. 67, p.3773, 2021.
- ALEXANDERSEN, S., CHAMINGS, A. & BHATTA, T.R. SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. **Nature communications**, v. 11, p. 6059, 2020.
- ALLEN, T. *et al.* Global hotspots and correlates of emerging zoonotic diseases. **Nature communications**, v. 8, n.1, 1-10, 2017.
- ANDERSEN, K. G. *et al.* The proximal origin of SARS-CoV-2. **Nature medicine**, v. 26, n. 4, p. 450-452, 2020.
- ARYA, R. *et al.* Structural insights into SARS-CoV-2 proteins. **Journal of molecular biology**, v. 433, n.2, p. 166725, 2021.
- ASTUTI, I. *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response.* **Diabetes & Metabolic Syndrome: Clinical Research & Reviews**, v.14, n.4, p. 407-412, 2020.
- BAKER, D. J. *et al.* CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes. **Genome medicine**, v.13, n.1, 1-11, 2021.
- BAUER, D. C. *et al.* Supporting pandemic response using genomics and bioinformatics: A case study on the emergent SARS-CoV-2 outbreak. **Transboundary and emerging diseases**, v. 67, n. 4, p. 1453-1462, 2020.
- BHOYAR, R. C. *et al.* High throughput detection and genetic epidemiology of SARS-CoV-2 using COVIDSeq next-generation sequencing. **PloS one**, v. 16, n. 2, p. e0247115, 2021.
- BIANCHI, M. *et al.* SARS-CoV-2 envelope and membrane proteins: structural differences linked to virus characteristics? **BioMed research international**, 2020.
- BOTELHO-SOUZA, L. F. *et al.* SARS-CoV-2 genomic surveillance in Rondônia, Brazilian Western Amazon. **Scientific reports**, v. 11, n.1, p. 1-12, 2021.
- CALLAWAY, E. Heavily mutated Omicron variant puts scientists on alert. **Nature**, v.600, n. 7887, p.21-21, 2021.
- CALLAWAY, E.; LEDFORD, H.; MALLAPATY, S. Six months of coronavirus: the mysteries scientists are still racing to solve. **Nature**, v. 583, n. 7815, p. 178-179, 2020.
- CANDIDO, D. S. *et al.* Evolution and epidemic spread of SARS-CoV-2 in Brazil.

Science, v. 369, n. 6508, p. 1255-1260, 2020.

CANNATARO M.; HARRISON, A. Bioinformatics helping to mitigate the impact of COVID-19 – Editorial. **Briefings in Bioinformatics**, v.22, n.2, p. 613–615, 2021.

CARMO, R. F. *et al.* Expansion of COVID-19 within Brazil: the importance of highways. **Journal of travel medicine**, v. 27, n. 5, p. taaa106, 2020

CHAN, J. F. W. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. **Emerging microbes & infections**, v.9, n.1, p. 221-236, 2020.

CHEN, A. T. *et al.* COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. *Elife*, v. 10, p. e63409, 2021.

CHEN, C. *et al.* CoV-Spectrum: Analysis of globally shared SARS-CoV-2 data to Identify and Characterize New Variants. **Bioinformatics**, 2021.

CHEN, Z. *et al.* Genomic and evolutionary comparison between SARS-CoV-2 and other human coronaviruses. **Journal of virological methods**, v. 289, p.114032, 2021.

CUI, J.; LI, F.; SHI, Z. L. Origin and evolution of pathogenic coronaviruses. **Nature Reviews Microbiology**, n. 17, n.3, p.181-192, 2019.

DAVIDSON, A.D. *et al.* Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the *Spike* glycoprotein. **Genome medicine**, v. 12, n. 1, p. 1-15, 2020.

DELAUNE, D. *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. **Nature Communications**, v. 12, n. 1, p. 1-7, 2021.

DELLICOUR, S. *et al.* A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. **Molecular biology and evolution**, v. 38, n. 4, p. 1608-1613, 2021.

DHAKAL, S. *et al.* Could the severity of COVID-19 be enhanced by ORF10 accessory proteins?. **Current Topics in Peptide & Protein Research**, v. 21, p. 97, 2020.

DROSTEN, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. **New England journal of medicine**, v. 348, n. 20, p. 1967-1976, 2003.

FARIA, N. R. *et al.* Genomics and epidemiology of the P. 1 SARS-CoV-2 lineage in Manaus, Brazil. **Science**, v. 372, n. 6544, p. 815-821, 2021.

FINKEL, Y. *et al.* The coding capacity of SARS-CoV-2. **Nature**, v. 589, p. 125–130, 2021.

FLOWER, T. G. *et al.* Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. **Proceedings of the National Academy of Sciences**, v. 118, n. 2, p. e2021785118, 2021.

FORNI, D. *et al.* Molecular evolution of human coronavirus genomes. **Trends in microbiology**, v. 25, n.1, p. 35-48, 2017.

FRANCESCHI, V. B. *et al.* Genomic epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil. **BMC genomics**, v. 22, n. 1, p. 1-17, 2021.

FURUSE, Y. Genomic sequencing effort for SARS-CoV-2 by country during the pandemic. **International Journal of Infectious Diseases**, v. 103, p. 305-307, 2021.

GORBALENYA, A, E. *et al.* Severe acute respiratory syndrome-related coronavirus: The species and its viruses—a statement of the Coronavirus Study Group. 2020.

HASSAN, S. S.; CHOUDHURY, P. P.; ROY, B. Rare mutations in the accessory proteins orf6, orf7b, and orf10 of the SARS-CoV-2 genomes. **Meta Gene**, v. 28, p. 100873, 2021.

HOQUE, M. N. *et al.* Genomic diversity and evolution, diagnosis, prevention, and therapeutics of the pandemic COVID-19 disease. **PeerJ**, v.8, p. e9689, 2020.

HU, T. *et al.* Bioinformatics resources for SARS-CoV-2 discovery and surveillance. **Briefings in Bioinformatics**, v. 22, n. 2, p. 631-641, 2021.

HUFISKY, F. *et al.* Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. **Briefings in bioinformatics**, v. 22, n. 2, p. 642-663, 2021.

INTERNATIONAL COMMITTEE ON TAXONOMY OF VIRUSES (ICTV). **Virus Taxonomy: 2020 Release**. Disponível em: <https://talk.ictvonline.org/taxonomy>. Acesso em: 13 ago. 2021.

JOHN, G. *et al.* Next-generation sequencing (NGS) in COVID-19: A tool for SARS-CoV-2 diagnosis, monitoring new strains and phylodynamic modeling in molecular epidemiology. **Current issues in molecular biology**, v. 43, n. 2, p. 845-867, 2021.

KHAN, M. T. *et al.* Structures of SARS-CoV-2 RNA-Binding Proteins and Therapeutic Targets. **Intervirolgy**, p.1-14, 2020.

KUO, L. *et al.* Analyses of coronavirus assembly interactions with interspecies membrane and nucleocapsid protein chimeras. **Journal of virology**, v.90, n.9, p. 4357-4368, 2016.

LAI, C. C. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Coronavirus Disease-2019 (COVID-19): **The Epidemic and the Challenges**. *International Journal of Antimicrobial Agents*, v. 55, n.3, p. 105924, 2020.

LAM, T.T.Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, v. 583, n. 7815, p. 282-285, 2020.

LAURING, A. S.; HODCROFT, E. B. Genetic variants of SARS-CoV-2—what do they mean? *Jama*, v. 325, n. 6, p. 529-531, 2021.

LEAO, J. C. *et al.* Coronaviridae—Old friends, new enemy! *Oral Diseases*, 2020.

LLANES, A. *et al.* Betacoronavirus genomes: how genomic information has been used to deal with past outbreaks and the COVID-19 pandemic. *International journal of molecular sciences*, v. 21, n. 12, p. 4546, 2020.

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

NAVECA, F. G. *et al.* COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P. 1 emergence. *Nature Medicine*, p. 1-9, 2021.

MERCATELLI, D. *et al.* Coronapp: a web application to annotate and monitor SARS-CoV-2 mutations. *Journal of medical virology*, v. 93, n. 5, p. 3238-3245, 2021.

MINISTÉRIO DA SAÚDE BRASIL. **COVID-19 Painei Coronavírus**. 2021. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 31 out. 2021.

MOLLENTZE, N.; STREICKER, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences*, v.117, n.17, p.9423-9430, 2020.

MOUSAVIZADEH, L.; GHASEMI, S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of Microbiology, Immunology and Infection*, v. 54, n. 2, p. 159-163, 2021.

MUÑOZ, M. *et al.* Characterizing SARS-CoV-2 genome diversity circulating in South American countries: Signatures of potentially emergent lineages?. *International Journal of Infectious Diseases*, v. 105, p. 329-332, 2021.

MURAKAMI, S. *et al.* Detection and characterization of bat *Sarbecovirus* phylogenetically related to SARS-CoV-2, Japan. *Emerging infectious diseases*, v. 26, n. 12, p. 3025, 2020.

NAQVI, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, v. 1866, n. 10, p.165878, 2020.

NEUMAN, B. W. *et al.* A structural analysis of M protein in coronavirus assembly and morphology. *Journal of structural biology*, v.174, n.1, p. 11-22, 2011.

OLIVAL, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature*, v. 546, n.7660, 646-650, 2017.

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE (OPAS). **Brasil confirma primeiro caso de infecção pelo novo coronavírus**. 2020. Disponível em: https://www.paho.org/bra/index.php?option=com_content&view=article&id=6113:brasil-confirma-primeiro-caso-de-infeccao-pelo-novo-coronavirus&Itemid=812 :-:text=26%20de%20fevereiro%20de%202020,21%20de%20fevereiro%20dest e%20ano . Acesso em: 21 set. 2020.

PADEN, C. R. *et al.* Rapid, sensitive, full-genome sequencing of severe acute respiratory syndrome coronavirus 2. *Emerging infectious diseases*, v. 26, n. 10, p. 2401, 2020.

PAIVA, M. H. S. *et al.* Multiple introductions followed by ongoing community spread of SARS-CoV-2 at one of the largest metropolitan areas of Northeast Brazil. *Viruses*, v. 12, n. 12, p. 1414, 2020.

PEKAR, J. *et al.* Timing the SARS-CoV-2 index case in Hubei province. *Science*, v. 372, n. 6540, p. 412-417, 2021.

PENG, X. *et al.* Transmission Routes of 2019-NCoV and Controls in Dental Practice. *International Journal of Oral, Science*, v.12, n.1, p. 1–6, 2020.

PÉREZ-LOSADA, M. *et al.* Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution*, v. 30, p. 296-307, 2015.

RAHMAN, M. S. *et al.* Mutational insights into the envelope protein of SARS-CoV-2. *Gene reports*, v.22, p.100997, 2021.

RAMBAUT, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of *Spike* mutations. *Virological*; 2020. Disponível em: < <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-SARS-CoV-2-lineage-in-the-uk-defined-by-a-novel-set-of-Spike-mutations/563>>. Acesso em: 15 set. 2021.

RAY, M. *et al.* Essential interpretations of bioinformatics in covid-19 pandemic. *Meta Gene*, p. 100844, 2020.

RELMAN, D.A. Opinion: To stop the next pandemic, we need to unravel the origins of COVID-19. **Proceedings of the National Academy of Sciences**, v. 117, n. 47, p. 29246-29248, 2020.

RITCHIE, H. *et al.* Coronavirus pandemic (COVID-19). **Our World in Data**, 2020. Disponível em: <https://ourworldindata.org/covid-cases>. Acesso em: 31 out. 2021.

ROGERS, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. **Science**, v.369, n.6506, p. 956-963, 2020.

ROY, S. *et al.* Standards and guidelines for validating next-generation sequencing bioinformatics pipelines: a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. **The Journal of Molecular Diagnostics**, v. 20, n. 1, p. 4-27, 2018.

SHEREEN M. A *et al.* COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. **Journal of advanced research**, v. 24, p. 91-98, 2020.

SHU, Y. L.; MCCAULEY, J. GISAID: global initiative on sharing all influenza data - from vision to reality. **Eurosurveillance**, v. 22, p. 2–4, 2017.

SINGH, J.; RAHMAN, S. A. EHTESHAM, Nasreen Z.; HIRA, Subhash; HASNAIN, Seyed E. SARS-CoV-2 variants of concern are emerging in India. **Nature Medicine**, v. 27, n. 7, p. 1131–1133, 2021.

SU, S. *et al.* Epidemiology, genetic recombination, and pathogenesis of coronaviruses. **Trends in microbiology**, v.24, n.6, p. 490-502, 2016.

TEGALLY, H. *et al.* Sixteen novel lineages of SARS-CoV-2 in South Africa. **Nature Medicine**, v. 27, n. 3, p. 440-446, 2021.

THOMAS, S. The structure of the membrane protein of SARS-CoV-2 resembles the sugar transporter semiSWEET. **Pathogens and Immunity**, v.5, n.1, p.342, 2020.

VAN DORP, L. *et al.* Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. **Infection, Genetics and Evolution**, v. 83, p. 104351, 2020.

VANDERBILT, C.; MIDDHA, S. Role of Bioinformatics in Molecular Medicine. In: **Genomic Medicine**. Springer, Cham, p. 55-68, 2020.

VANWAMBEKE, S. O.; LINARD, C.; GILBERT, M. Emerging challenges of infectious diseases as a feature of land systems. **Current Opinion in Environmental Sustainability**, v. 38, p. 31-36, 2019.

WACHARAPLUESADEE, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. **Nature**

Communications, v. 12, n. 1, p. 1-9, 2021.

WALLS, A. C. *et al.* Structure, function, and antigenicity of the SARS-CoV-2 Spike glycoprotein. **Cell**, v.181, n.2, p.281-292, 2020.

WAMAN, V. P. *et al.* The impact of structural bioinformatics tools and resources on SARS-CoV-2 research and therapeutic strategies. **Briefings in Bioinformatics**, v. 22, n. 2, p. 742-768, 2021.

WANG, L. F. *et al.* Review of bats and SARS. **Emerging infectious diseases**, v.12, n. 12, p. 1834, 2006.

WANG, P. *et al.* Antibody resistance of SARS-CoV-2 variants B. 1.351 and B. 1.1. 7. **Nature**, v. 593, n. 7857, p. 130-135, 2021.

WASSENAAR, T. M.; ZOU, Y. 2019_nCoV/SARS-CoV-2: rapid classification of betacoronaviruses and identification of Traditional Chinese Medicine as potential origin of zoonotic coronaviruses. **Letters in Applied Microbiology**, v.70, n.5, 342-348. 2020.

WOO, P. C. *et al.* Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. **Journal of virology**, v. 86, n. 7, p.3995-4008. 2012.

WORLD HEALTH ORGANIZATION (WHO). **Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)**. Disponível em: [www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](http://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)). Acesso em: 01 nov. 2021.

WORLD HEALTH ORGANIZATION (WHO). **Coronavirus Disease (COVID-19): Dashboard**. 2021a. Disponível: <https://covid19.who.int>. Acesso em: 01 nov. 2021.

WORLD HEALTH ORGANIZATION (WHO). **WHO-convened global study of origins of SARS-CoV-2: China Part**. 2021b. Disponível em: <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>. Acesso em: 01 nov. 2021.

WORLD HEALTH ORGANIZATION (WHO). **Tracking SARS-CoV-2 Variants**. 2021c. Disponível em: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>. Acesso em: 30 dez. 2021.

WOROBAY, M. *et al.* The emergence of SARS-CoV-2 in europe and north america. **Science**, v. 370, n. 6516, p. 564-570, 2020.

WU, F. *et al.* A new coronavirus associated with human respiratory disease in China. **Nature**, v. 579, n. 7798, p. 265-269, 2020.

XIAO, M. *et al.* Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. **Genome medicine**, v. 12, n. 1, p. 1-15, 2020.

XU, K. *et al.* Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein. **Virology**, v. 388, n. 2, p. 279-285, 2009.

ZAKI, A. M. *et al.* Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi arabia. **New England Journal of Medicine**, v. 367, n. 19, p. 1814–20, 2012.

ZHOU, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**, v. 579, n. 7798, p. 270-273, 2020.

ZHOU, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. **Cell discovery**, v. 6, n. 1, p. 1-18, 2020.

ZHU, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. **New England Journal of Medicine**, 382, p.727-733, 2020.

ZIEGLER, K. *et al.* SARS-CoV-2 samples may escape detection because of a single point mutation in the N gene. **Eurosurveillance**, v. 25, n. 39, p. 2001650, 2020.

			20						
Location	Petrolina	Tuparetama	Petrolândia	Parnamirim	Bodocó	Itaíba	Itaíba	Custódia	Carnaíba
# contigs	1	1	1	1	1	1	1	1	1
Largest contig	29859	29861	29859	29903	29900	29903	29903	29903	29903
Total length	29859	29861	29859	29903	29900	29903	29903	29903	29903
Reference length	29903	29903	29903	29903	29903	29903	29903	29903	29903
GC (%)	37.98	37.98	38.00	37.95	37.97	37.95	37.96	37.93	37.97
Reference GC (%)	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97
N50	29859	29861	29859	29903	29900	29903	29903	29903	29903
Largest alignment	29834	29836	29834	29878	29875	29891	29878	29878	29878
Total aligned length	29834	29836	29834	29878	29875	29891	29878	29878	29878
N per sample	25	25	65	34	34	18	25	86	34
Genome fraction (%)	99.769	99.776	99.769	99.916	99.906	99.960	99.916	99.916	99.916
Coverage (%)	99.769	99.776	99.769	99.916	99.906	99.960	99.916	99.916	99.916
Assembly	UFPE090	UFPE091	UFPE092	UFPE093	UFPE094	UFPE095	UFPE096	UFPE097	UFPE098
Collection date	05/09/2020	05/09/2020	11/09/2020	11/09/2020	11/09/2020	11/09/2020	11/09/2020	11/09/2020	16/09/2020
Location	Petrolândia	Petrolândia	Brejinho	Brejinho	Brejinho	Brejinho	Brejinho	Brejinho	Exu
# contigs	1	1	1	1	1	1	1	1	1
Largest contig	29903	29903	29897	29903	29902	29859	29859	29859	29864

Total length	29903	29903	29897	29903	29902	29859	29859	29859	29864
Reference length	29903	29903	29903	29903	29903	29903	29903	29903	29903
GC (%)	37.95	37.93	37.92	37.95	37.95	38.00	37.99	37.98	37.99
Reference GC (%)	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97
N50	29903	29903	29897	29903	29902	29859	29859	29859	29864
Largest alignment	29878	29834	29872	29878	29877	29834	29834	29834	29841
Total aligned length	29878	29834	29872	29878	29877	29834	29834	29834	29841
N per sample	36	67	34	35	36	25	25	25	23
Genome fraction (%)	99.916	99.769	99.896	99.916	99.913	99.769	99.769	99.769	99.793
Coverage (%)	99.916	99.769	99.896	99.916	99.913	99.769	99.769	99.769	99.793
Assembly	UFPE099	UFPE100	UFPE101	UFPE102	UFPE103	UFPE104	UFPE105	UFPE106	UFPE107
Collection date	16/09/2020	09/10/2020	09/10/2020	09/10/2020	09/10/2020	09/10/2020	09/10/2020	10/10/2020	13/10/2020
Location	Petrolândia	Caetés	Caetés	Exu	Exu	Exu	Exu	Floresta	Sertânia
# contigs	1	1	1	1	1	1	1	1	1
Largest contig	29900	29859	29859	29875	29859	29854	29859	29903	29892
Total length	29900	29859	29859	29875	29859	29854	29859	29903	29892
Reference length	29903	29903	29903	29903	29903	29903	29903	29903	29903
GC (%)	37.95	37.98	37.98	37.97	37.98	37.99	37.97	37.94	37.95
Reference GC (%)	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97
N50	29900	29859	29859	29875	29859	29854	29859	29903	29892

Genome fraction (%)	99.776	99.786	99.769	99.769	99.769	99.769	99.786	99.716	99.716
Coverage (%)	99.776	99.786	99.769	99.769	99.769	99.769	99.786	99.716	99.716
Assembly	UFPE117	UFPE118	UFPE119	UFPE120	UFPE121	UFPE122	UFPE123	UFPE124	UFPE125
Collection date	26/10/2020	Caetés	Caetés	Caetés	Bonito	Bonito	Bonito	03/11/2020	03/11/2020
Location	Parnamirim	28/10/2020	28/10/2020	28/10/2020	30/10/2020	30/10/2020	30/10/2020	Feira Nova	Feira Nova
# contigs	1	1	1	1	1	1	1	1	1
Largest contig	29903	29903	29859	29858	29843	29890	29859	29843	29804
Total length	29903	29903	29859	29858	29843	29890	29859	29843	29804
Reference length	29903	29903	29903	29903	29903	29903	29903	29903	29903
GC (%)	37.94	37.96	37.98	37.99	37.98	37.94	37.97	37.98	37.99
Reference GC (%)	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97
N50	29903	29903	29859	29858	29843	29890	29859	29843	29804
Largest alignment	29818	29878	29834	29833	29818	29865	29834	29819	29779
Total aligned length	29818	29878	29834	29833	29818	29865	29834	29819	29779
N per sample	50	26	25	46	25	35	25	24	25
Genome fraction (%)	99.716	99.916	99.769	99.766	99.716	99.873	99.769	99.719	99.719
Coverage (%)	99.716	99.916	99.769	99.766	99.716	99.873	99.769	99.719	99.719
Assembly	UFPE126	UFPE127	UFPE128	UFPE129	UFPE130	UFPE131	UFPE132	UFPE133	UFPE134
Collection date	Feira Nova	05/11/2020	05/11/20	05/11/2020	06/11/2020	06/11/2020	06/11/2020	06/11/2020	06/11/2020

			20						
Location	03/11/2020	Floresta	Floresta	Floresta	Caetés	Caetés	Arcoverde	Arcoverde	Orobó
# contigs	1	1	1	1	1	1	1	1	1
Largest contig	29903	29859	29859	29903	29903	29903	29860	29863	29843
Total length	29903	29859	29859	29903	29903	29903	29860	29863	29843
Reference length	29903	29903	29903	29903	29903	29903	29903	29903	29903
GC (%)	37.96	37.97	37.97	37.93	37.94	37.94	37.98	37.99	37.99
Reference GC (%)	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97	37.97
N50	29903	29859	29859	29903	29903	29903	29860	29863	29843
Largest alignment	29878	29834	29834	29878	29818	29878	29835	29838	29818
Total aligned length	29878	29834	29834	29878	29818	29878	29835	29838	29818
N per sample	32	25	25	25	50	35	25	25	35
Genome fraction (%)	99.916	99.769	99.769	99.916	99.716	99.916	99.773	99.783	99.716
Coverage (%)	99.916	99.769	99.769	99.916	99.716	99.916	99.773	99.783	99.716
Assembly	UFPE135	UFPE136	UFPE137	UFPE138					
Collection date	06/11/2020	07/11/2020	07/11/2020	07/11/2020					
			20						
Location	Orobó	Caetés	Caetés	Caetés					
# contigs	1	1	1	1					
Largest contig	29903	29864	29859	29862					
Total length	29903	29864	29859	29862					

Reference length	29903	29903	29903	29903					
GC (%)	37.94	38.03	37.97	37.98					
Reference GC (%)	37.97	37.97	37.97	37.97					
N50	29903	29864	29859	29862					
Largest alignment	29878	29839	29834	29837					
Total aligned length	29878	29839	29834	29837					
N per sample	33	241	62	25					
Genome fraction (%)	99.916	99.786	99.769	99.779					
Coverage (%)	99.916	99.786	99.769	99.779					

APÊNDICE B – MUTAÇÕES COM FREQUÊNCIAS MAIORES QUE 5% ENTRE OS 67 GENOMAS DE PERNAMBUCO OBTIDOS NESTE ESTUDO E OS 293 RECUPERADAS DO GISAID.

Mutação	Gene/ORF	Tipo de mutação	Alteração aa	Freq. (abr-jun)	Freq. (jul-set)	Freq. (out-dez)
C100T	ORF1ab	5' UTR	-	0,00%	0,00%	19,82%
C241T	ORF1ab	5' UTR	-	98,68%	98,8%	100,00%
C3037T	ORF1ab	Sinônima	F924F	98,68%	100%	100%
T10667G	ORF1ab	missense	L3468V	0,00%	0,00%	17,12%
C11824T	ORF1ab	Sinônima	I3853I	0,00%	0,00%	17,12%
C12053T	ORF1ab	missense	L3930F	7,89%	27,17%	42,34%

A12964G	ORF1ab	Sinônima	G4233G	0,00%	0,00%	17,12%
C14408T	ORF1ab	Sinônima	P4715L/P32 3L	96,05%	99,42%	99,10%
G20292A	ORF1ab	missense	G676D	9,21%	5,78%	5,41%
G23012A	S	missense	E484K	0,00%	0,00%	18,02%
A23403G	S	missense	D614G	98,68%	100%	100%
G25088T	S	missense	V1176F	10,53%	30,64%	51,35%
G25906T	ORF3a	missense	G172C	0,00%	13,29%	11,71%
T27299C	ORF6	missense	I33T	6,58%	14,45%	25,23%

C28253T	ORF8	Sinônima	F120F	0,00%	1,73%	20,72%
G28628T	N	missense	A119S	0,00%	0,00%	18,02%
G28881A	N	missense	R203K	94,74%	100,00%	99%
G28882A	N	Sinônima	R203R	94,74%	100,00%	99%
G28883C	N	missense	G204R	94,74%	100,00%	99%
G28975T	N	missense	M234I	0,00%	0,00%	19,82%
T29148C	N	missense	I292T	6,58%	14,45%	25,23%
C29754T	-	3' UTR	-	0,00%	0,00%	20,72%

APÊNDICE C – MUTAÇÕES DEFINIDORAS DE CADA CLADO

Cluster	Sítio Mutado	Região genômica	Mutação sinônima	Localização aa	Linhagem
Clado 1					
	C241T	5' UTR	—	—	B.1; B.1.195; B.1.212; B.1.91
	C1059T	ORF1ab	Não	T265I	B.1; B.1.499;
	C3037T	ORF1ab	Sim	—	B.1; B.1.195; B.1.212; B.1.91
	A6466G	ORF1ab	Sim	—	B.1.195
	C14408T	ORF1ab	Não	P4715L	B.1; B.1.195; B.1.212; B.1.91
	C15324T	ORF1ab	Sim	—	B.1; B.1.212;
	C16428T	ORF1ab	Sim	—	B.1.212

Clado 2	A22320G	S	Não	D253G	B.1.499;
	A23403G	S	Não	D614G	B.1; B.1.195; B.1.212; B.1.91
	G24077T	S	Não	D839Y	B.1.91
	G25563T	ORF3a	Não	Q57H	B.1 e B.1.499
	C28863T	N	Não	S197L	B.1.499; B.1
	G29540A	3' UTR	—	—	B.1.499; B.1
	C19895T	ORF1ab	Não	A2143V	B.1.1 LABBE
	C23604A	S	Não	P681H	B.1.1 LABBE
	G28881A	N	Não	R203K	B.1.1

Clado 3

G28882A	N	Sim	—	B.1.1
G28883C	N	Não	G204R	B.1.1
G1264T	ORF1ab	Sim	—	N.9
C7851T	ORF1a	Não	A2529V	N.9
C7600T	ORF1ab	Sim	—	N.9
T11078C	ORF1ab	Não	F3605L (F36L)	N.9
C14408T	ORF1ab	Não	P4715L	B.1.1.33; N.9; B.1.1.161
G23012A	S	Não	E484K	N.9
A23403G	S	Não	D614G	N.9; B.1.1.33

Clado 4	T27299C	ORF6	Não	I33T	B.1.1.33; N.9
	A27853C	ORF7b	Não	E33A	N.9; B.1.1.33
	G28881A	N	Não	R203K	B.1.1.33; B.1.1.161
	G28882A	N	Sim	R203R	B.1.1.33; B.1.1.161
	G28883C	N	Não	G204R	B.1.1.33; B.1.1.161
	T29148C	N	Não	I292T	B.1.1.33; N.9; B.1.1.161
	A2969G	ORF1a	Não	M902V	B.1.1.28 de PE
	A6319G	ORF1ab	Sim	—	B.1.1.28
	C12053T	ORF1ab	Não	L3930F	B.1.1.28; P.2; B.1.1.332

Sub-clado P1	C16308T	ORF1ab	Sim	—	B.1.1.28 de PE
	G25088T	S	Não	V1176F	B.1.1.28; P.1; P.2; B.1.1.332
	T26149C	ORF3a	Não	S253P	B.1.1.28; P.1
	C3828T	ORF1ab	Não	S1188L	P.1
	A5648C	ORF1ab	Não	K1795Q	P.1
	G17259T	ORF1b	Não	E1264D	P.1
	C21614T	S	Não	L18F	P.1
	C21621A	S	Não	T20N	P.1
	C21638T	S	Não	P26S	P.1

G21974T	S	Não	D138Y	P.1
G22132T	S	Não	R190S	P.1
A22812C	S	Não	K417T	P.1
G23012A	S	Não	E484K	P.1; P.2; B.1.1.28
A23063T	S	Não	N501Y	P.1
C23525T	S	Não	H655Y	P.1
C24642T	S	Não	T1027I	P.1
T26149C	ORF3a	Não	S253P	P.1; B.1.1.28
G28167A	ORF8	Não	E92K	P.1

Sub-clado P2	C28512G	N	Não	P80R	P.1
	C100T	5' UTR	—	—	P.2
	T10667G	ORF1ab / nsp5A-B	Não	L3468V	P.2
	C12053T	ORF1ab	Não	L3930F	B.1.1.28; P.2; B.1.1.332
	G23012A	S	Não	E484K	P.1; P.2; B.1.1.28
	G25088T	S	Não	V1176F	B.1.1.28; P.1; P.2; B.1.1.332
	G28628T	N	Não	A119S	P.2; B.1.1.28
	G28975T	N	Não	M234I	P.2
	C29754T	noncoding	—	—	P.2

Sub-clado P7	C1288T	ORF1ab	Sim	—	P.7
	G10870T	ORF1ab	Sim	—	P.7
	C7124T	ORF1ab	Não	P2287S	P.7
	C9344T	ORF1a	Não	L3027F	P.7
	G8027T	ORF1a	Não	V2588F	P.7
	G11596T	ORF1ab	Não	Q3777H	P.7
	C12053T	ORF1ab	Não	L3930F	B.1.1.28; P.2; B.1.1.332; P.7
	G25088T	S	Não	V1176F	B.1.1.28; P.1; P.2; P.7; B.1.1.332
	C25844T	ORF3a	Não	T151I	P.7; P.2

C28311T

N

Não

P13L

P.7

APÊNDICE D – DEFINIÇÕES DE HAPLÓTIPOS DE SARS-COV-2 ANALISADOS E SUAS FREQUÊNCIAS

Haplótipos	Frequência	ID das sequências
Hap_1	1	B_NC_045512.2
Hap_2	1	B.1.212_Brazil/PB-IEC161853
Hap_3	1	B.1.212_Brazil/PA-IEC173005
Hap_4	1	B.1.212_Brazil/PA-IEC173002
Hap_5	2	B.1.212_Brazil/AC-IEC162526, B.1.212_Brazil/CE-FIOCRUZ-17848
Hap_6	1	B.1.212_Brazil/AC-IEC162533
Hap_7	1	B.1_Brazil/RR-IEC176343
Hap_8	5	B.1_Colombia/VAC-UdeV-01-2, B.1_Brazil/SC-FIOCRUZ-769, B.1_Colombia/VAC-GVI-93381, B.1_Argentina/PAIS-G0002, B.1_Argentina/PAIS-D0012
Hap_9	1	B.1.212_Brazil/PA-IEC166687
Hap_10	1	B.1_Chile/MA-ISPCH-4
Hap_11	4	B.1_Brazil/SP-06, B.1_Brazil/RJ-DCVN3, B.1_Brazil/RJ-FIOCRUZ-314, B.1_Chile/RM-CMM-0000E
Hap_12	9	B.1_Colombia/NSA-NS-79852, B.1_Brazil/BA-24, B.1_Chile/RM-ISPCH-70, B.1_Chile/RM-ISPCH-30, B.1_Chile/Santiago-PUC_MVL_0519, B.1_Chile/Santiago-PUC_MVL_0043, B.1_Chile/Santiago-PUC_MVL_0047, B.1_Chile/Santiago-PUC_MVL_0077, B.1_Chile/Puerto_Montt-PUC_MVL_0246
Hap_13	4	B.1_Peru/LIM-INS-155, B.1_Chile/AR-ISPCH-14, B.1_Colombia/BOL-INS-84730, B.1_Colombia/BOY-INS-88859
Hap_14	1	B_Brazil/un-SP02cc

Hap_15	1	B.1_Chile/MA-ISPCH-11
Hap_16	1	B.1_Brazil/SE-FIOCRUZ-6555
Hap_17	1	B.1_Chile/Santiago-PUC_MVL_0540
Hap_18	1	B.1_Brazil/SE-FIOCRUZ-6539
Hap_19	1	B.1.212_Brazil/CE-FIOCRUZ-17833
Hap_20	1	B.1_Chile/Santiago-PUC_MVL_0271
Hap_21	1	B.1_Chile/Santiago-PUC_MVL_0289
Hap_22	1	B.1_Chile/RM-87464-TC1
Hap_23	1	B.1_Peru/JUN-INS-074
Hap_24	1	B.1.1.33_Brazil/DF-FIOCRUZ-861
Hap_25	19	B.1.1.33_Brazil/RJ-FIOCRUZ-2203, B.1.1.33_Brazil/DF-FIOCRUZ-891, B.1.1.33_Brazil/PR-FIOCRUZ-33316, B.1.1.33_Brazil/RJ-FIOCRUZ-2044, B.1.1.33_Chile/Santiago-PUC_MVL_0379, B.1.1.33_Chile/Santiago-PUC_MVL_0159, B.1.1.33_USA/FL-QDX-786, B.1.1.33_USA/FL-QDX-787, B.1.1.33_Brazil/SP-860, B.1.1.33_Brazil/RJ-INCA-C181, B.1.1.33_Brazil/RJ-INCA-C68, B.1.1.33_Brazil/RJ-00305, B.1.1.33_Brazil/RJ-00346, B.1.1.33_Brazil/SP-862, B.1.1.33_Chile/RM-134034, B.1.1.33_Brazil/RS-0313CVL, B.1.1.33_Brazil/RS-0372CVL, B.1.1.33_Brazil/AP-IEC161167, B.1.1.33_Brazil/AP-IEC163359
Hap_26	2	B.1.1.33_Brazil/RJ-FIOCRUZ-13416, B.1.1.33_Brazil/RJ-FIOCRUZ-12854
Hap_27	1	B.1.1.33_Brazil/RJ-FIOCRUZ-31113
Hap_28	1	B.1.1.33_Brazil/RJ-FIOCRUZ-2115
Hap_29	1	B.1.1.33_Brazil/SP-HIAE-ID29
Hap_30	1	B.1.1.33_Brazil/RJ-FIOCRUZ-7336A4
Hap_31	1	B.1.1.33_Brazil/CE-FIOCRUZ-17866

Hap_32	1	B.1.1.33_Australia/QLD1096
Hap_33	1	B.1.1.33_Brazil/PR-FIOCRUZ-10582-TG-109
Hap_34	1	B.1.1_Brazil/SE-FIOCRUZ-6561
Hap_35	22	B.1.1_Brazil/PR-FIOCRUZ-5616, B.1.1_Chile/Santiago-PUC_MVL_0024, B.1.1_Ecuador/USFQ-066, B.1.1.28_Brazil/AL-FIOCRUZ-837, B.1.1.28_Brazil/SP-516, B.1.1.28_Brazil/SP-831, B.1.1.28_Brazil/SP-629, B.1.1.28_Brazil/SP-826, B.1.1.28_Brazil/SP-810, B.1.1.28_Brazil/SP-459, B.1.1_Peru/LIM-INS-308, B.1.1.28_Brazil/SP-293, B.1.1.28_Brazil/BA-93, B.1.1_UFPE079, B.1.1_Brazil/PE-FIOCRUZ-IAM221, B.1.1_Brazil/PE-FIOCRUZ-IAM103, B.1.1_Peru/LIM-INS-340, B.1.1.28_Brazil/SP-585, B.1.1.28_Brazil/PA-IEC173014, B.1.1_Brazil/PE-UFPE054, B.1.1_Brazil/PE-UFPE056, B.1.1_Brazil/PR-IPEC-103
Hap_36	1	B.1.1_Brazil/PB-FIOCRUZ-16577
Hap_37	1	B.1.1.33_Brazil/PR-FIOCRUZ-33294
Hap_38	1	B.1.1.33_Brazil/PR-FIOCRUZ-10623-TG-112
Hap_39	1	B.1.1_Brazil/PB-FIOCRUZ-16583
Hap_40	1	B.1.1.33_Brazil/PR-FIOCRUZ-11065-TG-121
Hap_41	1	B.1.1.33_Brazil/PR-FIOCRUZ-11212
Hap_42	1	B.1.1_Brazil/SE-FIOCRUZ-6530
Hap_43	1	B.1.1_Brazil/PB-FIOCRUZ-16524
Hap_44	1	B.1.1.33_Brazil/RJ-FIOCRUZ-1574
Hap_45	1	B.1.1_Chile/RM-87391-TC1
Hap_46	1	B.1.1_Peru/LIM-INS-147
Hap_47	1	B.1.1.33_Brazil/PR-FIOCRUZ-10994-TG-118
Hap_48	1	B.1.1.33_Brazil/RJ-FIOCRUZ-13870

Hap_49	1	B.1.1.33_Brazil/PR-FIOCRUZ-9931-TG-176
Hap_50	1	B.1.1.33_Brazil/RJ-FIOCRUZ-1464
Hap_51	1	B.1.1.33_Brazil/RJ-FIOCRUZ-2000
Hap_52	1	B.1.1.33_Brazil/GO-FIOCRUZ-13916
Hap_53	1	B.1.1_Brazil/PB-FIOCRUZ-16590
Hap_54	1	B.1.1_Chile/Santiago-GPJ5829
Hap_55	1	B.1_Chile/Santiago-MAA978K
Hap_56	1	B.1_Chile/Santiago-PUC_MVL_0026
Hap_57	1	B.1_Chile/Santiago-PUC_MVL_0068
Hap_58	1	B.1.1_Colombia/VAC-GVI-93467
Hap_59	1	B.1.1_Venezuela/VEN-89312
Hap_60	1	B.1.1_Chile/Santiago-PUC_MVL_0305
Hap_61	1	B.1.1_Colombia/GUR-06537
Hap_62	1	B.1.1_Chile/Santiago-PUC_MVL_0279
Hap_63	1	B.1.1_Chile/Santiago-AGR8546
Hap_64	1	B.1.1_Colombia/NAR-GVI-97417
Hap_65	1	B.1_Chile/Santiago-PUC_MVL_0096
Hap_66	1	B.1_Colombia/GUR-06406
Hap_67	1	B.1.1_Brazil/PR-FIOCRUZ-15066
Hap_68	1	B.1.212_Brazil/CE-L16-CD346
Hap_69	1	B.1.1.28_Brazil/SP-L4-CAMPI65
Hap_70	2	B.1.1.28_England/MILK-ABBF00, B.1.1.28_England/MILK-ABBBBE
Hap_71	1	B.1.1.33_Canada/ON-PHL-20-01468

Hap_72	1	B.1.1.33_USA/IL-IDPH-COO-C-000829
Hap_73	2	B.1.1.33_USA/IL-IDPH-MCD-S-000879, B.1.1.33_USA/IL-IDPH-S-000369
Hap_74	1	B.1.1.33_USA/TX-HMH098
Hap_75	2	B.1.1.33_USA/WI-UW-198, B.1.1.33_Chile/RM-ISPCH-71
Hap_76	1	B.1.1_Chile/RM-CMM-A2P533884024
Hap_77	2	B.1.1.33_Canada/ON-PHL-20-01438, B.1.1.33_Canada/ON-SC0139
Hap_78	1	B.1.1.33_Canada/ON-UHTC_0215
Hap_79	1	B.1_Argentina/PAIS-E0045
Hap_80	3	B.1_Argentina/PAIS-E0134, B.1_Argentina/PAIS-C0035, B.1_Argentina/PAIS-C0030
Hap_81	1	B.1_Argentina/PAIS-E0144
Hap_82	1	B.1_Ecuador/USFQ-097
Hap_83	1	B.1_Ecuador/USFQ-515
Hap_84	1	B.1_Ecuador/USFQ-539
Hap_85	2	B.1.1_Ecuador/USFQ-045, B.1.1_Ecuador/USFQ-114
Hap_86	1	B.1.1_Ecuador/USFQ-147
Hap_87	2	B.1.1.33_England/MILK-9A7F6F, B.1.1.33_England/QEUA-A1F1A7
Hap_88	1	B.1.1_Colombia/GUR-02256
Hap_89	1	B.1.1.28_Brazil/SC-FIOCRUZ-34070-R1
Hap_90	1	B.1.1.28_Brazil/PR-FIOCRUZ-11653-TG-148
Hap_91	1	B.1.1.28_Brazil/PR-FIOCRUZ-11632-TG-134
Hap_92	1	B.1.1.28_Brazil/RS-123_LABRESIS
Hap_93	1	B.1.1.28_Brazil/SP-820
Hap_94	1	B.1.1.28_Brazil/SP-438

Hap_95	1	B.1.1.28_Brazil/SP-832
Hap_96	1	B.1.1.28_Brazil/SP-340
Hap_97	1	B.1.1.28_Brazil/SP-445
Hap_98	1	B.1.1.28_Brazil/SP-818
Hap_99	1	B.1.1.28_Brazil/PE-UFPE032
Hap_100	1	B.1.1.28_Brazil/SP-136
Hap_101	1	B.1.1.28_Brazil/SP-218
Hap_102	1	B.1.1.28_Brazil/SP-559
Hap_103	1	B.1.1.28_Brazil/SP-822
Hap_104	1	B.1.1.33_Brazil/SP-177
Hap_105	1	B.1.1.33_Brazil/SP-163
Hap_106	1	B.1.1.33_Brazil/SP-164
Hap_107	1	B.1.1.33_Brazil/BA-96
Hap_108	1	B.1.1_Peru/LIM-INS-318
Hap_109	1	B.1.1.33_Brazil/SP-299
Hap_110	1	B.1.1.28_Brazil/SP-301
Hap_111	1	B.1.1.33_Brazil/BA-01
Hap_112	1	B.1.1.33_Brazil/BA-83
Hap_113	1	B.1.1.28_Brazil/BA-88
Hap_114	1	B.1.1_Brazil/PE-UFPE049
Hap_115	1	B.1.1_Brazil/PE-UFPE023
Hap_116	1	B.1.1.28_Brazil/PE-UFPE052
Hap_117	1	B.1.1_Kenya/COV-MBB-07-075

Hap_118	1	B.1.1_UFPE097
Hap_119	1	B.1.1_UFPE074
Hap_120	2	B.1.1_UFPE072, B.1.1_UFPE075
Hap_121	1	B.1.1_UFPE080
Hap_122	1	B.1.1_Brazil/PE-UFPE071
Hap_123	1	B.1.1_Brazil/PE-UFPE005
Hap_124	1	B.1.1_UFPE110
Hap_125	1	B.1.1_UFPE111
Hap_126	1	B.1.1_UFPE119
Hap_127	3	B.1.1.28_UFPE094, B.1.1.28_UFPE095, B.1.1.28_UFPE096
Hap_128	1	B.1.1.28_UFPE083
Hap_129	1	B.1.1.28_UFPE081
Hap_130	1	B.1.1.28_UFPE100
Hap_131	1	B.1.1.28_UFPE101
Hap_132	2	B.1.1.28_UFPE138, B.1.1_UFPE136
Hap_133	1	B.1.1_Brazil/PE-FIOCRUZ-IAM109
Hap_134	1	B.1.1_Brazil/PE-FIOCRUZ-IAM311
Hap_135	1	B.1.1.33_Brazil/RJ-INCA-C23
Hap_136	1	B.1.1_Brazil/PE-FIOCRUZ-IAM2116
Hap_137	1	B.1.1_Peru/LIM-INS-315
Hap_138	1	B.1.1_Brazil/PE-FIOCRUZ-IAM2245
Hap_139	5	B.1.1_Brazil/PE-FIOCRUZ-IAM2236, B.1.1_Brazil/PE-FIOCRUZ-IAM2160, B.1.1_Brazil/PE-FIOCRUZ-IAM2140, B.1.1_Brazil/PE-FIOCRUZ-IAM2238, B.1.1_UFPE087

Hap_140	2	B.1.1.33_Brazil/PE-FIOCRUZ-IAM2182, B.1.1.33_Brazil/PE-FIOCRUZ-IAM2229
Hap_141	1	B.1.1_Brazil/RJ-00402
Hap_142	1	B.1.1_Brazil/PE-FIOCRUZ-IAM2256
Hap_143	1	B.1.1_Brazil/PE-FIOCRUZ-IAM1775
Hap_144	1	B.1.1.28_Brazil/SP-963
Hap_145	1	B.1.1.28_Brazil/SP-523
Hap_146	1	B.1.1.33_Brazil/SP-399
Hap_147	1	B.1.1.33_Brazil/RJ-00307
Hap_148	1	B.1.1.28_Brazil/RJ-00409
Hap_149	1	B.1.1.33_Brazil/RJ-00329
Hap_150	1	B.1.1.33_Brazil/RS-32833
Hap_151	1	B.1.1_Peru/LIM-UPCH-0149
Hap_152	1	B.1.1_Peru/ARE-INS-631
Hap_153	1	B.1.1.28_Brazil/GO-HLAGYN-870588_R1
Hap_154	1	B.1.1.28_Brazil/RS-UFCSPA_LACEN-89673
Hap_155	1	B.1.1.28_Scotland/CVR3315
Hap_156	1	B.1.1.28_Brazil/SP-140
Hap_157	1	B.1.1.28_Brazil/BA-LACEN-121
Hap_158	1	B.1.1_Brazil/PE-UFPE037
Hap_159	1	B.1.1.33_Brazil/RJ-FIOCRUZ-2811
Hap_160	1	B.1.1.33_Brazil/MS-1310
Hap_161	1	B.1.1.33_Brazil/MS-1289
Hap_162	1	B.1.1.33_Brazil/RS-20020CVL

Hap_163	1	B.1.1_Chile/RM-112882
Hap_164	1	B.1.1_Chile/BI-206407
Hap_165	1	B.1.1.33_Brazil/RS-17663CVL
Hap_166	1	B.1.1.33_Brazil/RS-17114CVL
Hap_167	1	B.1.1.33_Brazil/RS-0200CVL
Hap_168	1	B.1.1.33_Brazil/RS-0237CVL
Hap_169	1	B.1.1.33_Brazil/RS-1432CVL
Hap_170	1	B.1.1.33_Brazil/RS-0343CVL
Hap_171	1	B.1.1.33_Brazil/RS-5377CVL
Hap_172	1	B.1.1.33_Brazil/RS-13407CVL
Hap_173	1	B.1.1.33_Brazil/RS-26276CVL
Hap_174	1	B.1.1.33_Brazil/RS-31975CVL
Hap_175	1	B.1.1.33_Brazil/RS-19274CVL
Hap_176	1	B.1.1.33_Chile/RM-82488
Hap_177	1	B.1.1.33_Chile/Santiago-PUC_MVL_0384
Hap_178	1	B.1.212_Brazil/RO-04
Hap_179	1	B.1.212_Brazil/RO-06
Hap_180	1	B.1.212_Brazil/AM-FIOCRUZ-20892300TV
Hap_181	1	B.1.195_Brazil/AM-FIOCRUZ-20140481LB
Hap_182	1	B.1.195_Brazil/AM-FIOCRUZ-20140522MA
Hap_183	3	B.1.195_Brazil/AM-FIOCRUZ-20890023DP, B.1.195_Brazil/AM-FIOCRUZ-20890128FE, B.1.195_Brazil/AM-FIOCRUZ-20890140CD
Hap_184	1	B.1.195_Brazil/AM-FIOCRUZ-20892241JS

Hap_185	1	B.1.195_Brazil/AM-FIOCRUZ-20890223JC
Hap_186	1	B.1.1.28_Brazil/AM-FIOCRUZ-20141355VB
Hap_187	1	B.1.1.28_Brazil/AM-FIOCRUZ-20891898RL
Hap_188	1	B.1.1.28_Brazil/AM-FIOCRUZ-20891680JR
Hap_189	1	B.1.1_Chile/RM-72815
Hap_190	1	B.1.1.33_Brazil/SE-FIOCRUZ-6556
Hap_191	1	B.1.212_Brazil/CE-FIOCRUZ-17832
Hap_192	1	B.1.1.33_Brazil/RJ-FIOCRUZ-7633
Hap_193	1	B.1.1.33_Brazil/RJ-FIOCRUZ-7634
Hap_194	1	B.1.1.33_Brazil/RJ-FIOCRUZ-14012
Hap_195	1	B.1.1.33_Brazil/GO-FIOCRUZ-13602s2
Hap_196	1	B.1.1.33_Brazil/BA-FIOCRUZ-33265
Hap_197	1	B.1.1.28_Brazil/MG-FIOCRUZ-34172
Hap_198	1	B.1.1.28_Brazil/PR-FIOCRUZ-11074-TG-122
Hap_199	1	B.1.1.28_Portugal/PT1822
Hap_200	1	B.1.1.28_Brazil/TO-1173
Hap_201	1	B.1.1.161_Brazil/RS-11JCJHCPA
Hap_202	1	B.1.1.161_Brazil/RS-130_LABRESIS
Hap_203	1	B.1.1.33_Brazil/RS-141_LABRESIS
Hap_204	1	B.1.1.161_Brazil/RS-127_LABRESIS
Hap_205	1	B.1.1.161_Brazil/RS-138_LABRESIS
Hap_206	1	B.1.1.161_Brazil/RS-151_LABRESIS
Hap_207	1	B.1.1.161_Brazil/RS-137_LABRESIS

Hap_208	1	B.1.1.161_Brazil/RS-117_LABRESIS
Hap_209	1	B.1.1.161_Brazil/RS-26AVGFHCPA
Hap_210	1	B.1.1_Bolivia/26937
Hap_211	1	B.1.1.161_Brazil/RS-169_LABRESIS
Hap_212	1	B.1.1.161_Brazil/RS-142_LABRESIS
Hap_213	1	B.1.1.33_Brazil/RS-FIOCRUZ-6187
Hap_214	1	B.1.1_Brazil/RJ-FIOCRUZ-11493
Hap_215	1	B.1.1_Bolivia/26946
Hap_216	1	B.1.1.33_Brazil/SP-318
Hap_217	1	B.1.1_Brazil/PB-FIOCRUZ-29644
Hap_218	1	B.1.1.33_Brazil/RJ-FIOCRUZ-17588
Hap_219	1	B.1.1_Chile/AR-208476
Hap_220	1	B.1.1.33_Brazil/RN-FIOCRUZ-34124
Hap_221	1	B.1.1.33_Brazil/RJ-FIOCRUZ-20195
Hap_222	1	B.1.1.33_Brazil/RJ-FIOCRUZ-12794
Hap_223	1	B.1.1.33_Brazil/MG-FIOCRUZ-69
Hap_224	1	B.1.1.33_Brazil/SP-957
Hap_225	1	B.1.1.28_Brazil/RS-FIOCRUZ-15284
Hap_226	1	B.1.1.28_Brazil/SP-817
Hap_227	1	B.1.1.28_Brazil/SP-857
Hap_228	1	B.1.1.28_Uruguay/CUY1-000134-UYTA
Hap_229	1	B.1.1.33_Brazil/BA-23
Hap_230	1	B.1.1_Ecuador/UEES-ECU404

Hap_231	1	B.1.1_Ecuador/UEES-ECU469
Hap_232	1	B.1.1.33_Argentina/PAIS-G0079
Hap_233	1	B.1.1_Peru/LIM-UPCH-0128
Hap_234	1	B.1_Peru/LIM-UPCH-0056
Hap_235	1	B.1_Peru/ARE-INS-628
Hap_236	1	B.1_Peru/LIM-INS-523
Hap_237	1	B.1_Chile/RM-88024-TC1
Hap_238	1	B.1_Argentina/PAIS-G0085
Hap_239	1	B.1.1.28_Brazil/SP-423
Hap_240	1	B.1.1.28_Brazil/SP-1104
Hap_241	1	B.1.1.28_Qatar/QA-WCMQ_FD18190004
Hap_242	1	B.1.1.28_Brazil/SP-702
Hap_243	1	B.1.1.28_Brazil/AM-FIOCRUZ-20898137RM
Hap_244	1	B.1.1.28_Brazil/AM-FIOCRUZ-20142424SS
Hap_245	1	B.1.1.28_Brazil/AM-FIOCRUZ-20892948LS
Hap_246	1	B.1.1.28_Brazil/SP-208
Hap_247	1	B.1.1.28_Brazil/SP-661
Hap_248	1	B.1.1.28_Brazil/SP-359
Hap_249	1	B.1.1_Brazil/PE-UFPE003
Hap_250	1	B.1.1.28_Brazil/SP-417
Hap_251	1	B.1.1.28_Brazil/SP-667
Hap_252	1	B.1.1.33_Brazil/SP-594
Hap_253	1	B.1.1.33_Brazil/SP-658

Hap_254	1	B.1.1.33_Brazil/RJ-UFRJ-57721
Hap_255	1	B.1.1.33_Brazil/AC-IEC168034
Hap_256	1	B.1.1.33_Chile/RM-ISPCH-73
Hap_257	1	B.1.1.33_UFPE078
Hap_258	1	B.1.1.33_Brazil/RS-24277
Hap_259	1	B.1.1.33_Brazil/RJ-00456
Hap_260	2	B.1.1.33_Uruguay/RIV-M5, B.1.1.33_Uruguay/RIV-M13
Hap_261	1	B.1.1.28_Brazil/PA-IEC172962
Hap_262	1	B.1.1.28_UFPE073
Hap_263	1	B.1.1_UFPE089
Hap_264	1	B.1.1.33_USA/MA-MGH-01491
Hap_265	1	B.1.1.33_Brazil/RR-IEC176364
Hap_266	1	B.1.173_UFPE090
Hap_267	1	B.1.173_UFPE135
Hap_268	1	B.1.1.28_Brazil/RJ-00419
Hap_269	2	B.1.1_Brazil/PE-UFPE013, B.1.1_Brazil/PE-UFPE012
Hap_270	1	B.1.1_Brazil/PE-UFPE025
Hap_271	1	B.1.1_Brazil/PE-UFPE061
Hap_272	1	B.1.1_Brazil/PE-UFPE068
Hap_273	1	B.1.1_Peru/LAM-INS-721
Hap_274	1	B.1.1.33_Brazil/PE-UFPE018
Hap_275	1	B.1.1.33_Brazil/RJ-00342
Hap_276	1	B.1.1_Chile/AT-CMM-ATAC05

Hap_277	1	B.1.1.371_UFPE118
Hap_278	1	B.1.1.371_UFPE126
Hap_279	1	B.1.1_Brazil/PE-UFPE067
Hap_280	1	B.1.1.33_Brazil/PR-IPEC-135
Hap_281	1	B.1.1.33_Brazil/PR-IPEC-144
Hap_282	1	B.1.1_Brazil/PR-IPEC-139
Hap_283	1	B.1.1.33_Australia/QLD1184
Hap_284	1	B.1.1.33_UFPE106
Hap_285	1	B.1.1.33_UFPE084
Hap_286	1	B.1.1.33_Brazil/RJ-00458
Hap_287	1	B.1.1.28_Indonesia/JI-ITD-26463NTv
Hap_288	1	B.1.1.28_UFPE120
Hap_289	1	B.1.1.28_UFPE131
Hap_290	1	B.1.1.28_UFPE086
Hap_291	1	B.1.1.28_Philippines/PH-PGC-01559
Hap_292	1	B.1.1_Ecuador/USFQ-206
Hap_293	1	B.1.1_Peru/LIM-INS-527
Hap_294	1	B.1.1_Peru/LIM-INS-530
Hap_295	1	B.1.1_UFPE076
Hap_296	1	B.1.1.192_UFPE132
Hap_297	1	B.1.1_UFPE133
Hap_298	1	B.1.1.28_UFPE099
Hap_299	1	B.1.1.28_UFPE085

Hap_300	1	B.1.1.28_UFPE093
Hap_301	2	B.1.1_UFPE128, B.1.1_UFPE129
Hap_302	1	B.1.1.28_UFPE088
Hap_303	1	B.1.1.28_Brazil/PR-IPEC-048
Hap_304	1	B.1.1.28_Brazil/PE-FIOCRUZ-IAM1833
Hap_305	1	B.1.1.28_Brazil/PE-FIOCRUZ-IAM2233
Hap_306	1	B.1.1.28_UFPE082
Hap_307	1	B.1.1.28_Brazil/PE-FIOCRUZ-IAM1834
Hap_308	1	B.1.1.28_UFPE102
Hap_309	1	B.1.1.28_UFPE103
Hap_310	1	B.1.1.28_UFPE105
Hap_311	1	B.1.1.28_UFPE104
Hap_312	1	B.1.1.28_UFPE127
Hap_313	1	B.1.1_Brazil/PE-FIOCRUZ-IAM2131
Hap_314	1	B.1.1.28_UFPE077
Hap_315	1	B.1.1.28_UFPE108
Hap_316	1	B.1.1.28_UFPE109
Hap_317	1	B.1.1.371_Brazil/PE-FIOCRUZ-IAM1818
Hap_318	1	B.1.1_Brazil/PE-FIOCRUZ-IAM1825
Hap_319	1	B.1.1_Brazil/PE-FIOCRUZ-IAM2247
Hap_320	2	B.1.1_Brazil/PE-FIOCRUZ-IAM1804, B.1.1_Brazil/PE-FIOCRUZ-IAM1820
Hap_321	1	B.1.1_Brazil/PE-FIOCRUZ-IAM2147
Hap_322	1	B.1.1.33_Brazil/MS-1275

Hap_323	1	B.1.1.28_Brazil/PR-IPEC-016
Hap_324	1	B.1.1.33_Brazil/SP-1099
Hap_325	1	B.1.1.28_UFPE113
Hap_326	1	B.1.1.28_UFPE114
Hap_327	1	B.1.1.33_Brazil/RJ-FIOCRUZ-22685
Hap_328	1	B.1.1.33_Brazil/AM-FIOCRUZ-20140447AV
Hap_329	1	B.1.1_Argentina/PAIS-A0005
Hap_330	1	B.1.1_Peru/LIM-INS-184
Hap_331	1	B.1.1_Peru/LIM-INS-283
Hap_332	1	B.1.1_Peru/LIM-INS-769
Hap_333	1	B.1.1_Peru/MOQ-INS-836
Hap_334	1	B.1.1_Peru/LIM-INS-828
Hap_335	1	B.1.1_Peru/LIM-UPCH-0238
Hap_336	1	B.1.1.28_Brazil/SP-L7-CD79
Hap_337	1	B.1.1.28_Brazil/RS-8698CVL
Hap_338	1	B.1.1_Brazil/RS-UFCSPA_LACEN-46396
Hap_339	1	B.1.1_Argentina/PAIS-E0124
Hap_340	1	B.1.499_Argentina/PAIS-G0114
Hap_341	1	B.1.499_Argentina/PAIS-G0112
Hap_342	1	B.1_Argentina/PAIS-A0179
Hap_343	1	B.1_Argentina/PAIS-E0023
Hap_344	1	B.1.1_Chile/AR-192891
Hap_345	1	B.1.1.33_Brazil/RS-18490CVL

Hap_346	1	B.1.1.33_Brazil/RS-13715CVL
Hap_347	1	B.1.1.28_Brazil/DF-L124-CD4042
Hap_348	1	B.1.1.28_Brazil/PE-FIOCRUZ-IAM1790
Hap_349	1	B.1.1.28_Brazil/RS-FUNED-904220
Hap_350	1	P.7_Brazil/RS-00666
Hap_351	1	P.7_Brazil/GO-HLAGYN-1308717
Hap_352	1	P.7_Brazil/PB-BA1200-250032126
Hap_353	1	P.7_Brazil/RS-34614CVL
Hap_354	1	P.7_Brazil/SC-FIOCRUZ-33484
Hap_355	1	P.7_Japan/TY7-416-P1
Hap_356	1	B.1_Brazil/PR-FIOCRUZ-5622
Hap_357	1	B.1_Peru/LIM-INS-090
Hap_358	1	B.1.1.33_Brazil/RJ-FIOCRUZ-13784
Hap_359	1	B.1.1_Brazil/RJ-FIOCRUZ-11485
Hap_360	1	B.1.1_Brazil/PB-FIOCRUZ-29628
Hap_361	1	B.1.1.33_Brazil/MT-FIOCRUZ-21770
Hap_362	1	B.1.1.33_Brazil/BA-FIOCRUZ-21786
Hap_363	1	B.1.1.33_Chile/RM-171340
Hap_364	1	B.1.1_Chile/RM-110987
Hap_365	1	B.1.1_Chile/RM-157016
Hap_366	1	B.1.1_Chile/RM-170990
Hap_367	1	B.1.1_Chile/RM-192950
Hap_368	1	B.1.1.28_Brazil/SC-FIOCRUZ-34078-R1

Hap_369	1	B.1.1_Brazil/RJ-FIOCRUZ-2930
Hap_370	1	B.1.1.28_Brazil/BA-FIOCRUZ-33254
Hap_371	1	B.1.1.28_Brazil/SP-1100
Hap_372	1	B.1.1.28_Brazil/RJ-FIOCRUZ-13013
Hap_373	1	B.1_Colombia/AMA-GUR-06747
Hap_374	1	B.1_Colombia/AMA-GUR-06818
Hap_375	1	B.1_Colombia/AMA-GUR-06777
Hap_376	1	B.1.1.33_Brazil/ES-FIOCRUZ-33712
Hap_377	1	N.9_Brazil/SP-883
Hap_378	1	N.9_Brazil/BA-FIOCRUZ-33636
Hap_379	1	B.1.1.28_Brazil/SP-803
Hap_380	1	B.1.1.28_Brazil/SP-878
Hap_381	1	B.1.1.28_Brazil/RJ-FIOCRUZ-23375
Hap_382	1	B.1.1.28_Portugal/PT1957
Hap_383	1	B.1.1.28_Brazil/RJ-FIOCRUZ-18566
Hap_384	1	B.1.1.28_Brazil/RS-34418CVL
Hap_385	1	B.1.1.33_Brazil/RS-16174CVL
Hap_386	1	B.1.1.33_Brazil/RS-7774CVL
Hap_387	1	B.1.1.28_Brazil/PB-IEC-177526
Hap_388	1	B.1.1.28_Brazil/SP-UW-651271524001
Hap_389	1	B.1.1.28_Brazil/RN-FIOCRUZ-34120
Hap_390	1	B.1.1.33_Brazil/MG-UW-651313414501
Hap_391	1	B.1.1.28_Brazil/SP-1146

Hap_392	3	B.1.1.28_UFPE115, B.1.1.28_UFPE116, B.1.1.28_UFPE117
Hap_393	1	B.1.1.28_Brazil/TO-1184
Hap_394	1	B.1.1_Brazil/SP-781
Hap_395	1	B.1.1.33_Brazil/RJ-00574
Hap_396	1	B.1.1.33_Brazil/RJ-FIOCRUZ-16446
Hap_397	1	B.1.1.33_Brazil/RJ-FIOCRUZ-7560
Hap_398	1	B.1.1.33_Suriname/SR-94
Hap_399	1	B.1.1.33_Switzerland/VS-ETHZ-270176
Hap_400	1	B.1.1_Ecuador/EC-E-124319
Hap_401	1	B.1.1.33_UFPE112
Hap_402	1	B.1.1.33_UFPE137
Hap_403	1	B.1.1.33_UFPE107
Hap_404	1	B.1.1.33_UFPE098
Hap_405	1	B.1.1.33_Brazil/AC-43TFR
Hap_406	1	B.1.1_Brazil/PB-FIOCRUZ-29631
Hap_407	1	B.1.1.28_Brazil/PR-FIOCRUZ-15062-TG-158
Hap_408	1	B.1.1_Brazil/PE-FIOCRUZ-IAM215
Hap_409	1	B.1.1_Peru/LOR-INS-584
Hap_410	1	B.1.1_Peru/LOR-INS-601
Hap_411	1	B.1.1.28_Brazil/RJ-00420
Hap_412	1	B.1_Peru/LIM-INS-136
Hap_413	1	B.1.1.33_Brazil/RJ-00479
Hap_414	1	B.1.1.33_Brazil/PR-IPEC-072

Hap_415	1	B.1.1.33_Brazil/RJ-00497
Hap_416	1	B.1.1.33_UFPE091
Hap_417	1	B.1.1.33_Brazil/RS-1061CVL
Hap_418	1	B.1.1_Brazil/PA-IEC176486
Hap_419	1	B.1.1.28_Philippines/PH-PGC-01943
Hap_420	1	B.1.1_Brazil/PR-IPEC-152
Hap_421	1	B.1.1.28_Brazil/PR-IPEC-101
Hap_422	1	B.1.1.33_Brazil/PR-IPEC-021
Hap_423	1	B.1.1.33_Brazil/PR-IPEC-078
Hap_424	1	B.1.212_Brazil/PE-UFPE022
Hap_425	1	B.1.212_Brazil/RN-FIOCRUZ-34132
Hap_426	1	B.1.91_Brazil/RS-FIOCRUZ-34195
Hap_427	1	B.1.91_Brazil/RS-FIOCRUZ-6237
Hap_428	1	B.1_Chile/Santiago-PUC_MVL_0254
Hap_429	1	B.1_Colombia/VAC-GVI-97181
Hap_430	1	B.1.1_Brazil/RN-BA1200-240140562
Hap_431	1	N.9_Brazil/PB-BA1200-250036191
Hap_432	1	N.9_Brazil/MA-IEC1200-177755
Hap_433	1	B.1.1.33_Brazil/MA-IEC1200-177752
Hap_434	1	B.1_Argentina/PAIS-A0205
Hap_435	1	B.1.1.28_Brazil/TO-1186
Hap_436	1	B.1.1.28_Portugal/PT1952
Hap_437	1	P.2_Brazil/RJ-00537

Hap_438	1	P.2_Brazil/RJ-00553
Hap_439	1	P.2_Brazil/RS-00633
Hap_440	1	P.2_Brazil/RS-00653
Hap_441	1	P.2_Brazil/RS-31800
Hap_442	1	P.2_Brazil/PR-IPEC-022
Hap_443	1	P.2_Brazil/PR-IPEC-039
Hap_444	1	P.2_Brazil/SP-891
Hap_445	1	P.2_Brazil/CE-IEC-177343
Hap_446	3	P.2_Brazil/MG-FUNED-828-20US, P.2_UFPE124, P.2_UFPE125
Hap_447	1	P.2_Brazil/AP-IEC1200-178185
Hap_448	1	P.2_Brazil/AP-IEC1200-178192
Hap_449	1	P.2_Brazil/PA-IEC-176581
Hap_450	1	P.2_Brazil/PA-IEC-177042
Hap_451	1	P.2_Brazil/PR-FUNED-51538-20
Hap_452	1	P.2_UFPE121
Hap_453	2	P.2_UFPE122, P.2_UFPE123
Hap_454	1	P.2_UFPE134
Hap_455	1	P.2_UFPE130
Hap_456	1	P.2_Brazil/AP-IEC1200-178194
Hap_457	1	P.2_Brazil/RJ-FUNED-330157046
Hap_458	1	P.2_Brazil/AM-996
Hap_459	1	P.2_Brazil/PR-FUNED-57868-20
Hap_460	1	P.2_Brazil/PR-FUNED-56917-20

Hap_461	1	P.2_Brazil/PR-FUNED-55927-20
Hap_462	1	P.2_Brazil/RS-FUNED-837150
Hap_463	1	P.2_Brazil/RJ-FIOCRUZ-25304
Hap_464	1	P.2_Brazil/RJ-FIOCRUZ-32593
Hap_465	1	P.2_Brazil/RN-FIOCRUZ-34150
Hap_466	1	P.2_Brazil/MA-FIOCRUZ-34251
Hap_467	1	P.2_Brazil/PR-FIOCRUZ-28608
Hap_468	2	P.2_Brazil/RJ-FIOCRUZ-27541-1P, P.2_Brazil/RJ-FIOCRUZ-30991
Hap_469	1	P.2_Brazil/PR-FIOCRUZ-33174-TG-169
Hap_470	1	P.2_Brazil/SC-FIOCRUZ-33477
Hap_471	1	P.2_Brazil/SC-FIOCRUZ-33492
Hap_472	2	P.2_Brazil/RJ-FIOCRUZ-32807, P.2_Brazil/RJ-FIOCRUZ-32806
Hap_473	1	P.2_Brazil/PB-00762
Hap_474	1	B.1.1.28_Brazil/SP-IB_103268/2021
Hap_475	1	P.2_Brazil/PA-IEC1200-177707
Hap_476	1	P.2_Brazil/BA-LACEN-164
Hap_477	1	P.2_Brazil/SP-986
Hap_478	1	P.2_Brazil/MA-FIOCRUZ-34257
Hap_479	1	P.2_Brazil/RJ-LVM29614
Hap_480	1	P.2_Brazil/RJ-LVM29784
Hap_481	1	P.2_Brazil/RJ-00560
Hap_482	1	P.2_Brazil/MG-FIOCRUZ-31822
Hap_483	1	P.2_Brazil/GO-HLAGYN-1299078

Hap_484	1	P.2_Brazil/RS-UFCSPA_LACEN-106932
Hap_485	1	P.2_Brazil/CE-FIOCRUZ-33681
Hap_486	1	P.2_Brazil/SC-FIOCRUZ-33502
Hap_487	1	P.2_Brazil/SC-FIOCRUZ-33468
Hap_488	1	P.2_Brazil/RS-00613
Hap_489	1	P.2_Brazil/PA-IEC1200-177706
Hap_490	1	P.2_Brazil/RJ-FUNED-330173180
Hap_491	1	P.2_Brazil/RS-UFCSPA_LACEN-106835
Hap_492	1	B.1.1.332_Brazil/SP-HIAE-ID16
Hap_493	1	B.1.1.332_Brazil/MA-FIOCRUZ-34262
Hap_494	1	B.1.1.28_Brazil/RS-00691
Hap_495	1	B.1.1.28_Brazil/RS-23278CVL
Hap_496	1	B.1.1.28_Brazil/MG-FUNED-829-20US
Hap_497	1	P.7_Brazil/SP-884
Hap_498	1	B.1.1.28_Brazil/31GGPHCPA
Hap_499	1	B.1.1.28_Brazil/RS-00602
Hap_500	1	B.1.1.28_Brazil/MG-FIOCRUZ-34210
Hap_501	1	B.1.1_UFPE092
Hap_502	1	B.1.1_Brazil/PE-FIOCRUZ-IAM1798
Hap_503	1	B.1.1_Chile/ML-260119
Hap_504	1	B.1.1_Brazil/RS-12423
Hap_505	1	B.1.1.28_Brazil/MS-1251
Hap_506	1	B.1.1.33_Brazil/PB-FIOCRUZ-29608

Hap_507	1	B.1.1.33_USA/NY-PRL-2021_0702_51G01/2021
Hap_508	1	B.1.1.33_Brazil/SP-HIAE-ID19
Hap_509	1	B.1.1.33_Brazil/MG-FIOCRUZ-34167
Hap_510	1	B.1.1.33_USA/AZ-ASPHL-2655
Hap_511	1	B.1.1.28_Brazil/PB-BA1200-250028130
Hap_512	1	B.1.1_Brazil/PA-IEC-177041
Hap_513	1	B.1.1.28_Brazil/SP-HIAE-ID02
Hap_514	1	B.1.1.28_Brazil/GO-HLAGYN-1299079
Hap_515	1	B.1.1.33_Brazil/PE-FIOCRUZ-IAM2174
Hap_516	1	B.1.1_Chile/RM-251740
Hap_517	1	B.1.1_Colombia/SAN-576900-032-RNA
Hap_518	1	B.1.1.33_Brazil/RJ-00512
Hap_519	1	B.1.1.33_Brazil/RJ-00578
Hap_520	1	B.1.1_Brazil/PB-00797
Hap_521	1	B.1.1.28_Brazil/AM-FIOCRUZ-20843023JA
Hap_522	1	B.1.1_Chile/RM-183871
Hap_523	1	B.1.1.28_Brazil/RJ-LVM29757
Hap_524	1	P.7_Brazil/RS-FIOCRUZ-33192
Hap_525	1	P.7_Brazil/RS-FIOCRUZ-33200
Hap_526	1	B.1.1_Brazil/RJ-FIOCRUZ-12197
Hap_527	1	B.1.195_Brazil/AM-FIOCRUZ-20891507CM
Hap_528	2	B.1.195_Brazil/AM-FIOCRUZ-20892520WS, B.1.195_Brazil/AM-FIOCRUZ-20892521AG
Hap_529	1	B.1.1_Chile/RM-131478

Hap_530	1	B.1.1.28_Brazil/MG-FIOCRUZ-34181
Hap_531	1	B.1.1.28_England/ALDP-C3AC67
Hap_532	1	P.1_Brazil/AM-FIOCRUZ-20143093FL
Hap_533	1	P.1_Brazil/BA-FIOCRUZ-33769
Hap_534	1	P.1_Brazil/AM-IEC-177294
Hap_535	1	P.1_Brazil/AM-FIOCRUZ-20890585EFV
Hap_536	1	B.1.1.33_Brazil/RS-13648CVL
Hap_537	1	B.1.1.28_South_Korea/KDCA0101
Hap_538	1	B.1_Colombia/CES-INS-VG-248
Hap_539	1	B.1.212_Brazil/PI-BA1200-220043219
Hap_540	1	P.7_Brazil/RS-34618CVL
Hap_541	1	B.1.91_Brazil/RS-00311HM_LMM52637