



# SARS-CoV-2 Omicron BA.1, BA.2, and XAG identification during routine surveillance on a university campus in Belo Horizonte, Brazil, 2022

Daniel Costa Queiroz<sup>1</sup> · Natália Virtude Carobin<sup>2</sup> · Luiza Campos Guerra de Araújo e Santos<sup>1</sup> · Paula Luize Camargos Fonseca<sup>1</sup> · Isabela Lorraine Braga-Paz<sup>1</sup> · Rillery Calixto Dias<sup>1</sup> · Jorge Gomes Goulart Ferreira<sup>1</sup> · Túlio Resende Freitas<sup>2</sup> · Diego Menezes<sup>1</sup> · Sílvia Cristina Verde Mendes Nolasco<sup>2</sup> · Simone Rodrigues Ribeiro<sup>2</sup> · Thaís Salviana Ribeiro<sup>2</sup> · Laura do Amaral Xavier<sup>2</sup> · Karina Braga Gomes<sup>2</sup> · Adriano de Paula Sabino<sup>2</sup> · Renato Santana de Aguiar<sup>1,3</sup> · Renan Pedra de Souza<sup>1</sup>

Received: 29 July 2022 / Accepted: 12 October 2022  
© The Author(s) under exclusive licence to Sociedade Brasileira de Microbiologia 2022

## Abstract

We report SARS-CoV-2 genomic surveillance results between Belo Horizonte, Brazil's third and fourth case waves. Samples were obtained through a routine university monitoring COVID-19 program from the 9th to the 22nd epidemiological weeks (March and June 2022). We identified ten samples from the BA.1 clade (BA.1, BA.1.1, and BA.1.14.1 lineages) and 45 samples from the BA.2 clade (BA.2, BA.2.56, BA.2.9, BA.2.62, BA.2.23, BA.2.81, and BA.2.10). We observed progressive replacement of the BA.1 by the BA.2 clade. Furthermore, two XAG recombinants were found in the 22nd week. Diversification of the omicron variant seems to have contributed to the resurgence of cases in Belo Horizonte, similarly to what has been reported in South Africa.

**Keywords** Genomic surveillance · Variants of concern · Phylogeny · Viral evolution · Recombinant variant

---

Responsible Editor: Luis Augusto Nero

---

Daniel Costa Queiroz and Natália Virtude Carobin contributed equally.

- ✉ Adriano de Paula Sabino  
adriansabin01@gmail.com
- ✉ Renato Santana de Aguiar  
santanarnt@gmail.com
- ✉ Renan Pedra de Souza  
renanpedra@gmail.com

- <sup>1</sup> Laboratório de Biologia Integrativa Departamento de Genética Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
- <sup>2</sup> Laboratório Institucional de Pesquisa Em Biomarcadores Laboratório de Hematologia Clínica Departamento de Análises Clínicas E Toxicológicas Faculdade de Farmácia, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 ICB – Pampulha, Minas Gerais 31270901 Belo Horizonte, Brazil
- <sup>3</sup> Instituto D'OR de Pesquisa E Ensino, Rio de Janeiro, Brazil

## Introduction

COVID-19 has strongly impacted Brazil. The country presents the second-largest death toll worldwide and a significant reduction in life expectancy due to the pandemic [1]. Brazil, a country with a continental proportion, showed substantial heterogeneity in the number of cases and deaths across its territory. Minas Gerais, a state with more than 20,000,000 inhabitants, reported more than 60,000 deaths reaching almost 10% of the Brazilian total. Belo Horizonte, the capital of the Minas Gerais state, presented one of the lowest state capital mortality rates in Brazil [2].

SARS-CoV-2 variants played a significant role in the COVID-19 waves in Brazil [3]. Four waves could be identified in Minas Gerais: the first was in mid-2020 when B.1.1.28 and B.1.1.33 variants were circulating [4–6]; the second was after the introduction of alpha (B.1.1.7) and gamma (P.1) variants in late 2020 [7, 8]. The other two waves were observed in early 2022 after introducing the omicron variant and its subvariants. Interestingly, the delta variant in Brazil did not result in increased cases [9, 10].

All genomic surveillance studies published from Minas Gerais have used samples from the Brazilian Public Health

System or private laboratories. Here, we conducted a study sampling from a university COVID-19 monitoring project that attends academic community members (students and staff). We aimed to evaluate SARS-CoV-2 variant circulation between the third and fourth waves (between March and June 2022) to understand whether viral lineages may have changed during the resurgence of cases.

## Methods

Epidemiological data from Belo Horizonte was obtained from the Secretary of Health. We estimated the changepoint in the number of cases time series with the *mcp* package. Medians were compared with the Wilcoxon test. Samples were obtained through the MonitoraCOVID (<https://monitracovid.ufmg.br/>), an outreach COVID-19 monitoring project from the largest university in the state, the Universidade Federal de Minas Gerais. Symptomatic subjects were initially remotely clinically assessed and had their tests scheduled. Out of the 111 SARS-CoV-2 positive samples collected between the 9th and 22nd epidemiological weeks (March 3 until June 1, 2022), fifty-seven nasopharyngeal swab specimens with an RT-qPCR cycle threshold below 30 were randomly selected. No samples were obtained from weeks 12, 13, 15, 16, and 18. All samples were residual positive COVID-19 clinical diagnostic samples de-identified before receipt by the researchers. The study was conducted according to the Declaration of Helsinki and approved by the Ethics Committee (protocol number 33202820.7.1001.5348).

Sequencing was performed using the Illumina platform, and bioinformatic analysis was carried out as described elsewhere [7, 9]. Sequencing libraries were assembled using the QIAseq SARS-CoV-2 Primer Panel (QIAGEN) with the ARTIC V4.1 primer pools and quantified by the QIAseq Library Quant Assay kit (QIAGEN). We successfully sequenced all selected samples. The data generated were processed with a custom pipeline (available on GitHub—<https://github.com/filiperomero2/ViralUnity>). The 57 consensus genomes were classified into Pango lineages (pangolin tool V. 4.1.2) [11] and Nextclade web application (v.2.3.0) (Supplementary Table 1). A dataset ( $n = 1468$ ) containing public reference genomes classified by Nextstrain as Omicron clades (Supplementary Table 2) was aligned using Minimap2, and a maximum likelihood phylogeny was inferred using IQ-tree v2.0.3 [12] under the GTR + F + I + G4 model to corroborate the classification.

## Results

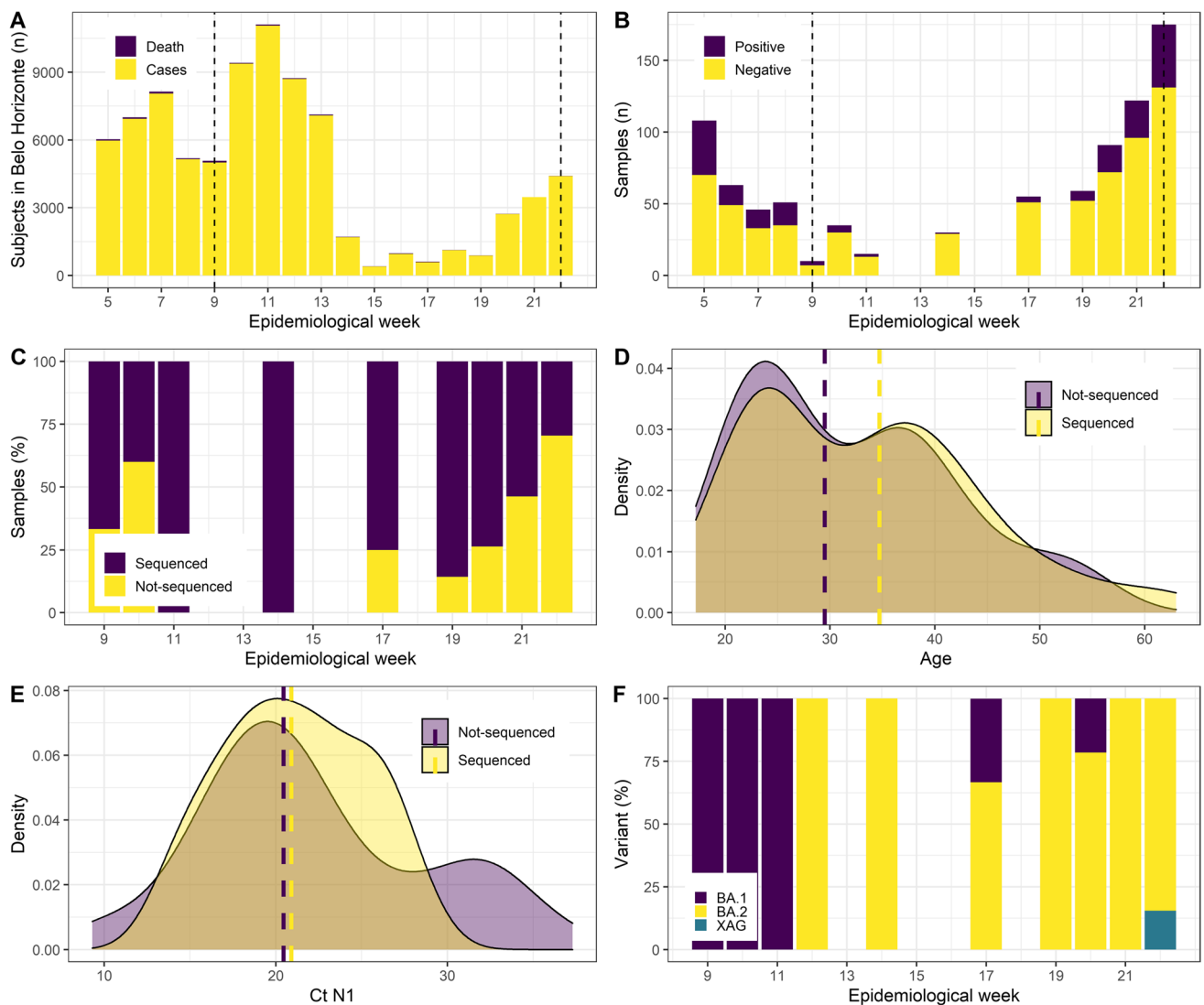
There was a significant change in the Belo Horizonte confirmed cases time series in the 20th week (Fig. 1A). Deaths in the period ranged from 0.03 to 4.65% of the

reported cases in the week. A reduction was observed in the MonitoraCOVID project from the 8th week and an increase (19th week) in diagnostic tests and their positivity (Fig. 1B). Of note, there was a student break from the 8th until the 13th week. One hundred and eleven diagnostic tests were positive (out of 481 tests performed) in the sampling period for sequencing (9th to 22nd week). The most frequently reported symptoms were odynophagia (79/111; 71.2%), cough (77/111; 69.4%), headache (64/111; 57.7%), myalgia (50/111; 45.0%), nasal congestion (47/111; 42.3%), coryza (44/111; 39.6%), and fever (25/111; 22.5%). Most subjects had completed the first vaccination scheme (110/111; 99.1%) and received the third dose (100/111; 90.0%). The first vaccination schemes included two doses of ChAdOx1 (68/110; 61.9%), BNT162b2 (27/110; 24.5%), and CoronaVac (15/110; 13.6%), while the third doses were with BNT162b2 (74/100; 74.0%), Ad26.COV2.S (16/100; 16.0%), and ChAdOx1 (10/100; 10.0%).

57/111 samples (51.35%) were successfully sequenced with weekly representation ranging from 29.55% (22nd week) to 100% (11th and 14th week) (Fig. 1C). Median age (Fig. 1D) and cycle threshold (Fig. 1E) between sequenced and non-sequenced samples did not differ ( $p = 0.514$  and  $p = 0.645$ , respectively). 38/57 (66.67%) of the sequenced samples were from females. The median genome coverage was 79% (ranging from 51 to 94%), and sequencing depth was 670.60 times (ranging from 608.36 to 5832.09 times) (Supplementary Table 1).

According to the Nextclade classification, we identified 10 samples from the BA.1 clade: BA.1 ( $n = 2$ ), BA.1.1 ( $n = 6$ ), and BA.1.14.1 ( $n = 2$ ) and 45 samples in the BA.2 clade: BA.2 ( $n = 33$ ), BA.2.9 ( $n = 2$ ), BA.2.23 ( $n = 1$ ), BA.2.10 ( $n = 1$ ), BA.2.56 ( $n = 2$ ), BA.2.62 ( $n = 2$ ), and BA.2.81 ( $n = 4$ ) (Supplementary Table 1). Two samples were from the XAG recombinant of BA.1/BA.2 in the 22nd week. The distribution throughout the period indicates that BA.1 clade was the most present when the number of cases declined, and the resurgence happened with BA.2 clade variants circulating (Fig. 1D).

Our phylogenetic analysis corroborated the assignment of the genomes generated in our study in BA.1, BA.2, and XAG recombinant lineages (Fig. 2A). The XAG is considered a recombination between BA.1 and BA.2, showing characteristic mutations of both variants. ORF1ab gene C241T, A2832G, C2857T, C3037T, T5386G, and C5585A mutations are likely derived from the BA.1, while the non-synonymous mutations in spike gene (T19I, D405N, 5408S) and the absence of the H69/V70 deletions derived from the BA.2. XAG also presents some other mutations: T2790C, A4184G, A334G, and C17502G (Fig. 2B).



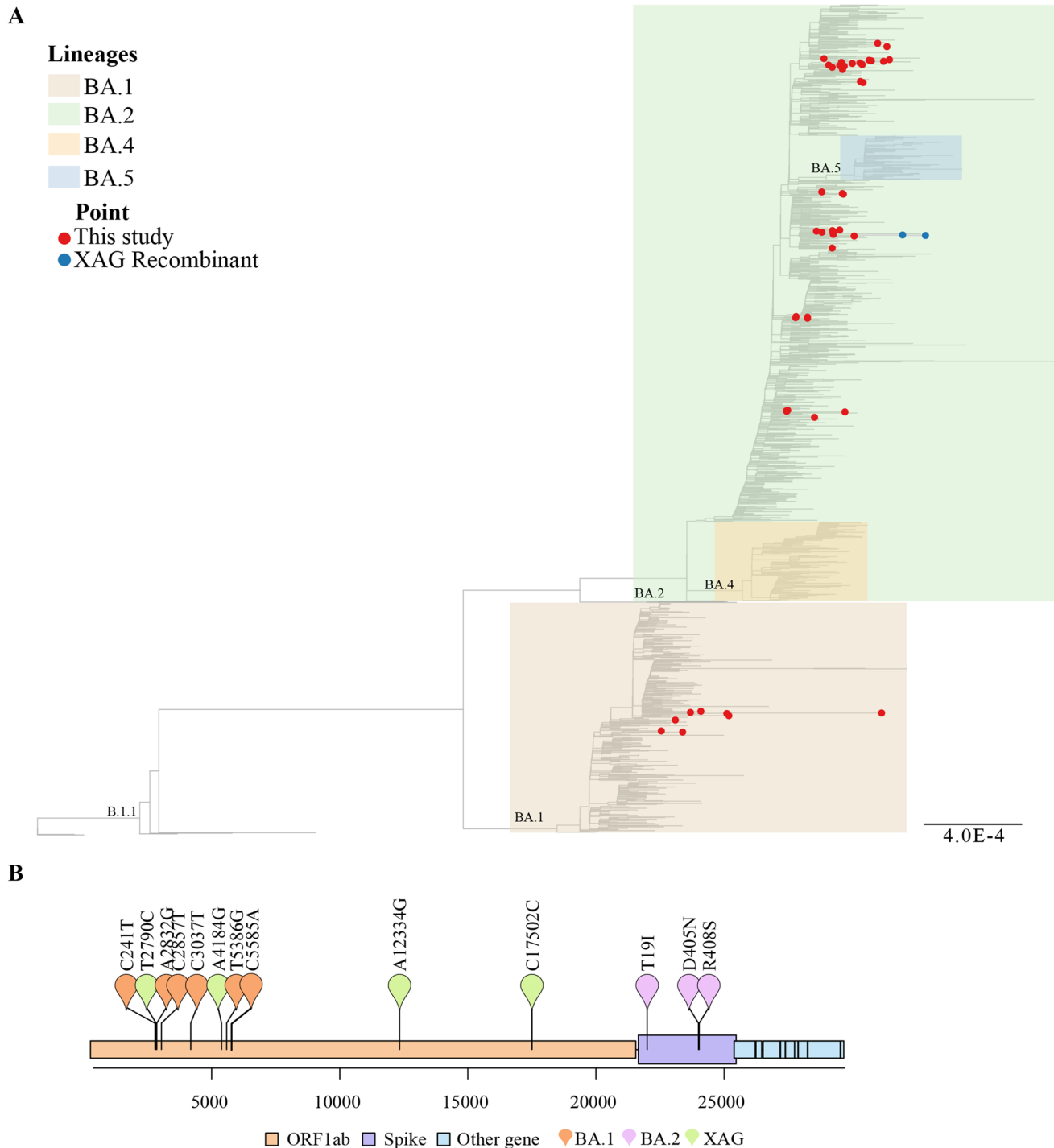
**Fig. 1** Epidemiological and genomic surveillance results. **A** Distribution of the number of cases and deaths in the city of Belo Horizonte, capital of Minas Gerais state. Dashed lines indicate the period of sampling for sequencing. **B** Distribution of the number of tests conducted in the MonitoraCOVID project between the 5th and 22nd epidemiological week and their results. Dashed lines indicate the period of sampling for sequencing. **C** Percentage of the positive samples that were successfully sequenced between the 9th and 22nd epidemiological weeks. **D** Age distribution between sequenced (median 34.7) and not-sequenced samples (median 29.5). Dashed lines indicate the median age in each group. **E** N1 target cycle threshold (Ct) distribution between sequenced (median 20.9) and not-sequenced (median 20.4) samples. Dashed lines indicate the median Ct in each group. **F** Distribution of the BA.1, BA.2, and XAG variants found in the study period

## Discussion

Genomic surveillance has always been a powerful tool in epidemiological investigations, but its use has grown substantially during the COVID-19 pandemic. Tracking the SARS-CoV-2 variants has proven to be challenging. Sequencing results had to be obtained and communicated promptly and efficiently so genomic findings could be helpful in public policy [13]. SARS-CoV-2 evolution has contributed to changes in viral transmissibility and infectivity, leading the World Health Organization to designate certain

strains as a variant of interest or concern. Omicron is the only variant of concern currently circulating, and its subvariants are under close monitoring.

BA.1, BA.2, and BA.3 variants contributed to South Africa's fourth COVID-19 wave [14]. The BA.1 was associated with increased cases in late 2021–early 2022 in the Amazonas and the Rio Grande do Sul states, according to virological.org posts 783 and 785. Similarly, our results indicate that the decline observed between weeks 9th and 11th was associated with higher BA.1 circulation. When the cases showed the second increase in 2022, between



**Fig. 2** Lineage classification results. **A** Maximum likelihood inference phylogeny using a global reference dataset. Red circles represent the genomes generated in our study ( $n=57$ ). Blue circles represent the XAG recombinant genomes. Different colors represent BA.1, BA.2, BA.4, and BA.5. **B** Schematic representation of SARS-CoV-2

XAG recombinant lineage. Orange dots (ORF1ab gene) represent conserved mutations frequently found in BA.1 lineage. Purple dots (spike gene) represent conserved mutations commonly found in BA.2 lineage, while green dots (ORF1ab) represent mutations found in XAG recombinant variant

weeks 19th and 22nd, the BA.2 clade was the most common. The replacement of BA.1 by the BA.2 clade has also been reported in Japan [15] and England [16]. Many omicron subvariants and lineages have been described [14].

Our results indicated two BA.1 and five BA.2 sub-lineages circulating between the third and fourth waves suggesting a greater genomic diversification of these two subvariants.

We also report two genomes classified as omicron XAG recombinant. The XAG clade is the first BA.1 and BA.2 recombinant lineage found in Brazil. The first Brazilian identification happened in the Rio Grande do Sul state in March 2022. Until recently, 260 genomes (186 identified in Brazil) classified as XAG recombinant by Pango lineages were available in the GISAID database (available at <https://www.gisaid.org/> and accessed on July 22, 2022). The XAG has already been identified in six other Brazilian states (Distrito Federal, Paraná, Pernambuco, Rio de Janeiro, Santa Catarina, and São Paulo) and other countries (e.g., Argentina, Canada, Chile, Colombia, USA, Israel) (available at <https://github.com/cov-lineages/pango-designation/issues/709> and accessed on July 22, 2022). The description of other SARS-CoV-2 recombinants has already been reported, such as XD and XF, derived from Delta and Omicron BA.1, and XE, derived from subvariants Omicron BA.1 and BA.2 [17]. However, most data has been unpublished thus far. Since the number of omicron genomic diversification and recombinants has increased significantly, continuous genomic monitoring is of great importance to help in epidemiological surveillance studies and understanding these mutations' effects on viral behavior.

Our study presents limitations. Since samples were obtained from a university testing program, younger subjects were overrepresented compared to the city age distribution. Of note, at the time of the sampling (9th to 22nd week), younger subjects had not had a fourth vaccination shot available in the city. Another selection bias may have happened since the project tests symptomatic subjects. Although BA.2 infections seem to lead to a higher viral load than BA.1 [15], no clinical outcome difference has been suggested between BA.1 and BA.2 variants [18, 19]. The inability to sequence samples with lower viral levels may have also led to bias. We could also not estimate the BA.2 introduction date due to insufficient or inexistent sampling in some epidemiological weeks. Despite the mentioned limitations, integrating genomic surveillance with routine university monitoring was successfully achieved. Further studies are necessary to follow omicron subvariants dispersion and track new variants.

## Data sharing

All consensus genome sequences from this study have been deposited on GISAID and are publicly available (IDs: EPI\_ISL\_13948217—EPI\_ISL\_13948273). The Supplementary Tables and phylogeny are available on our GitHub repository page <https://github.com/LBI-lab/SARS-CoV-2-Omicron-BA.1-and-BA.2-during-routine-surveillance-on-a-university-campus-in-Belo-Horizont>.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s42770-022-00848-5>.

**Acknowledgements** We want to thank all authors who submitted their data to GISAID and the EpiCoV curation team for their work. The accession numbers from all sequences used in this study are available in Supplementary Table 2.

**Funding** We acknowledge support from the Rede Corona-ômica BR MCTI/FINEP affiliated with RedeVírus/MCTI (01.20.0029.000462/20 404096/2020–4; 1227/21 01.22.0074.00), Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq (315592/2021–4), Financiadora de Estudos e Projetos—FINEP (0494/20 01.20.0026.00; 1228/21 01.22.0082.00; 1139/20 01.20.0076.00; Rede Diagnóstico), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES (Finance Code 001), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro—FAPERJ (202.922/2018), Fundação de Amparo à Pesquisa do Estado de Minas Gerais—FAPEMIG (APQ-00800–19), Centro de Laboratórios Multiusuários—CELAM-ICB-UFGM, and Laboratório Institucional de Pesquisa em Biomarcadores—LINBIO-FAFAR-UFGM.

## Declarations

**Conflict of interest** The authors declare no competing interests.

## References

1. Castro MC, Gurzenda S, Turra CM, Kim S, Andrasfay T, Goldman N (2021) Reduction in life expectancy in Brazil after COVID-19. *Nat Med* 27(9):1629–1635
2. Brizzi A, Whittaker C, Servo LMS et al. (2022) Spatial and temporal fluctuations in COVID-19 fatality rates in Brazilian hospitals. *Nat Med* 28(7):1476–1485
3. Alcantara LCJ, Nogueira E, Shuab G et al (2022) SARS-CoV-2 epidemic in Brazil: how the displacement of variants has driven distinct epidemic waves. *Virus Res* 315:198785
4. Xavier J, Giovanetti M, Adelino T et al (2020) The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *Emerg Microbes Infect* 9(1):1824–1834
5. Candido DS, Claro IM, de Jesus JG et al (2020) Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369(6508):1255–1260
6. Silva AVFG, Menezes D, Moreira FRR et al (2022) Seroprevalence, prevalence, and genomic surveillance: monitoring the initial phases of the SARS-CoV-2 pandemic in Betim, Brazil. *Front Microbiol* 13:799713
7. Moreira FRR, Bonfim DM, Zauli DAG et al (2021) Epidemic spread of SARS-CoV-2 lineage B.1.1.7 in Brazil. *Viruses* 13(6):984
8. Giovanetti M, Fonseca V, Wilkinson E, et al. (2022) Replacement of the Gamma by the Delta variant in Brazil: impact of lineage displacement on the ongoing pandemic. *Virus Evol* 8(1):veac024
9. Silva JP, de Lima AB, Alvim LB et al (2022) Delta variant of SARS-CoV-2 replacement in Brazil: a national epidemiologic surveillance program. *Viruses* 14(5):847
10. Fonseca PLC, Moreira FRR, de Souza RM et al (2022) Tracking the turnover of SARS-CoV-2 VOCs gamma to delta in a Brazilian state (Minas Gerais) with a high-vaccination status. *Virus Evol* 8(2):veac064

11. O'Toole Á, Scher E, Underwood A, et al. (2021) Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* 7(2):veab064
12. Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37(5):1530–1534
13. Romano CM, Melo FL (2021) Genomic surveillance of SARS-CoV-2: a race against time. *Lancet Reg Health Am* 1:100029
14. Viana R, Moyo S, Amoako DG et al (2022) Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 603(7902):679–686
15. Hirotsu Y, Maejima M, Shibusawa M, et al. (2022) SARS-CoV-2 Omicron sublineage BA.2 replaces BA.1.1: genomic surveillance in Japan from September 2021 to March 2022. *J Infect* 85(2):174–211
16. Elliott P, Eales O, Steyn N, et al. (2022) Twin peaks: the Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England. *Science* 376(6600):eabq4411
17. Wang L, Zhou HY, Li JY, et al (2022) Potential intervariant and intravariant recombination of delta and omicron variants. *J Med Virol* 94(10):4830–4838
18. Lewnard JA, Hong VX, Patel MM, Kahn R, Lipsitch M, Tartof SY (2022) Clinical outcomes associated with SARS-CoV-2 Omicron (B.1.1.529) variant and BA.1/BA.1.1 or BA.2 subvariant infection in southern California. *Nat Med* 28(9):1933–1943
19. Wolter N, Jassat W, DATCOV-Gen author group, von Gottberg A, Cohen C (2022) Clinical severity of omicron lineage BA.2 infection compared with BA.1 infection in South Africa. *Lancet* 400(10346):93–96

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.