SARS-CoV-2 genomic monitoring in the São Paulo State unveils new sublineages of the AY.43 strain

Alex Ranieri Jerônimo Lima^{1¶}, Gabriela Ribeiro^{1¶}, Vincent Louis Viala¹, Loyze Paola Oliveira de Lima¹, Antonio Jorge Martins¹, Claudia Renata dos Santos Barros¹, Elaine Cristina Marqueze¹, Jardelina de Souza Todao Bernardino¹, Debora Botequiao Moretti¹, Evandra Strazza Rodrigues², Elaine Vieira Santos², Ricardo Augusto Brassaloti³, Raquel de Lello Rocha Campos Cassano³, Pilar Drummond Sampaio Corrêa Mariani⁴, Luan Gaspar Clemente³, Patricia Akemi Assato⁵, Felipe Allan da Silva da Costa⁵, Mirele Daiana Poletti⁶, Jessika Cristina Chagas Lesbon⁶, Elisangela Chicaroni Mattos⁶, Cecilia Artico Banho⁷, Lívia Sacchetto⁷, Marília Mazzi Moraes⁷, Melissa Palmieri⁸, Fabiana Erica Vilanova da Silva⁹, Rejane Maria Tommasini Grotto^{5,10}, Jayme A. Souza-Neto⁵, Marta Giovanetti^{11,12}, Luiz Carlos Junior de Alcantara^{11,12}, Maurício Lacerda Nogueira⁷, Heidge Fukumasu⁶, Luiz Lehmann Coutinho³, Simone Kashima², Raul Machado Neto¹, Dimas Tadeu Covas^{1,2}, Svetoslav Nanev Slavov^{2*}, Maria Carolina Elias^{1*}, Sandra Coccuzzo Sampaio^{1*}

¹*Butantan Institute, São Paulo, Brazil*

²*University of São Paulo, Ribeirão Preto Medical School, Blood Center of Ribeirão Preto, Ribeirão Preto, SP, Brazil*

³*University of São Paulo, Centro de Genômica Funcional da ESALQ, Piracicaba, SP, Brazil*

⁴*NGS Soluções Genômicas, Piracicaba, SP, Brazil*

⁵*São Paulo State University (UNESP), School of Agricultural Sciences, Department of Bioprocesses and Biotechnology, Botucatu, Brazil*

⁶*Department of Veterinary Medicine, School of Animal Science and Food Engineering, University of São Paulo, Pirassununga, São Paulo, Brazil*

⁷*Medicine School of São José do Rio Preto (FAMERP), São José do Rio Preto, São Paulo, Brazil*

⁸*Coordenação de Vigilância em Saúde - Secretaria Municipal da Saúde, São Paulo*

⁹*Assistência Laboratorial/Coordenação de Atenção Básica - Secretaria Municipal da Saúde, São Paulo*

¹⁰*Molecular Biology Laboratory, Applied Biotechnology Laboratory, Clinical Hospital of the Botucatu Medical School, Brazil*

¹¹*Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*

¹²*Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil*

[¶]*These authors contributed equally to this work*

*Correspondence to: svetoslav.slavov@hemocentro.fmrp.usp.br, carolina.elias@sabbaga@butantan.gov.br, sandra.coccuzzo@butantan.gov.br

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Abstract

Delta VOC is highly diverse and more than 120 sublineages have been identified in Pango lineages with the continuous description of emerging ones. Brazil is now one of the most vaccinated countries against SARS-CoV-2 in the world which can enhance the emergence of viral mutations related to improved viral fitness. In this study, we identified two novel sublineages of the AY.43 lineage which were classified as AY.43.1 and AY.43.2 as observed on the specific clustering on the obtained phylogenetic tree. The novel sublineages were defined by the following characteristic nonsynonymous mutations ORF1ab:A4133V and ORF3a:T14I for AY.43.1 and ORF1ab:G1155C for AY.43.2. The majority of the analyzed sequences of both lineages were Brazilian, which shows that probably these two emerging sublineages have Brazilian origin. It is still unknown how these two sublineages are disseminated in São Paulo State and Brazil and their potential impact on the ongoing vaccination process. However, the performed study reinforces the importance of the SARS-CoV-2 genome monitoring for timely identification of emerging SARS-CoV-2 variants which can impact the ongoing SARS-CoV-2 vaccination and public health policies.

Introduction

Currently in Brazil have been applied more than 157 million doses of anti-SARS-CoV-2 vaccine and the individuals who have been fully vaccinated are approximately 124 million, which makes Brazil one of the most vaccinated nations in the world. In Brazil, on the background of this solid process of vaccination, the first cases of Delta VOC introductions were reported (1). Although now there is a full substitution of the dominant Brazilian gamma VOC by the Delta sublineages. Despite the full substitution of the pre-existing gamma VOC in Brazil with the delta VOC, no exponential growth of the new cases has been observed, most probably due to the ongoing vaccination. Delta VOC is highly diversified, and more than 120 sublineages have been classified within the Pango lineages (<https://cov-lineages.org/>) with the frequent description of novel lineages and sublineages.

We assume that the presence of massive vaccination in Brazil may also help for the selection of novel SARS-CoV-2 lineages and sublineages due to the accumulation of viral mutations leading to viral fitness. The molecular surveillance of the SARS-CoV-2 variants is of crucial importance for tracking the genomic profile of this virus and the accumulation of mutations which on one hand can alter the viral functions in terms of infectivity and transmissibility and on the other to be important for the emergence of novel variants, lineages, and sublineages which can exert significant pressure on the healthcare system of a given country (2).

In this study, we identified two novel Brazilian sublineages belonging to the AY.43 lineage, which were named AY.43.1 and AY.43.2, suggested due to their specific clade clustering within the main AY.43 lineage. Most of these sequences originated from the city of São Paulo and the genome monitoring was performed by the Network for Pandemic Alert of SARS-CoV-2 Variants of the São Paulo State.

Materials and methods

The raw sequence data obtained were submitted to quality control analysis using the FastQC (3) software version 0.11.8. Trimming was performed using Trimmomatic (4) version 0.3.9 in order to select the sequences with the best quality. Only sequences with quality scores >30 were used. We mapped the trimmed sequences against the SARS-CoV-2 reference (Genbank RefSeq NC_045512.2) using BWA (5) software and samtools (6) for read indexing. The mapped files were submitted to refinement using Pilon (7) to obtain the correct indels and insertions. The trimmed sequences were subjected to a remap against the genome refined by Pilon. Finally, we use bcftools (8) for variant calling and seqtk (9) for the assembly of the consensus SARS-CoV-2 genomes. Positions covered by fewer than 10 reads (DP<10) and bases of quality lower than 30 were considered a gap in coverage and converted to Ns. Coverage values for each genome were calculated using samtools v1.12. We assessed the consensus genome sequence quality using Nextclade v0.8.1 (<https://clades.nextstrain.org>).

Phylogenetic analysis was performed using the Nextstrain v3.0.3 SARS-CoV-2 workflow (10). To reconstruct the phylogeny of the AY.43 hypothetical sublineages we used 1,616 SARS-CoV-2 sequences generated from the Butantan Network for Pandemic Alert of SARS-CoV-2 Variants distributed between the epidemiological weeks 41 to 43 (GISAID accession ID in Supplementary File 1). The analyzed sequences were obtained from all 17 Health Divisions of the São Paulo State. A representative global dataset was retrieved from GISAID (3,711 sequences, downloaded from nextregions global in 2021, 10th of November - GISAID accession ID in Supplementary File 2) (<https://www.gisaid.org/>) which was used as background for our analysis.

Results

In the performed phylogenetic analysis we observed a large cluster of the AY.43 lineage containing 492 sequences, the majority of which (n=359, 73%) were from Brazil (Figure 1). From the Brazilian sequences, only one sequence was not generated from the Butantan Network for Pandemic Alert of SARS-CoV-2 Variants. An interactive tree can be accessed at the following link: https://nextstrain.org/fetch/repositorio.butantan.gov.br/bitstream/butantan/3990/1/ncov_EpiWeek_forty-one-to-forty-three.json. Additionally, the dissemination of these sequences in the São Paulo State showed that the majority were obtained from the metropolitan zone of São Paulo city.

Figure 1. Nextstrain tree colored by continent. The AY.43 lineage clade is indicated by the lines, and the box contain its information, such as number of sequences in the clade and the amino acid mutations present in all sequences.

The Brazilian AY.43 sequences were divided as two subclades within the AY.43 cluster: (i) 104 sequences sharing the amino acid mutations in ORF1ab: A4133V and in ORF3a:T14I; named AY.43.1 (ii) and 99 sequences, which presented the amino acid mutation in ORF1ab:G1155C and named AY.43.2 (Figure 2). We additionally observed other subclusters in the AY.43 clade, including in the proposed AY.43.1 and AY.43.2 clades. However, these were composed of a limited number of sequences and did not show sufficient support to be suggested as novel sublineages of AY.43.

Figure 2. AY.43 clade zoomed from Figure 1. The yellow clade contains 359 Brazilian sequences, which form two subclades: AY.43.1 in purple and AY.43.2 in blue. Dots in branches indicates amino acid mutations present in all sequences from that clade.

Discussion

In this study, we provide a preliminary report of the characterization of two novel AY.43 sublineages with probable Brazilian origin, due to the observed specific clustering and specific mutation profile observed in the phylogenetic tree generated by Nextstrain. Additionally, depending on the type of mutation (localization in ORF1a or ORF3a) we hypothesized the subdivision of the AY.43 strain into two sublineages designated as AY.43.1 and AY.43.2. The obtained data shows the importance of the SARS-CoV-2 genomic surveillance for the identification of emerging lineages. This is particularly important because SARS-CoV-2 emerging lineages can exert an enormous impact on the public health systems due to increased infectivity and transmission (2). In the newly characterized sublineages, we defined the mutations A4133V and G1155C in the ORF1a and the T14I mutation in the ORF3a, which were non-synonymous. To our knowledge, the impact of these mutations is unknown, except for the T14I, which shows deleterious effects on the viral proteins (11). A nonsynonymous mutation in ORF3a, which is a conserved protein involved in viral replication and release (12), may affect viral functions in addition to the mutational constellation defined for the Delta VOC. Based on the performed analysis the newly classified AY.43.1 and AY.43.2 sublineages probably have Brazilian origin as were composed of sequences obtained from Brazil. Further studies are necessary to investigate the dissemination in these two sublineages in Brazil, but preliminary results show that the majority of AY.43.1 and AY.43.2 sequences originated from the city of São Paulo.

In conclusion, we show that SARS-CoV-2 genome monitoring is crucial for the prompt characterization of SARS-CoV-2 novel lineages and sublineages. By this approach, we can timely detect the presence of novel SARS-CoV-2 variants and implement strategies for preventing their dissemination which can have further implications on the ongoing SARS-CoV-2 vaccination and public health policies.

Acknowledgements

We thank all contributors from GISAID. We also thank Gabriela Ribeiro and Glaucia Borges for their help with the Instituto Butantan's local database repository.

Funding

This work was supported by Butantan Foundation, Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant Numbers: 2020/10127; 2020/06441-2), Fundação Hemocentro Ribeirão Preto

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of the Faculty of Medicine of Ribeirão Preto, University of São Paulo (CAAE: 50367721.7.1001.5440).

Informed Consent Statement

Patient consent was waived once the diagnostic test had already been performed and the result has been communicated to the patients. Viral RNA was used, and the results

obtained will not compromise private information, neither bring any new clinical/therapeutic outcome for the patient.

References:

1. Patané J, Viala V, Lima L, Martins A, Barros C, Marqueze E, et al. SARS-CoV-2 Delta variant of concern in Brazil - multiple introductions, community transmission, and early signs of local evolution. medRxiv [Internet]. 2021 Oct 12 [cited 2021 Nov 24];2021.09.15.21262846. Available from: <https://www.medrxiv.org/content/10.1101/2021.09.15.21262846v2>
2. Khateeb J, Li Y, Zhang H. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. Critical Care. 2021.
3. Andrews S. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. 2010.
4. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;
5. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;
7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;
8. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. Bioinformatics. 2011;
9. Li H. seqtk Toolkit for processing sequences in FASTA/Q formats. GitHub. 2012;
10. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: Real-time tracking of pathogen evolution. Bioinformatics. 2018;
11. Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. mSystems. 2020;
12. Bianchi M, Borsetti A, Ciccozzi M, Pascarella S. SARS-Cov-2 ORF3a: Mutability and function. Int J Biol Macromol. 2021;