

## A importância da vigilância genômica no contexto da pandemia de COVID-19 para o Estado do Amazonas

### The importance of genomic surveillance in the context of the COVID-19 pandemic for the State of Amazonas

Yasmim Fujimoto Gordiano<sup>1</sup>, Maeda Batista dos Anjos<sup>1</sup>, Whendel Mesquita do Nascimento<sup>2</sup>, Roberto Alexandre Alves Barbosa Filho<sup>3\*</sup>

---

#### RESUMO

Desde março de 2020 até outubro de 2021, o Amazonas registrou 427.523 casos de COVID-19, presenciando duas ondas epidêmicas (em março de 2020 e janeiro de 2021). Este artigo relata uma análise de dados genômicos e sociais de SARS-CoV-2 provenientes do Amazonas a partir de um banco de dados público. Observou-se que, apesar do número de amostras sequenciadas corresponderem somente a 0,46% do total de casos de COVID-19 identificados no estado durante o mesmo período, os dados permitem concluir que as variantes virais possuem relação significativa com a idade e o ano de coleta e que as regiões genômicas possuem variabilidade genética mínima de 7%. Além disso, os dados sugerem que os genes de SARS-CoV-2 estão sofrendo seleção direcional e apresentando gargalos populacionais, causados possivelmente pela trajetória do vírus durante a pandemia. Desse modo, ressaltamos a importância da vigilância genômica para o combate à pandemia de COVID-19 no Amazonas como ferramenta eficaz no monitoramento das variantes circulantes e de eventos que possam contribuir para o surgimento de novas variantes potencialmente mais infectantes.

**Palavras-chave:** Vigilância genômica; COVID-19; SARS-CoV-2; Amazonas.

---

#### ABSTRACT

From March 2020 to October 2021, Amazonas registered 427,523 cases of COVID-19, witnessing two epidemic waves (in March 2020 and January 2021). This article reports an analysis of genomic and social data on SARS-CoV-2 from Amazonas from a public database. It was observed that, despite the number of sequenced samples corresponding to only 0.46% of the total number of COVID-19 cases identified in the state during the same period, the data allow us to conclude that viral variants have a significant relationship with age and year of collection and that the genomic regions have a minimum genetic variability of 7%. Furthermore, the data suggest that the genes of SARS-CoV-2 are undergoing directional selection and presenting population bottlenecks, possibly caused by the trajectory of the virus during the pandemic. Thus, we emphasize the importance of genomic surveillance to combat the COVID-19 pandemic in Amazonas as an effective tool in monitoring circulating variants and events that may contribute to the emergence of new, potentially more infective variants

**Keywords:** Genomic surveillance; COVID-19; SARS-CoV-2; Amazon.

---

<sup>1</sup> Colégio Militar de Manaus.

<sup>2</sup> Universidade Federal do Amazonas.

<sup>3</sup> Universidade do Estado do Amazonas.

\*E-mail: robertoaabfilho@gmail.com

---

## INTRODUÇÃO

Os coronavírus constituem um grupo diversificado de vírus de RNA que infectam muitos animais diferentes e podem causar infecções respiratórias em humanos. Em 2002 e 2012, respectivamente, dois coronavírus altamente patogênicos de origem zoonótica, coronavírus com Síndrome Respiratória Aguda Grave (SARS-CoV) e coronavírus com síndrome respiratória do Oriente Médio (MERS-CoV), surgiram em humanos e causaram doença respiratória fatal, tornando os coronavírus emergentes um novo problema de saúde pública no século XXI. No final de 2019, um novo coronavírus designado como SARS-CoV-2 surgiu na cidade de Wuhan, China, e causou um surto de pneumonia viral incomum. Por ser altamente transmissível, esta nova doença, também conhecida como COVID-19, se espalhou rapidamente por todo o mundo (HU *et al.*, 2021).

Até o final de outubro de 2021, foram confirmados 246.407.251 casos de COVID-19 no mundo. Estados Unidos foi o país com o maior número de casos acumulados de infecção (45.953.186), seguido pela Índia (34.273.300), Brasil (21.804.094), Reino Unido (9.062.710) e Rússia (8.338.053). Em relação aos óbitos, foram confirmados 4.995.464 no mundo até o dia 30 de outubro de 2021. Estados Unidos apresentou o maior número acumulado de óbitos (745.668), seguido do Brasil (607.694), Índia (458.186), México (288.276) e Rússia (233.063) (Ministério da Saúde, 2021).

De acordo com dados da Fundação de Vigilância em Saúde (FVS/AM), o Amazonas ocupa a 20ª posição no ranking de estados com a maior taxa de incidência, com 5.465 casos por 100.000 habitantes, e a 12ª entre os estados com maior mortalidade pela doença, com uma taxa de 204 óbitos por 100.000 habitantes. Desde março de 2020 até outubro de 2021, o Amazonas registrou 427.523 casos de COVID-19, presenciando duas ondas epidêmicas (em março de 2020 e janeiro de 2021), a primeira provocada principalmente pelas variantes B.1.1.28 e B.1.195 e a segunda pela P.1 (FVS/AM, 2021; NAVECA *et al.*, 2021).

O SARS-CoV-2 sofre frequentes mutações, que ocorrem durante sua replicação devido à microevolução e pressões seletivas e que podem levar ao surgimento de novas variantes, que atualmente podem ser classificadas como VOI (*variant of interest*) ou VOC (*variant of concern*). As VOI apresentam mutações que podem aumentar a transmissão

comunitária, já as VOC apresentam mutações associadas ao aumento da transmissibilidade, alteração prejudicial na epidemiologia da COVID-19, aumento na virulência/mudança clínica da doença ou a diminuição da eficácia das medidas sociais e de saúde pública, diagnósticos, vacinas e terapias disponíveis (RAMBAUT *et al.*, 2020; Secretaria de Estado de Saúde, 2021).

Sabe-se que quanto maior o nível de transmissão, maior será a probabilidade de ocorrência de mutações no vírus, visto que mutação se refere a qualquer mudança no genoma viral. Como as variações genômicas do SARS-CoV-2 podem influenciar diretamente as características epidemiológicas da COVID-19, o acompanhamento das mudanças na sequência genética do vírus é fundamental para rastrear suas mutações, entender a evolução da epidemia e discutir as possibilidades de contenção da infecção. Este monitoramento recebe o nome de vigilância genômica (VG), que possibilita acompanhar as linhagens e mutações genéticas do novo coronavírus e contribuir para o desenvolvimento de vacinas eficazes e diagnósticos mais precisos (HU *et al.*, 2021; Secretaria de Estado de Saúde, 2021; TOYOSHIMA *et. al.*, 2020).

Considerando os cinco países com mais casos de infecção, no período entre março de 2020 e setembro de 2021, uma pesquisa prévia no banco de dados *Global Initiative for Sharing All Influenza Data* (GISAID, disponível em <https://www.gisaid.org>) revelou que o Brasil foi responsável por aproximadamente 1,69% das 3.352.281 amostras sequenciadas no mundo, com cerca de 56.523 sequenciamentos do genoma do SARS-CoV-2 realizados (genomas completos e com boa qualidade). Já Estados Unidos, Índia, México e Rússia foram responsáveis por 28,8% (965.329 sequências), 1,07% (35.922 sequências), 1% (33.444 sequências) e 0,14% (4.640 sequências), respectivamente, dos sequenciamentos realizados no mundo para o referido período, totalizando somente cerca de 32,7% dos 3.352.281 sequenciamentos registrados. Dessa forma, é possível observar que os países com mais casos de COVID-19 não foram os países que mais sequenciaram o genoma do SARS-CoV-2, o que pode comprometer a eficiência da vigilância genômica desse vírus, especialmente nesses países.

A primeira sequência do genoma do novo coronavírus foi publicada no site *Virological* em 10 de janeiro de 2020. A partir daí, sequências de genoma quase completas determinadas por diferentes institutos de pesquisa vêm sendo então divulgadas através de bancos de dados como o GISAID, o *National Center for Biotechnology Information* (NCBI, disponível em [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), e o *Virus Pathogen Resource*

(ViPR, disponível em [www.viprb.rc.org](http://www.viprb.rc.org)). Esse artigo descreve uma análise de dados sociais e biológicos de registros de amostras de SARS-CoV-2 provenientes do Amazonas, coletados do banco de dados GISAID (HAMED *et al.*, 2021; HU *et al.*, 2021).

## MATERIAL E MÉTODOS

### 1. Coleta dos dados

As sequências dos genomas do SARS-CoV-2 foram obtidas a partir do banco de dados do GISAID, disponível em: <https://www.gisaid.org/>, utilizando-se como parâmetros de busca: 01) sequências provenientes do Amazonas, 02) registradas no período de 01 de março de 2020 a 30 de setembro de 2021, 03) sequências completas (aproximadamente 30.000 nucleotídeos) e 04) com alta qualidade (percentual de nucleotídeos identificados como N < 1%). Além da sequência genômica, também foram coletadas informações sobre os locais e datas de coleta das amostras, variante viral identificada, gênero e idade dos pacientes (ELBE *et al.*, 2017).

### 2. Alinhamento e edição das sequências genômicas

Em seguida, o alinhamento múltiplo das sequências genômicas foi realizado utilizando a ferramenta MAFFT, disponível em <https://mafft.cbrc.jp>. O alinhamento foi editado utilizando-se a ferramenta BioEdit versão 7.2 (HALL, 1999; KATOH *et al.*, 2019).

### 3. Análise dos genomas

Os genomas foram analisados integralmente e de forma fragmentada (a partir das regiões gênicas do vírus: orf3a, orf7b, orf8, orf3b, orf6, N, orf10, E, M, S e orf1ab), utilizando-se os programas BioEdit e MEGA X, de acordo com o percentual de mutações, percentual de sítios conservados e o teste de neutralidade de Tajima (KUMAR *et al.*, 2018; TAJIMA, 1989).

### 4. Análise estatística

Os dados referentes a variantes, idade, gênero, local e ano de coleta foram analisados utilizando-se a plataforma estatística Rstudio 4.0.2. A correlação entre a prevalência das variantes e as variáveis gênero, idade, local e ano foi realizada utilizando-se o teste exato de Fisher (RSTUDIO TEAM, 2021).

## RESULTADOS E DISCUSSÃO

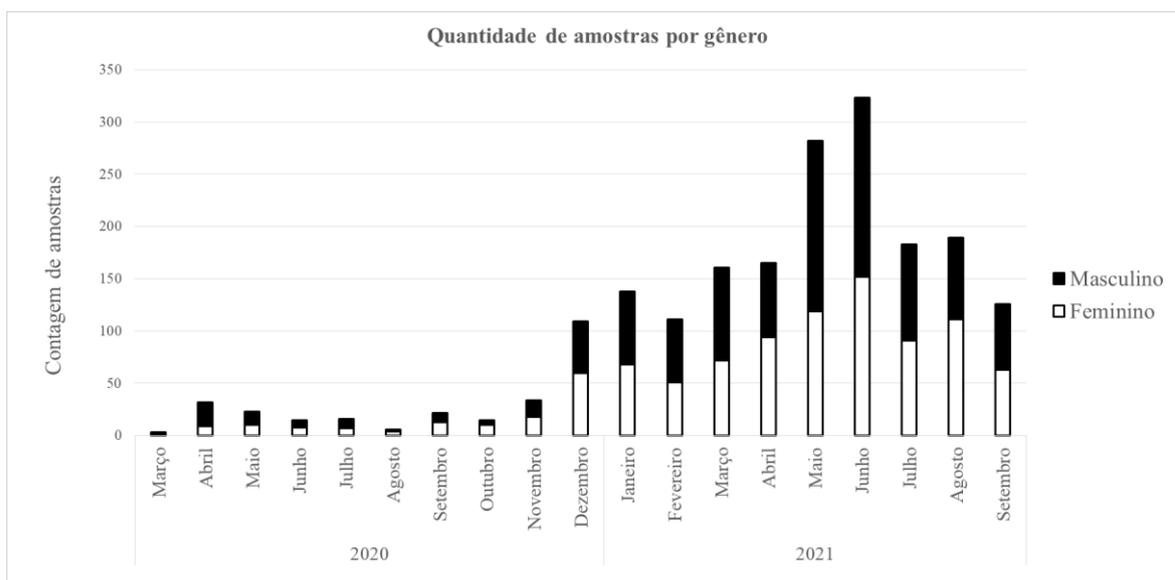
A busca realizada no banco de dados GISAID resultou em 1984 registros. Esses dados foram organizados numa planilha (cujo fragmento está representado pela Figura 1), que foi utilizada como base para as análises posteriores.

**Figura 1** – Fragmento da planilha do banco de dados contendo (da esquerda para a direita) a identificação das amostras no GISAID, a cepa, localização, mês da coleta em número, nome do mês da coleta, ano da coleta, sequência genômica, tamanho do genoma e variante.

Amostra	Cepa	Localização	Mês da coleta	Mês da coleta	Ano da coleta	Sequência genômica	Tamanho do genoma	Variante
EPI_ISL_417034	hCoV-19/Brazil/AM-FIOCRUZ-02/2020	Manaus	3	Março	2020	TCTCTGTAGATCTGTTCTCTA	29789	A.2
EPI_ISL_1068078	hCoV-19/Brazil/AM-FIOCRUZ-20140047PR/2020	Manaus	3	Março	2020	TCTCTGTAGATCTGTTCTCTA	29712	B.1.195
EPI_ISL_811148	hCoV-19/Brazil/AM-FIOCRUZ-20140055FN-R1/2020	Manaus	3	Março	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068079	hCoV-19/Brazil/AM-FIOCRUZ-20140253FP/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068080	hCoV-19/Brazil/AM-FIOCRUZ-20140273IA/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29780	B.1.195
EPI_ISL_1068178	hCoV-19/Brazil/AM-FIOCRUZ-20890128FE/2020	Itacoatiara	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29737	B.1.195
EPI_ISL_1068179	hCoV-19/Brazil/AM-FIOCRUZ-20890130MH/2020	Itacoatiara	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29780	B.1.195
EPI_ISL_1068193	hCoV-19/Brazil/AM-FIOCRUZ-20890610SL/2020	Tabatinga	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068081	hCoV-19/Brazil/AM-FIOCRUZ-20140283VA/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29780	B.1.195
EPI_ISL_1068117	hCoV-19/Brazil/AM-FIOCRUZ-20250329LS/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068166	hCoV-19/Brazil/AM-FIOCRUZ-20890033RD/2020	Maues	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.319
EPI_ISL_1068165	hCoV-19/Brazil/AM-FIOCRUZ-20890024PL/2020	Anori	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068194	hCoV-19/Brazil/AM-FIOCRUZ-20890632AV/2020	Tabatinga	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068210	hCoV-19/Brazil/AM-FIOCRUZ-20891585IS/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29588	B.1.1.28
EPI_ISL_1068162	hCoV-19/Brazil/AM-FIOCRUZ-20890007MR/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068176	hCoV-19/Brazil/AM-FIOCRUZ-20890111DM/2020	Presidente Figueiredo	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068118	hCoV-19/Brazil/AM-FIOCRUZ-20250373JO/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29643	B.1.195
EPI_ISL_1068119	hCoV-19/Brazil/AM-FIOCRUZ-20250374DA/2020	Manaus	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29589	B.1.195
EPI_ISL_1068181	hCoV-19/Brazil/AM-FIOCRUZ-202503167EC/2020	Santo Antônio do	4	Abril	2020	TCTCTGTAGATCTGTTCTCTA	29588	B.1.195

Verificou-se que 50% dos dados analisados eram provenientes de pacientes do gênero masculino (Figura 2), 48,44% feminino e 1,56% não possuíam essa informação. Esses dados corroboram com os dados de Naveca *et al.* (2021), no qual não houve diferença estatisticamente significativa entre os gêneros com relação à infecção pela variante P.1. Notou-se também o aumento significativo do volume de sequenciamentos no ano de 2021, com ápice no mês de junho (323 sequenciamentos). Tal aumento ocorreu em resposta ao incremento da capacidade de geração e análise de dados genômicos da rede de diagnóstico molecular para COVID-19, coordenado pelo Laboratório Central de Saúde Pública do Amazonas (LACEN), devido a aportes financeiros de diferentes órgãos públicos e privados. A rede de laboratórios de referência é composta pelo Instituto Leônidas & Maria Deane (ILMD/Fiocruz Amazônia), além do laboratório da Fundação de Medicina Tropical Heitor Vieira Dourado (FMT-HVD) e da Universidade Federal do Amazonas (UFAM) (NAVECA *et al.*, 2021; Secretaria de Estado de Saúde, 2021).

**Figura 2** – Gráfico que relaciona a quantidade de amostras sequenciadas por gênero a cada mês ao longo dos anos de 2020 e 2021. Na parte inferior das barras, a quantidade de amostras sequenciadas do gênero feminino e na superior, a quantidade de amostras sequenciadas do gênero masculino.



Os dados revelaram 27 variantes para o período analisado (Tabela 1). Entre elas, vale destacar as variantes associadas à primeira onda epidêmica em 2020, B.1.1.28 e B.1.195, além das VOC's AY, B.1.617.2 e P.1 (associada à segunda onda epidêmica em 2021). Porém, os dados da Secretaria do Estado de Saúde (2021) também identificaram a variante VOC B.1.1.7, encontrada em Manaus, originária no Reino Unido. Contudo, como a qualidade da sequência da variante B.1.1.7 não atendia a um dos parâmetros estabelecidos para essa pesquisa (percentual de nucleotídeos identificados como N<1%), seus dados foram excluídos da análise. Essa discrepância no resultado da pesquisa ressalta a necessidade premente de realizar mais sequenciamentos, a fim de produzir sequências genômicas com alta qualidade.

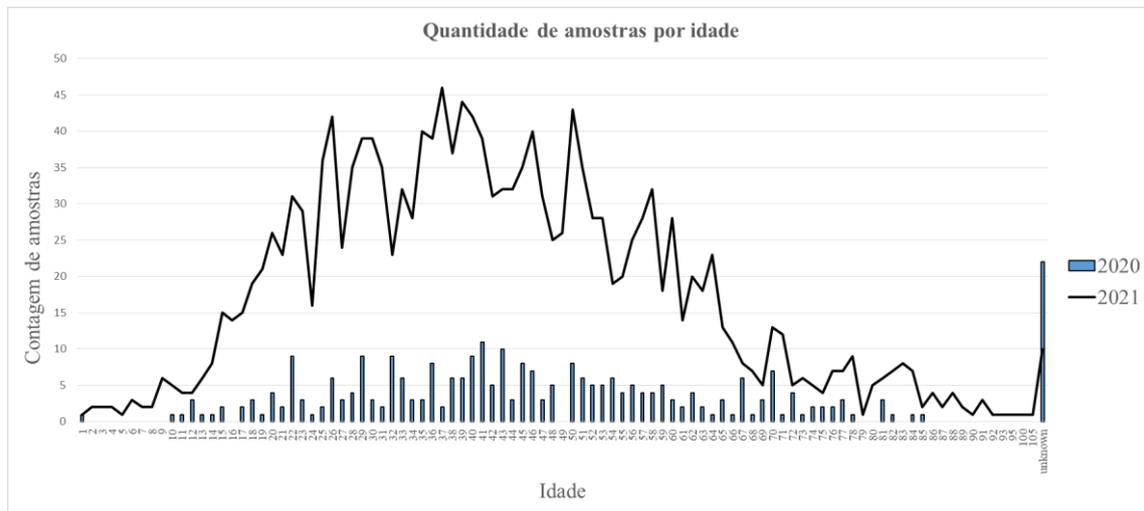
**Tabela 1** – Tabela de referência cruzada que relaciona as variantes identificadas entre as 1984 amostras analisadas nos anos de 2020 e 2021.

Variante	2020	2021	Total
A.2	1	0	1
AY.4	0	14	14
AY.6	0	1	1
AY.9	0	1	1
B.1.1	8	0	8

B.1.1.28	75	0	75
B.1.1.33	26	1	27
B.1.1.37	32	0	32
B.1.111	3	0	3
B.1.195	51	0	51
B.1.212	1	0	1
B.1.319	1	0	1
B.1.617.2	0	51	51
N.1	0	1	1
N.9	4	0	4
P.1	79	690	769
P.1.10	0	1	1
P.1.12	0	14	14
P.1.2	0	23	23
P.1.3	0	31	31
P.1.4	0	539	539
P.1.5	0	10	10
P.1.6	0	298	298
P.1.7	0	4	4
P.1.8	0	2	2
P.1.9	0	1	1
P.2	16	5	21
Total	297	1687	1984

A média de idade em 2020 dos pacientes da amostra foi de  $45 \pm 17,19$  anos, com moda de 41 anos. Já em 2021 essa média caiu para  $42 \pm 17,4$  anos, com moda de 37 anos (Figura 3). De acordo com a FVS/AM (2021), com o avanço da vacinação, houve redução na incidência de casos em idosos (60 anos ou mais), e as pessoas com idade entre 20 e 59 anos vêm apresentando a maior taxa de incidência de COVID-19, com mais de 4 mil casos por 100 mil habitantes. Porém, ao analisar a variação da taxa de incidência de setembro para outubro de 2021, observa-se redução em todas faixas etárias, principalmente em idosos, com 50% e na faixa etária de 20 a 59 anos (46%) (FVS/AM, 2021).

**Figura 3** – Gráfico que relaciona a quantidade de amostras sequenciadas por idade durante os anos de 2020 e 2021. As barras representam a quantidade de amostras sequenciadas no ano de 2020 e, a linha, a quantidade de amostras sequenciadas no ano de 2021.



Para realizar o teste exato de Fisher as variáveis analisadas (variante viral, idade, gênero, local e ano) foram dicotomizadas: grupos "P.1" e "não P.1" para as variantes; grupos "idoso" e "não idoso" para a idade; grupos "masculino" e "feminino" para o gênero; grupos "Manaus" e "não Manaus" para o local de coleta; grupos "2020" e "2021" para o ano. Após a organização dos dados, o teste demonstrou que as variantes identificadas possuem relação significativa com a idade e o ano de coleta ( $p < 0,05$ ), mas não com o gênero dos pacientes e local da coleta. Esses resultados confirmam o que está ilustrado nas figuras 2 e 3.

As sequências genômicas foram analisadas de forma global e fragmentada entre as regiões gênicas (Tabela 2), revelando variabilidade mínima de 7% dentro dos genes. O genoma de SARS-CoV-2 compreende cerca de 30.000 nucleotídeos, organizados em genes específicos que codificam proteínas estruturais e proteínas não estruturais (Nsps). As proteínas estruturais incluem proteínas da espícula (S), envelope (E), membrana (M) e nucleocapsídeo (N). Os Nsps, que são gerados como produtos de clivagem das poliproteínas virais do quadro de leitura aberta 1ab (ORF1ab), se reúnem para facilitar a replicação e a transcrição viral. Além disso, proteínas acessórias são codificadas pelos genes orf3a, orf3b, orf6, orf7a orf8 e orf10. O tamanho médio dos genomas sequenciados em 2020 foi de  $29.698 \pm 90,1$  nucleotídeos. Já em 2021, houve aumento no tamanho médio

dos genomas sequenciados ( $29.777 \pm 59,79$  nucleotídeos), indicando aumento na precisão dos sequenciamentos realizados. Apesar do total de amostras sequenciadas (1.984) representar uma parcela estatisticamente significativa dos 427.523 casos registrados no Amazonas, com margem de erro de 2,2%, ela corresponde somente a 0,46% do total de casos registrados de infecção, negligenciando, dessa maneira, 99,54% dos casos registrados para o estado no mesmo período (TOYOSHIMA *et. al.*, 2020).

**Tabela 2** – Tabela demonstrando os valores associados ao teste de neutralidade de Tajima para as sequências genômicas inteiras e regiões gênicas separadas.

Região	Tamanho (n)	Sítios conservados	Sítios variáveis (S)	% Variação (S/n)	Diversidade nucleotídica ( $\pi$ )	Estatística de Tajima (D)
Genoma	30038	27387	2532	8,43	0,000661	-2,68
orf3a	831	695	133	16,00	0,000909	-2,6
orf7b	133	113	19	14,28	0,001458	-2,02
orf8	368	314	52	14,13	0,001462	-2,34
orf3b	456	397	59	12,94	0,001006	-2,43
orf6	186	163	23	12,36	0,000467	-2,21
N	1260	1109	151	11,98	0,001563	-2,46
orf10	117	104	13	11,11	0,001472	-1,8
E	228	206	22	9,65	0,000316	-2,2
M	669	613	56	8,37	0,000474	-2,46
S	3834	3517	317	8,27	0,001213	-2
orf1ab	21290	19754	1536	7,21	0,000479	-2,7

Nessa análise, a orf3a apresentou o maior percentual de variação (16,0%) e a orf1ab o menor (7,21%). Como descrito anteriormente, os genes de SARS-CoV-2 vêm sofrendo seleção direcional, o que pode repercutir nos percentuais de variação deles, tornando-se mais evidente nos genes que codificam proteínas estruturais. Já os genes das proteínas acessórias tendem a apresentar percentuais maiores. O gene N foi exceção a esse fato, apresentando 11,98% de variação, sendo maior que a da orf10, com 11,11%. Além disso, o gene N também apresentou a maior diversidade nucleotídica (0,001563) e o gene E a menor (0,000316). Quanto menor for o valor da diversidade nucleotídica, maior é a probabilidade de existência de gargalos populacionais ou de eventos fundadores recentes, o que pode ser verificado em praticamente todos os genes em decorrência da

trajetória do SARS-CoV-2 durante a pandemia. Os valores negativos do teste de Tajima para todos os genes sugerem que todos vêm sofrendo seleção direcional, favorecendo o aumento da frequência dos alelos mais comuns dos genes analisados, e que os vírus estão em processo de expansão populacional (LIU *et al.*, 2020).

## CONCLUSÃO

Com base nas análises realizadas, podemos concluir que houve mudança na prevalência das variantes de SARS-CoV-2 na população amazonense entre os anos de 2020 e 2021, que mostraram ter relação também com as idades dos pacientes. As análises genômicas revelaram que a baixa cobertura de sequenciamento pode ocultar potenciais novas variantes, considerando a variabilidade genética e as pressões de seleção observadas, reforçando a importância da vigilância genômica no estado do Amazonas para o entendimento da dinâmica de transmissão do SARS-CoV-2 e o desenvolvimento de estratégias de combate à pandemia no estado.

## REFERÊNCIAS

AMAZONAS (Estado). Fundação de Vigilância em Saúde (FVS/AM). **Boletim Nº 28 sobre a situação epidemiológica da COVID-19 no estado do Amazonas**. Amazonas, 26 de outubro de 2021. 10 p.

AMAZONAS (Estado). Secretaria de Estado de Saúde, Fundação de Vigilância em Saúde do Amazonas, Instituto Leônidas e Maria Deane (FIOCRUZ AMAZÔNIA). **Relatório sobre a Vigilância Genômica do SARS-CoV-2 no Amazonas**. Amazonas, 14 de setembro de 2021. 12 p.

BRASIL. Ministério da Saúde - Secretaria de Vigilância em Saúde. **Boletim Epidemiológico Especial – COVID-19 Nº 87**. Brasil, 30 de outubro de 2021. 109 p.

ELBE, S.; BUCKLAND-MERRETT, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. **Global Challenges**, v. 1, p. 33-46, 2017.

HALL, T.A. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, v. 41, p. 95-98, 1999.

HAMED, S. M.; ELKHATIB, W. F.; KHAIRALLA, A. S.; NOREDDIN, A. M. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. **Scientific Reports**, v. 11, n. 1, p. 1-8, 2021.

HU, B.; GUO, H.; ZHOU, P.; SHI, Z. Characteristics of SARS- CoV-2 and COVID-19. **Nature Reviews - Microbiology**, v. 19, p. 141-154, 2021.

KATOH, K.; ROZEWICKI, J.; Yamada, K. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1160-1166, 2019.

KUMAR, S.; STECHER, G.; LI, M.; KNYAZ, C.; TAMURA, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547-1549, 2018.

LIU, Q.; ZHAO, S.; SHI, C-M.; SONG, S.; ZHU, S.; SU, Y.; ZHAO, W.; LI, M.; BAO, Y.; XUE, Y.; CHEN, H. Population Genetics of SARS-CoV-2: Disentangling Effects of Sampling Bias and Infection Clusters. **Genomics, Proteomics & Bioinformatics**, v. 18, n. 6, p. 640-647, 2020.

NAVECA, F.G.; NASCIMENTO, V.; DE SOUZA, V.C. *et al.* COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. **Nature Medicine**, v. 27, p. 1230–1238, 2021.

RAMBAUT, A.; HOLMES, E. C.; O'TOOLE, Á.; HILL, V.; MCCRONE, J. T.; RUIS, C.; DU PLESSIS, L.; PYBUS, O. G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. **Nature Microbiology**, v. 5, p. 1403–1407, 2020.

RSTUDIO TEAM. RStudio: Integrated Development Environment for R. Boston, MA. Disponível em <http://www.rstudio.com>. Acesso em 30 de setembro de 2021.

TAJIMA, F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. **Genetics**, v. 123, n. 3, p. 585-95, 1989.

TOYOSHIMA, Y.; NEMOTO, K.; MATSUMOTO, S.; NAKAMURA, Y.; KIYOTANI, K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. **Journal of Human Genetics**, v. 65, p. 1075-1082, 2020.

*Recebido em: 01/11/2021*

*Aprovado em: 30/11/2021*

*Publicado em: 03/12/2021*