

Aluna: Emily dos Santos Silva e Gregório Kappaun Rocha (orientador)
Instituto Federal de Educação, Ciência e Tecnologia Fluminense – IFFluminense
Macaé - RJ
santos.emily@gsuite.iff.edu.br; gregorio.rocha@iff.edu.br

INTRODUÇÃO

Os coronavírus (CoV) pertencem a uma família de vírus chamada Coronaviridae (Decaro et al., 2020). Recentemente, em 2019, foi identificada a COVID-19, uma doença infecciosa causada por um novo CoV, chamado de SARS-CoV-2.

Neste trabalho, buscou-se aplicar ferramentas computacionais de bioinformática para estudar sequências genômicas e proteicas de CoVs depositadas em banco de dados internacionais públicos, objetivando contribuir para o avanço nos estudos sobre os CoV e doenças associadas.

Ademais, buscou-se identificar mutações em sequências depositadas, estudar a distribuição global de algumas mutações, bem como construir modelos tridimensionais de proteínas virais mutadas.

Um programa de computador que tem como objetivo auxiliar na análise e na manipulação de sequências biológicas também é proposto neste trabalho.

MATERIAIS E MÉTODOS

O banco de dados GISAID foi utilizado como fonte de sequências genômicas de SARS-CoV-2. A partir da ferramenta EpiCoV, buscou-se identificar a distribuição global das variantes, com foco na América do Sul e, principalmente, no Brasil, até a data 15/09/2021.

Para o estudo, foram selecionadas como referência as proteínas do SARS-CoV-2 pertencentes no vírus hCoV-19/Wuhan/WIV04/2019, coletado no início da pandemia na cidade Wuhan, na China.

As sequências virais coletadas foram alinhadas com a sequência do genoma de referência, para analisar diferenças significativas entre os resíduos de ambas, por meio do software CoVsurver.

Devido ao grande volume de dados disponível, foi desenvolvido um algoritmo, em linguagem C, para auxiliar nas análises e padronização das informações coletadas nas bases de dados.

A pesquisa bibliográfica foi realizada por meio de consultas a diversos artigos publicados em revistas científicas e periódicos online.

RESULTADOS E DISCUSSÃO

A cepa B.1.1.7, também denominada Alfa, foi detectada em mais de 30 países e há registros de 2.546 sequências na América do Sul, dessas ocorrências, 624 foram no Brasil. Outra variante notável é a linhagem B.1.351, responsável por um aumento na afinidade de ligação aos receptores ACE2.

De um total de 22 sequências, 6 foram submetidas no Brasil.

A variante de preocupação B.1.617.2, denominada como variante Delta, foi inicialmente identificada na Índia. De um total de 5.966 sequências na América do Sul, 3.525 delas foram identificadas no Brasil.

A América do Sul possui um total de 31.689 sequências pertencentes a variante P.1, ou Gama, sendo 23.605 no Brasil, um dos países que mais submeteram sequências pertencentes a essa linhagem.

Tabela 1. Variantes de SARS-CoV-2 identificadas no estudo e suas principais mutações características.

| DESIGNAÇÃO DE VARIANTE | PROTEÍNA:MUTAÇÃO | ORIGEM |
|-------------------------------------|---|----------------|
| B.1.1.7 (20I / 501Y.V1) | S: N501Y, D614G | Reino Unido |
| B.1.1.28.1 / P.1 (20J / 501Y.V3) | S: L18F, E484K, N501Y, D614, H655Y, T1027I, V1176F N: P80R ORF3a: S253P | Brasil e Japão |
| B.1.351 (20H / 501Y.V2) | S: E484K, N501Y, D614G | África do Sul |
| B.1.617.2 | S: D614G, P681R | Índia |

Fonte: Centro de Controle e Prevenção de Doenças (CDC).

Legenda: S = Spike; N = Nucleocapsídeo; ORF3a = Quadro de leitura 3a.

A mutação D614G é dominante em cepas circulantes em todo o mundo, presente em 98.32% das amostras da proteína Spike, em 192 países. Está relacionada ao aumento da replicação viral em células do trato respiratório superior, aumentando a infectividade dos vírus (Plante et al., 2021).

A mudança de aminoácidos V1176F na proteína S já se espalhou por 107 países desde a sua primeira ocorrência, em março de 2020, no Brasil. Foram contabilizadas 37.572 sequências com essa substituição na América do Sul. O Brasil conta com 28.929 ocorrências, sendo o estado de São Paulo a primeira região com maior incidência global e Rio de Janeiro, a sétima.

Tabela 2. Mutações de resíduos do vírus SARS-CoV- 2 presentes na América do Sul filtradas a partir de entradas do GISAID.

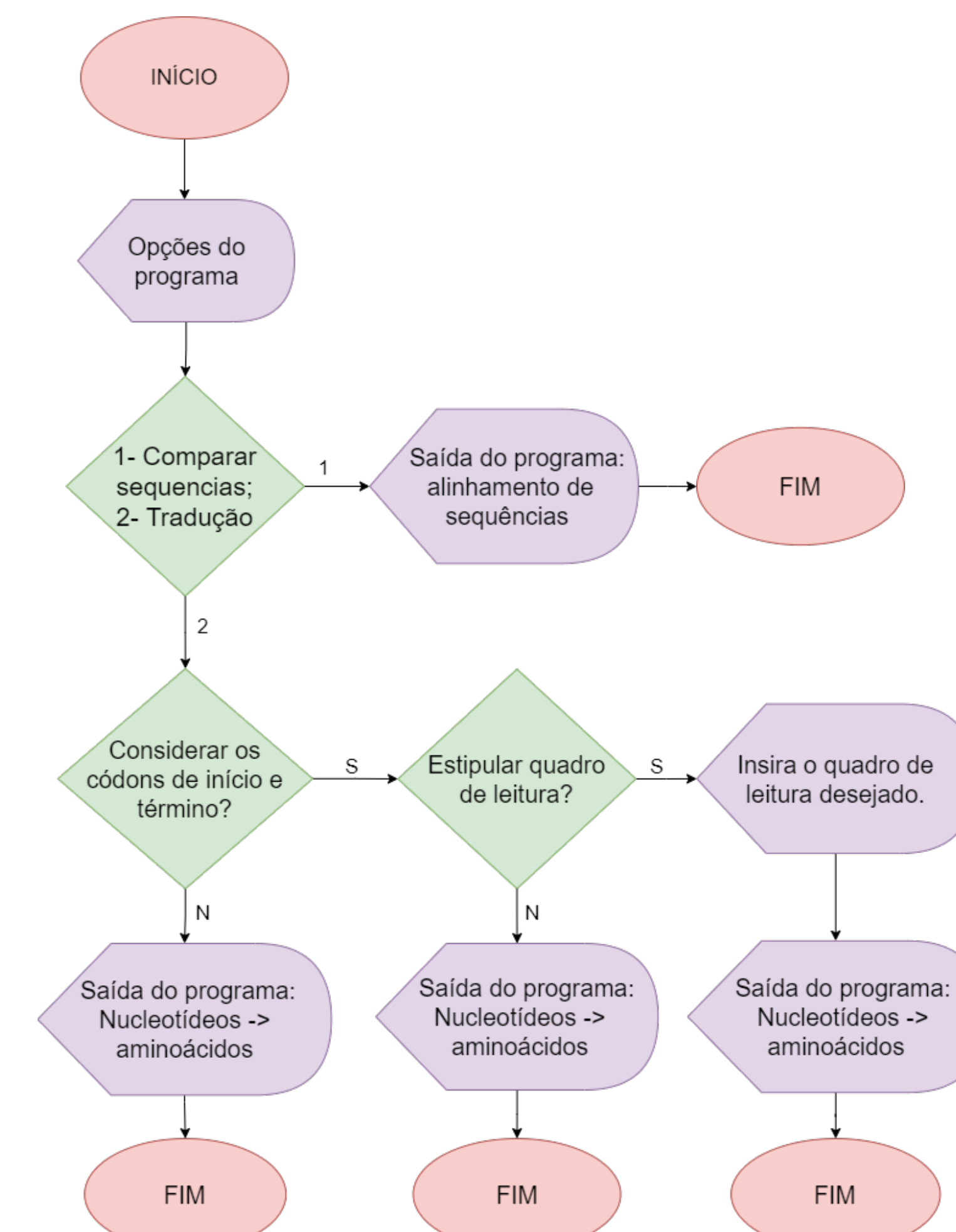
| MUTAÇÃO (PROTEÍNA: MUTAÇÃO) | ENTRADAS NA AMÉRICA DO SUL | ENTRADAS NO BRASIL |
|-----------------------------|----------------------------|--------------------|
| S: L18F | 31.052 | 23.489 |
| S: E484K | 37.373 | 25.675 |
| S: N501Y | 36.120 | 23.393 |
| S: D614G | 68.644 | 36.639 |
| S: H655Y | 32.882 | 23.689 |
| S: T1027I | 31.453 | 23.625 |
| S: V1176F | 37.571 | 28.929 |
| S: P681R | 6.141 | 3.660 |
| ORF3a: S253P | 31.984 | 23.782 |
| NSP3: K977Q | 31.700 | 23.665 |
| N: P80R | 31.781 | 23.572 |

Fonte: GISAID. Data de coleta de dados: 15/09/2021.

Legenda: S = Spike; N = Nucleocapsídeo; ORF3a = Quadro de leitura 3a.

Os repositórios contam com um grande volume de dados disponibilizados em diferentes formatos, o que pode impossibilitar a análise dessas informações de maneira manual. Para o presente estudo, foi desenvolvido um programa, em linguagem C, para realizar a conversão de códons para seus respectivos aminoácidos, segundo o código genético. Posteriormente, foi implementada uma opção para comparação de sequências, em relação à sequência de do estudo.

Figura 1. Fluxograma do algoritmo desenvolvido para a análise de sequências.



Legenda: S = sim; N = não.

CONSIDERAÇÕES FINAIS

Este trabalho contribuiu para identificar algumas mutações de relevância no cenário mundial da pandemia de COVID-19 e identificar a sua presença em países da América do Sul e no Brasil.

Também se iniciou o desenvolvimento de uma ferramenta de bioinformática para fornecer suporte à análise de dados de sequências biológicas. Para o futuro, espera-se incrementar o algoritmo com uma análise de mutações mais completa.

Como perspectiva futura, serão conduzidas análises de comparação de estruturas proteicas geradas a partir de sequências mutadas com o auxílio do programa Swiss-Model.

AGRADECIMENTOS

Agradecemos o apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ, por meio da concessão de bolsa de pré-iniciação científica no Programa Jovens Talentos. Além do Instituto Federal Fluminense – Campus Macaé, por possibilitar a realização dessa pesquisa.

REFERÊNCIAS

- Decaro, N., & Lorusso, A. (2020). Novel human coronavirus (SARS-CoV-2): A lesson from animal coronaviruses. *Veterinary microbiology*, 244, 108693.
- Elbe, S., and Buckland-Merrett, G. (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*, 1:33-46. DOI:10.1002/gch2.1018 PMCID: 31565258
- Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharton, D., Billelo, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., Plante, K. S., ... Shi, P. Y. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 592(7852), 116–121.