Evaluation of SARS-CoV-2 genomic architecture and its alteration pattern through ORF analysis

Avaliação da arquitetura genômica do SARS-CoV-2 e seu padrão de alteração por meio de análise de ORF

DOI:10.34119/bjhrv4n3-236

Recebimento dos originais: 05/05/2021
Aceitação para publicação: 10/06/2021

Yago Queiroz dos Santos
PhD
Limoeiro do Norte Campus, Federal Institute of Education, Science and Technology of Ceará
Institute of Tropical Medicine, Federal University of Rio Grande do Norte, Natal, Rio G rande do Norte, Brazil
E-mail: microbiologia.imt@gmail.com

Gabriella Silva Campos Carelli
BSc
Institute of Tropical Medicine, Federal University of Rio Grande do Norte, Natal, Rio G rande do Norte, Brazil
Department of Biochemistry, Federal University of Rio Grande do Norte, Natal, Rio Gr ande do Norte, Brazil
E-mail: microbiologia.imt@gmail.com

Bruno Oliveira de Veras
MSc
Department of Biochemistry, Federal University of Pernambuco, Recife, Pernambuco, Brazil
E-mail: microbiologia.imt@gmail.com

Anderson Felipe Jácome de França
PhD
Institute of Tropical Medicine, Federal University of Rio Grande do Norte, Natal, Rio G rande do Norte, Brazil
Department of Biochemistry, Federal University of Rio Grande do Norte, Natal, Rio Gr ande do Norte, Brazil
E-mail: microbiologia.imt@gmail.com

Elizeu Antunes dos Santos
PhD
Institute of Tropical Medicine, Federal University of Rio Grande do Norte, Natal, Rio G rande do Norte, Brazil
Department of Biochemistry, Federal University of Rio Grande do Norte, Natal, Rio Gr ande do Norte, Brazil
E-mail: microbiologia.imt@gmail.com

ABSTRACT

Members of the Coronaviridae family comprise four genera *Alphacoronavirus* and *Betacoronavirus*, which infect only mammals, and *Gammacoronavirus* and *Deltacoronavirus* infect birds and mammals. Since the end of 2019, humanity has witnessed the emergence of a new pandemic caused by a line of beta-coronavirus (SARS-CoV-2) responsible for causing a novel severe acute respiratory syndrome named coronavirus disease (COVID-19) affecting countries worldwide. In this context, this work aimed to investigate the main changes in SARS-CoV-2 genomic architecture through the time since the beginning of the infection using *in silico* analysis. A genomic database was built using complete and revised genomes from NCBI as well as analyzed through sequencing alignment and phylogenetic softwares. This study was able to show a change in the organizational pattern of the genome of the new coronavirus for the different regions studied, in addition to showing specific changes in the genomic sequence requiring further analysis of the collected genomes and may provide new evidence for the key protein changes like Spike protein.

Keywords: SARS-CoV-2, Genome, COVID-19.

RESUMO

Os membros da família Coronaviridae compreendem quatro gêneros Alfacoronavirus e Betacoronavirus, que infectam apenas mamíferos, e Gammacoronavirus e Deltacoronavirus infectam aves e mamíferos. Desde o final de 2019, a humanidade testemunhou o surgimento de uma nova pandemia causada por uma linha de beta-coronavírus (SARS-CoV-2) responsável por causar uma nova síndrome respiratória aguda chamada doença coronavírus (COVID-19) que afeta países em todo o mundo. Neste contexto, este trabalho teve como objetivo investigar as principais mudanças na arquitetura genômica do SRA-CoV-2 ao longo do tempo desde o início da infecção, utilizando em análise silico. Um banco de dados genômico foi construído usando genomas completos e revisados do NCBI, bem como analisado através de alinhamento de seqüenciamento e softwares filogenéticos. Este estudo foi capaz de mostrar uma mudança no padrão organizacional do genoma do novo coronavírus para as diferentes regiões estudadas, além de mostrar mudanças específicas na seqüência genômica que requerem uma análise mais aprofundada dos genomas coletados e pode fornecer novas evidências para as principais mudanças protéicas como a proteína Spike.

Palavras-chave: SARS-CoV-2, Genoma, COVID-19

1 INTRODUCTION

Since the end of 2019, humanity has witnessed the emergence of a new pandemic caused by a line of beta-coronavirus (SARS-CoV-2) responsible for causing a novel severe acute respiratory syndrome named coronavirus disease (COVID-19) affecting countries worldwide (Amsalem et al., 2021; Wu et al., 2020).

Members of the Coronaviridae family comprise four genera *Alphacoronavirus* and *Betacoronavirus*, which infect only mammals, and *Gammacoronavirus* and

Deltacoronavirus infect birds and mammals. They are enveloped, with positive single-stranded RNA between 26 and 32 kilobases. Due to these characteristics, they can present a high mutational rate, which can promote the ability to detect new cell types or even new species that can trigger serious lung diseases (Lvov & Alkhovsky, 2020).

Structurally coronaviruses mainly consist up to four group of proteins that form the nucleocapsid (N), the envelope (E), the membrane (M) and Spikes (S). The Spike protein forms elongated structures that protrude throughout the virion, forming a crown similar to the sun's rays, hence the name. These protrusions bind to receptors on host cells, thereby determining the types of cells and the variety of species that the virus can infect (Wrapp et al., 2020).

Initially coronavirus disease was presented as enzootic infections, restricted to their natural animal hosts, however, mutations allowed the Coronaviruses (CoVs) to establish themselves as a zoonotic disease in humans. As a consequence, outbreaks of severe acute coronavirus syndrome (SARS-CoV) in 2003 led to an almost pandemic with 8096 cases and 774 deaths reported worldwide, resulting in a 9.6% death rate. Since the outbreak of Middle Eastern respiratory syndrome by Coronavirus (MERS-CoV) in April 2012 to 2018, laboratory confirmed cases have been reported worldwide, including 791 deaths associated with a 35.5% death rate (Ahmadzadeh et al., 2020).

In this context, this work aimed to investigate the main changes in SARS-CoV-2 genomic architecture through the time since the beginning of the infection using *in silico* analysis.

2 MATERIALS AND METHODS

2.1 GENOME DATABASE

For the genome database assembly, the genomic information deposited and publicly disclosed at the National Center for Biotechnology Information (NCBI) was accessed for revised-only and complete genomes. To allow a global analysis of the genomic data, a selection of genomes from different regions (Brazil, Europe, USA and China) was chosen, each with 30 complete genomes and initial alignment was performed using *ClustalW*. For phylogenetic analysis, a database was constructed and the outgroup was represented by Bat SARS-CoV Rs672/2006 complete genomes. All genomes were checked for integrity and deposited in the database in FASTA format (Kryukov et al., 2020; Yang et al., 2020).

2.2 OPEN READING FRAME ANALYSIS

The analysis of open reading frames was performed based on the genomic sequences deposited in the database from different regions of the world as well as root groups containing the Bat SARS-CoV Rs672/2006 genome. The analysis was made using the online software Open Reading Frame Finder hosted on NCBI website. The search parameters were optimized for Minimal ORF length bigger than 150 nucleotides with a single initiation codon "ATG" for each ORF and standard genetic code validation (Rangwala et al., 2021).

2.3 PHYLOGENETIC ANALYSIS

The Maximum-likelihood (ML) tree was constructed under the appropriate nucleotide substitution model using the software MEGA. Using the same program, robustness of the tree was evaluated by the bootstrapping with 1000 replicates. Bat SARS-CoV Rs672/2006 genome was used to determine root group (Tamura et al., 2007).

3 RESULTS

3.1 SARS-CoV GENOME DATABASE

Considering all analyzed genomes worldwide, samples from China and United States constituted the highest percentage of deposited genomes in the NCBI. Each analyzed genome had, on average, a sequence of 30 Kbp, constituting a total genomic database of 3.6 Mbp (data not shown), considerably higher than the Human Genome itself of approximately 3 Mbp (Sawicki et al., 1993).

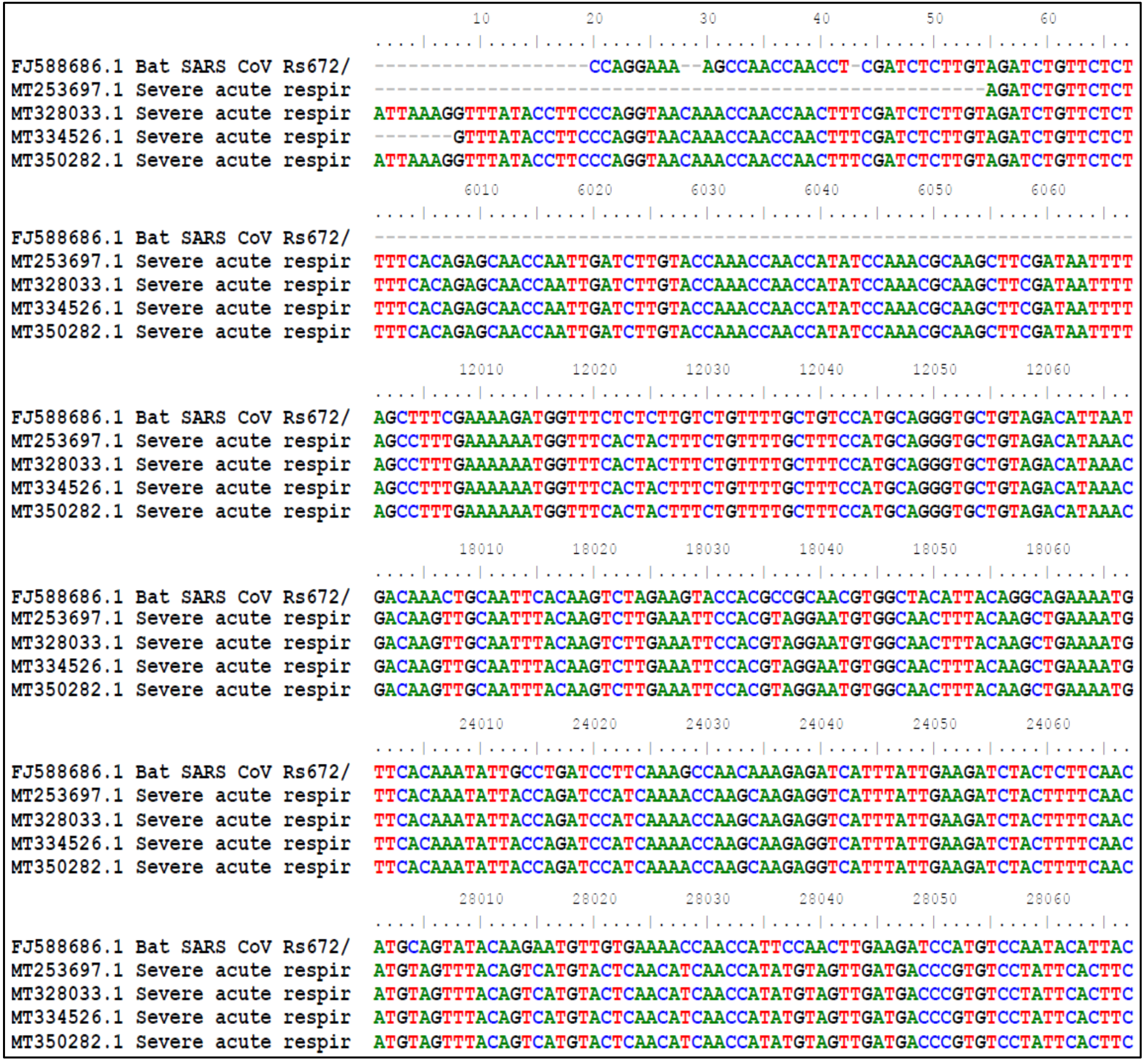
Table 1: Representative Bat SARS-CoV and SARS-CoV-2 genome sequences from four different regions worldwide.

NCBI Code	Region	Nucleotide Length	Organism
MT253697.1	Wuhan, China (2020)	29781 nt	SARS-CoV-2
MT328033.1	Greece, Europe (2020)	29875 nt	SARS-CoV-2
MT334526.1	Utah, United States (2020)	29883 nt	SARS-CoV-2
MT350282.1	São Paulo, Brazil (2020)	29883 nt	SARS-CoV-2
FJ588686.1	Southern China (2010)	29059 nt	SARS-CoV

nt: nucleotides

The smallest amount of complete genome samples was recorded for Brazil, which reflects a considerable underreporting of circulating variants. A preliminary alignment of representative sequences for each studied region (table 1) demonstrates a high degree of conservation of collected genomes (figure 1).

Figure 1: Overall genome alignment of representative SARS-CoV-2 sequences from four different regions worldwide compared with Bat SARS-CoV genome.

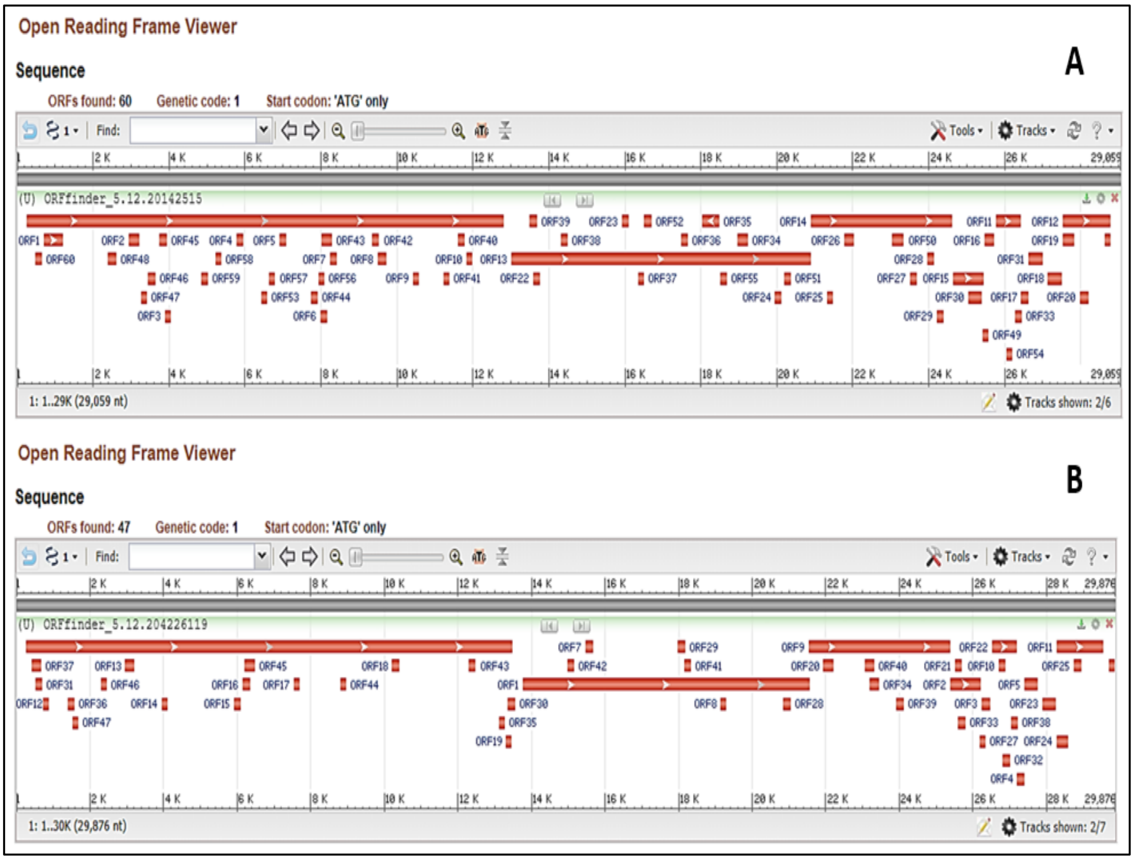


```
                                        10        20        30        40        50        60
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|..
FJ588686.1 Bat SARS CoV Rs672/  -------------------CCAGGAAA--AGCCAACCAACCT-CGATCTCTTGTAGATCTGTTCTCT
MT253697.1 Severe acute respir  -------------------------------------------------------AGATCTGTTCTCT
MT328033.1 Severe acute respir  ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCT
MT334526.1 Severe acute respir  -------GTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCT
MT350282.1 Severe acute respir  ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCT

                                      6010      6020      6030      6040      6050      6060
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|..
FJ588686.1 Bat SARS CoV Rs672/  -------------------------------------------------------------------
MT253697.1 Severe acute respir  TTTCACAGAGCAACCAATTGATCTTGTACCAAACCAACCATATCCAAACGCAAGCTTCGATAATTTT
MT328033.1 Severe acute respir  TTTCACAGAGCAACCAATTGATCTTGTACCAAACCAACCATATCCAAACGCAAGCTTCGATAATTTT
MT334526.1 Severe acute respir  TTTCACAGAGCAACCAATTGATCTTGTACCAAACCAACCATATCCAAACGCAAGCTTCGATAATTTT
MT350282.1 Severe acute respir  TTTCACAGAGCAACCAATTGATCTTGTACCAAACCAACCATATCCAAACGCAAGCTTCGATAATTTT

                                     12010     12020     12030     12040     12050     12060
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|..
FJ588686.1 Bat SARS CoV Rs672/  AGCTTTCGAAAAGATGGTTTCTCTCTTGTCTGTTTTGCTGTCCATGCAGGGTGCTGTAGACATTAAT
MT253697.1 Severe acute respir  AGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGTAGACATAAAC
MT328033.1 Severe acute respir  AGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGTAGACATAAAC
MT334526.1 Severe acute respir  AGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGTAGACATAAAC
MT350282.1 Severe acute respir  AGCCTTTGAAAAAATGGTTTCACTACTTTCTGTTTTGCTTTCCATGCAGGGTGCTGTAGACATAAAC

                                     18010     18020     18030     18040     18050     18060
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|..
FJ588686.1 Bat SARS CoV Rs672/  GACAAACTGCAATTCACAAGTCTAGAAGTACCACGCCGCAACGTGGCTACATTACAGGCAGAAAATG
MT253697.1 Severe acute respir  GACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTACAAGCTGAAAATG
MT328033.1 Severe acute respir  GACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTACAAGCTGAAAATG
MT334526.1 Severe acute respir  GACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTACAAGCTGAAAATG
MT350282.1 Severe acute respir  GACAAGTTGCAATTTACAAGTCTTGAAATTCCACGTAGGAATGTGGCAACTTTACAAGCTGAAAATG

                                     24010     24020     24030     24040     24050     24060
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|..
FJ588686.1 Bat SARS CoV Rs672/  TTCACAAATATTGCCTGATCCTTCAAAGCCAACAAAGAGATCATTTATTGAAGATCTACTCTTCAAC
MT253697.1 Severe acute respir  TTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAAC
MT328033.1 Severe acute respir  TTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAAC
MT334526.1 Severe acute respir  TTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAAC
MT350282.1 Severe acute respir  TTCACAAATATTACCAGATCCATCAAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAAC

                                     28010     28020     28030     28040     28050     28060
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|..
FJ588686.1 Bat SARS CoV Rs672/  ATGCAGTATACAAGAATGTTGTGAAAACCAACCATTCCAACTTGAAGATCCATGTCCAATACATTAC
MT253697.1 Severe acute respir  ATGTAGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTC
MT328033.1 Severe acute respir  ATGTAGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTC
MT334526.1 Severe acute respir  ATGTAGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTC
MT350282.1 Severe acute respir  ATGTAGTTTACAGTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTC
```

3.2 ORF ANALYSIS

Considering the open reading frames analysis, the genome architecture of SARS-CoV-2 differed significantly from the root group. A decrease in the amount of ORFs was observed in the SARS-CoV-2 genomes (figure 2), which indicates a condensation tendency towards new coronavirus open reading frames as well as a new organization in polynucleotides allowing a more compact sequence of genetic material when compared to ancestral coronaviruses (Holmes & Rambaut, 2004).

Figure 2: ORF architecture comparison. (A) New *betacoronavirus* theoretical ancestor showing a wider open reading frame with 60 ORFs with small polypeptides translational sites (small red bars). (B) Canonical SARS-CoV-2 genome showing a compact ORF organization with 47 entries and less translational sites (small red bars) than *betacoronavirus* ancestor.



Open Reading Frame Viewer

Sequence

ORFs found: 60 Genetic code: 1 Start codon: 'ATG' only

(U) ORFfinder_5.12.20142515

1: 1..29K (29,059 nt)

Open Reading Frame Viewer

Sequence

ORFs found: 47 Genetic code: 1 Start codon: 'ATG' only

(U) ORFfinder_5.12.204226119

1: 1..30K (29,876 nt)

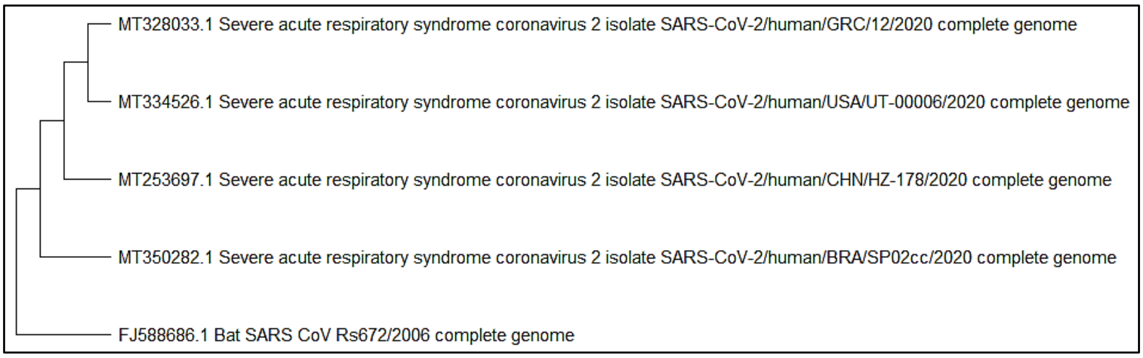
A
Find:
Tools
Tracks
1
2 K
4 K
6 K
8 K
10 K
12 K
14 K
16 K
18 K
20 K
22 K
24 K
26 K
29,059
ORF39 ORF23 ORF52 ORF35 ORF14 ORF11 ORF12
ORF1 ORF2 ORF45 ORF4 ORF5 ORF43 ORF42 ORF40 ORF38 ORF36 ORF34 ORF26 ORF50 ORF16 ORF19
ORF60 ORF48 ORF58 ORF7 ORF8 ORF10 ORF13 ORF28 ORF31
ORF46 ORF59 ORF57 ORF56 ORF9 ORF41 ORF22 ORF37 ORF55 ORF51 ORF27 ORF15 ORF18
ORF47 ORF53 ORF44 ORF24 ORF25 ORF30 ORF17 ORF20
ORF3 ORF6 ORF29 ORF33
ORF49
ORF54
Tracks shown: 2/6
B
Find:
Tools
Tracks
1
2 K
4 K
6 K
8 K
10 K
12 K
14 K
16 K
18 K
20 K
22 K
24 K
26 K
28 K
29,876
ORF7 ORF29 ORF9 ORF22 ORF11
ORF37 ORF13 ORF45 ORF18 ORF43 ORF42 ORF41 ORF20 ORF40 ORF21 ORF10 ORF25
ORF31 ORF46 ORF16 ORF17 ORF44 ORF1 ORF34 ORF2 ORF5
ORF12 ORF36 ORF14 ORF15 ORF30 ORF8 ORF28 ORF39 ORF3 ORF23
ORF47 ORF35 ORF33 ORF38
ORF19 ORF27 ORF24
ORF32
ORF4
Tracks shown: 2/7

Such a decrease in the genome may provide the key to understanding the greater degree of infectivity and viral load of current circulating variants when compared to SARS-CoV or even MERS-CoV which had generally larger ORF genomes and therefore spent more time and biochemical machinery for a cycle of replication. This greater infectivity rate highlights a bigger need for care with hand and surface hygiene thus preventing viral infection cycles, this practice is not always effective and comes up against infrastructure challenges inside and outside nosocomial environments (Bastian et al., 2021).

3.3 PHYLOGENETIC TREE CONSTRUCTION

The phylogeny construction (figure 3) for this new coronavirus variant showed a gradual formation of clades that probably reflect the evolutionary history of the virus. The known ancestor was considered the root group of the entire phylogeny, which corroborates the analysis while USA and Europe genomes formed a common clade followed by genomes from China. Moreover, a single clade was formed by Brazilian which indicates an existence of its own variants as observed in clinical and epidemiological findings (Lvov & Alkhovsky, 2020).

Figure 3: SARS-CoV-2 Evolutionary Analysis. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. The tree with the highest log likelihood is shown.


MT328033.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/GRC/12/2020 complete genome
MT334526.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/UT-00006/2020 complete genome
MT253697.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/CHN/HZ-178/2020 complete genome
MT350282.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/BRA/SP02cc/2020 complete genome
FJ588686.1 Bat SARS CoV Rs672/2006 complete genome

The phylogenetic tree with the highest log likelihood demonstrates uniformity in the branches of each clades which suggests small and occasional changes in genome organization, known as single nucleotide polymorphisms, being able to influence the distribution of different structural proteins isoforms of the capsid and associated envelope molecules that plays important role during the infection like the Spike protein (Hulswit et al., 2016).

4 DISCUSSION

Among the most varied groups of viruses infectious to humans, RNA viruses have stood out throughout history for their high mutation capacity and consequently a higher rate of evasion to innate and adaptive immunity mechanisms due to the production of different viral variants. Beta-coronaviruses are characterized by having genetic material consisting of coding RNA (RNA+) that serves as a template for new copies of viral genetic material as well as for the translation of viral proteins from the biochemical machinery of host cells (Blagova et al., 2020).

In addition to the difficulty presented by the high mutation rate in the genetic material of viral RNA, beta-coronaviruses bring an additional barrier in their detection and elimination during infectious processes since it is an enveloped virus that favors viral internalization and the concealment of proteins from the viral capsid. Despite the current state of the pandemic, few coronaviruses had caused large-scale harm to human health until then, which implies a recent change from previously harmless strains to highly infectious forms (Lvov & Alkhovsky, 2020).

The obtained data of genome organization in the most varied regions in the world demonstrated a tendency to decrease the size of nucleotide sequences when compared to the known ancestors for SARS-CoV-2 whereas ORFs also decreased in quantity and increased in its length. These findings suggests a specialization of the most derived forms of coronavirus that tends to increase the speed of replication and consequently the rate of infection, such variation of genomes has already been described in other works that point out that this might be the mechanism behind the formation of variants (Amsalem et al., 2021).

From a phylogenetic point of view, the built cladogram from the studied database demonstrates the formation of three representative clades with the ancestor bat coronavirus as a root. The first clade is the one that contains genomes from Asian origin showing a subpopulation of viral strains, followed by a European and North American clade that probably constitutes a unitary group due to the high movement among people between the United States and European countries that tends to spread the variants between these two regions (Holmes & Rambaut, 2004).

On the other hand, genomes originating in Brazil formed an isolated clade, probably for two reasons, such as the isolation that the other countries of the globe imposed on their air traffic with Brazil as well as the emergence of local variants such as P1 that quickly became the majority in the territory among Brazilian population, as can be seen from recently published epidemiological studies on this viral strain (Dos Santos et al., 2021).

5 CONCLUSION

In summary, the present work was able to show some significant changes in the organizational pattern of new coronaviruses genome at different studied regions, in addition to showing specific changes in the genomic sequence from the performed alignment, which evidences the occurrence of single nucleotide polymorphisms. The phylogenetic analysis corroborated the differences observed in the analysis of the ORFs and in the genomic alignment, forming three different groups of SARS-CoV-2 (Asian clade, Euro-American clade and Brazilian clade).

The present study serves as a starting point for further analysis of the collected genomes and may provide new evidence for the changes that are beginning to be described at the protein level - mainly regarding the emergence of isoforms of the Spike protein that has a fundamental role in the infection processes.

REFERENCES

Ahmadzadeh, J., Mobaraki, K., Mousavi, S. J., Aghazadeh-Attari, J., Mirza-Aghazadeh-Attari, M., & Mohebbi, I. (2020). The risk factors associated with MERS-CoV patient fatality: A global survey. *Diagnostic Microbiology and Infectious Disease*, *96*(3), 114876. https://doi.org/10.1016/j.diagmicrobio.2019.114876

Amsalem, D., Dixon, L. B., & Neria, Y. (2021). The Coronavirus Disease 2019 (COVID-19) Outbreak and Mental Health: Current Risks and Recommended Actions. In *JAMA Psychiatry* (Vol. 78, Issue 1, pp. 9–10). American Medical Association. https://doi.org/10.1001/jamapsychiatry.2020.1730

Bastian, M. S., Fonseca, C. D. da, & Barbosa, D. A. (2021). Os desafios da higienização das mãos de profissionais de saúde no pronto-socorro: revisão integrativa / The challenges of hand hygiene by healthcare professionals in the emergency room: integrative review. *Brazilian Journal of Health Review*, *4*(1), 485–499. https://doi.org/10.34119/bjhrv4n1-039

Blagova, O. V., Varionchik, N. V., Beraia, M. M., Zaidenov, V. A., Kogan, E. A., Sarkisova, N. D., & Nedostup, A. V. (2020). COVID-19 pneumonia in patients with chronic myocarditis (hbv-associated with infarct-like debute): specifics of the diseases course, the role of the basic therapy (Part II). *Rational Pharmacotherapy in Cardiology*, *16*(5), 730–736. https://doi.org/10.20996/1819-6446-2020-10-03

Dos Santos, C. A., Bezerra, G. V. B., Azevedo Marinho, A. R. R. A., Alves, J. C., Tanajura, D. M., & Martins-Filho, P. R. (2021). SARS-CoV-2 Genomic Surveillance in Northeast Brazil: Timing of Emergence of the Brazilian Variant of Concern P1. *Journal of Travel Medicine*, *2021*, 1–3. https://doi.org/10.1093/jtm/taab066

Holmes, E. G., & Rambaut, A. (2004). Viral evolution and the emergence of SARS coronavirus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1447), 1059–1065. https://doi.org/10.1098/rstb.2004.1478

Hulswit, R. J. G., de Haan, C. A. M., & Bosch, B. J. (2016). Coronavirus Spike Protein and Tropism Changes. In *Advances in Virus Research* (Vol. 96, pp. 29–57). Academic Press Inc. https://doi.org/10.1016/bs.aivir.2016.08.004

Kryukov, K., Ueda, M. T., Nakagawa, S., & Imanishi, T. (2020). Sequence compression benchmark (SCB) database-A comprehensive evaluation of reference-free compressors for FASTA-formatted sequences. *GigaScience*, *9*(7), 1–12. https://doi.org/10.1093/gigascience/giaa072

Lvov, D. K., & Alkhovsky, S. V. (2020). Source of the COVID-19 pandemic: Ecology and genetics of coronaviruses (Betacoronavirus: Coronaviridae) SARS-CoV, SARS-CoV-2 (subgenus Sarbecovirus), and MERS-CoV (subgenus Merbecovirus). *Voprosy Virusologii*, *65*(2), 62–70. https://doi.org/10.36233/0507-4088-2020-65-2-62-70

Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., Joukov, V., Lotov, V., Pannu, R., Rudnev, D., Shkeda, A., Weitz, E. M., & Schneider, V. A. (2021). Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Research*, *31*(1), 159–169. https://doi.org/10.1101/gr.266932.120

Sawicki, M. P., Samara, G., Hurwitz, M., & Passaro, E. (1993). Human Genome Project. *The American Journal of Surgery*, *165*(2), 258–264. https://doi.org/10.1016/S0002-9610(05)80522-7

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution*, *24*(8), 1596–1599. https://doi.org/10.1093/molbev/msm092

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260–1263. https://doi.org/10.1126/science.aax0902

Wu, Y. C., Chen, C. S., & Chan, Y. J. (2020). The outbreak of COVID-19: An overview. In *Journal of the Chinese Medical Association* (Vol. 83, Issue 3, pp. 217–220). Wolters Kluwer Health. https://doi.org/10.1097/JCMA.0000000000000270

Yang, M., Derbyshire, M. K., Yamashita, R. A., & Marchler-Bauer, A. (2020). NCBI’s Conserved Domain Database and Tools for Protein Domain Analysis. *Current Protocols in Bioinformatics*, *69*(1), e90. https://doi.org/10.1002/cpbi.90