



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Biologicals

journal homepage: [www.elsevier.com/locate/biologicals](http://www.elsevier.com/locate/biologicals)

## Molecular characterization and sequencing analysis of SARS-CoV-2 genome in Minas Gerais, Brazil

Giulia Magalhães Ferreira<sup>a</sup>, Ingra Morales Claro<sup>b,c</sup>, Victória Riquena Grosche<sup>a,d</sup>, Darlan Cândido<sup>b,e</sup>, Diego Pandeló José<sup>f</sup>, Esmenia Coelho Rocha<sup>b,c</sup>, Thaís de Moura Coletti<sup>b,c</sup>, Erika Regina Manuli<sup>b,c</sup>, Nelson Gaburo Jr.<sup>g</sup>, Nuno Rodrigues Faria<sup>b,e,h</sup>, Ester Cerdeira Sabino<sup>b,c</sup>, Jaqueline Goes de Jesus<sup>b,c</sup>, Ana Carolina Gomes Jardim<sup>a,d,\*</sup>

<sup>a</sup> Institute of Biomedical Science, Federal University of Uberlândia, Uberlândia, MG, Brazil

<sup>b</sup> Institute of Tropical Medicine, University of São Paulo Medical School, São Paulo, Brazil

<sup>c</sup> Department of Infectious and Parasitic Diseases, University of São Paulo Medical School, São Paulo, Brazil

<sup>d</sup> Institute of Bioscience, Humanities and Exact Sciences, São Paulo State University, São José do Rio Preto, São Paulo, Brazil

<sup>e</sup> Department of Zoology, University of Oxford, Oxford, UK

<sup>f</sup> Federal University of Triângulo Mineiro, Iturama, MG, Brazil

<sup>g</sup> DB, Diagnósticos do Brasil, São Paulo, Brazil

<sup>h</sup> MCR Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, UK

### ARTICLE INFO

#### Keywords:

SARS-CoV-2  
COVID-19  
Genomic surveillance  
Minas gerais  
Genome sequencing  
B.1 lineage

### ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first identified in Wuhan, China, is the causative agent of the coronavirus disease 2019 (COVID-19). Since its first notification in São Paulo state (SP) on 26<sup>th</sup> February 2020, more than 22,300,000 cases and 619,000 deaths were reported in Brazil. In early pandemic, SARS-CoV-2 spread locally, however, over time, this virus was disseminated to other regions of the country. Herein, we performed genomic sequencing and phylogenetic analysis of SARS-CoV-2 using 20 clinical samples of COVID-19 confirmed cases from 9 cities of Minas Gerais state (MG), in order to evaluate the molecular properties of circulating viral strains in this locality from March to May 2020. Our analyses demonstrated the circulation of B.1 lineage isolates in the investigated locations and nucleotide substitutions were observed into the genomic regions related to important viral structures. Additionally, sequences generated in this study clustered with isolates from SP, suggesting a dissemination route between these two states. Alternatively, monophyletic groups of sequences from MG and other states or country were observed, indicating independent events of virus introduction. These results reinforce the need of genomic surveillance for understand the ongoing spread of emerging viral pathogens.

### 1. Introduction

The World Health Organization (WHO) was informed on 31<sup>st</sup> December 2019 on the occurrence of unknown etiology respiratory disease cases in Wuhan, China [1]. Chinese authorities isolated and identified this pathogen as a novel coronavirus, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 2019 (COVID-19) [2]. SARS-Cov-2 belongs to the genus *Betacoronavirus*, subgenera *Sarbecovirus*, sub-family *Orthocoronavirinae* and family *Coronaviridae* [3]. Based on phylogenetic analysis, SARS-CoV-2 has been divided in two lineages, A and B,

according to the recent proposed lineage nomenclature [4,5]. Results from next-generation sequencing analysis have shown that SARS-Cov-2 has homology to others coronaviruses (CoVs), such as SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV) [2,6].

SARS-Cov-2 was declared a public health emergency of international concern in January 2020 [7], and, since then, COVID-19 caused over 293,000,000 cases and more than 5,450,000 deaths around the world [8]. The case fatality rates of this disease ranges from 1.2 to 1.6%, although, in over 60s the ratio considerably increases [9–11]. When there was no licensed antiviral for SARS-CoV-2 and vaccination coverage worldwide was limited, several countries mainly used

\* Corresponding author. Institute of Biomedical Science (ICBIM), Federal University of Uberlândia (UFU), Brazil.

E-mail address: [jardim@ufu.br](mailto:jardim@ufu.br) (A.C.G. Jardim).

<https://doi.org/10.1016/j.biologicals.2022.08.001>

Received 25 February 2022; Received in revised form 14 June 2022; Accepted 21 August 2022

Available online 2 September 2022

1045-1056/© 2022 International Alliance for Biological Standardization. Published by Elsevier Ltd. All rights reserved.

non-pharmaceutical interventions (NPIs) in an attempt to control the pandemic [12,13]. However, despite these measures have presented positive effects, studies have demonstrated that SARS-CoV-2 transmission from asymptomatic or pre-symptomatic individuals complicates public health efforts to combat COVID-19 [12,14,15].

On 8<sup>th</sup> December 2020, the United Kingdom became the first occidental country to vaccinate the population as a prophylaxis against COVID-19. After that, other countries initiated the COVID-19 vaccination programs, using different types of vaccine based on the methodology and technology of production, in a two doses protocol [16]. To date, the global efforts against COVID-19 resulted in 66,5% of global population vaccinated with a minimum one dose of vaccine, and 60,6% got the complete vaccination schedule, including the booster dose [17]. With the advancement of vaccination, the number of Covid-19 cases and related deaths significantly decreased, allowing to the countries to relax the NPIs measures [16,18]. Currently, 77,84% of Brazilian population are vaccinated with a completed initial protocol and 8.20% of people are only partially vaccinated [19].

SARS-CoV-2 replication can progress to different clinical manifestations of COVID-19, and, for this reason, numerous approaches to the development of licensed antiviral therapies have being investigated for the treatment of COVID-19 [13]. Up to date, Remdesivir is the only drug approved by the American Food and Drug Administration (FDA) for the treatment of COVID-19 [20]. Additionally, Ritonavir-boosted nirmatrelvir (Paxlovid), molnupiravir, and anti-SARS-CoV-2 monoclonal antibodies (mAbs) have received Emergency Use Authorizations from the FDA for the treatment of COVID-19 [20]. These drugs prevent viral replication through various mechanisms of action, including blocking SARS-CoV-2 entry, inhibiting the activity of the SARS-CoV-2 3-chymotrypsin-like protease (3CLpro) and RNA-dependent RNA polymerase (RdRp), and causing lethal viral mutagenesis [21,22]. In Brazil, the Brazilian Health Regulatory Agency (Anvisa) recommends the FDA-approved antiviral treatments for COVID-19 [23]. Molnupiravir was approved by the ANVISA in early May 2022 [24]. What is more, viral replication demonstrates to be active early in the course of COVID-19 [25,26], and as a consequence of that, antiviral therapy presents the greatest impact before the illness progresses to the hyper-inflammatory condition that can characterize the later stages of disease, including critical illness [25].

In Brazil, the first confirmed case of COVID-19 was reported in São Paulo state (SP) on 26<sup>th</sup> February 2020 [6]. Analysis of the first two whole-genome sequences of SARS-CoV-2, isolated from Brazilian patients who had recent returned from Italy, demonstrated two independent events of virus introduction into the country [27]. Since then, COVID-19 is responsible for causing 22,300,000 cases and more than 619,000 deaths in the country [8]. During the early pandemic in Brazil, SARS-CoV-2 spread locally. However, despite interventions to prevent the virus dissemination, afterwards, large urban centers have become responsible for spreading the virus to other locations. According to Candido and coworkers data, over 100 international introductions of virus were observed in Brazil in 2020, and most of Brazilian strains were classified in three clades [12]. Clade 1 circulated predominantly in SP state and presented a nucleotide substitution in the spike protein; clade 2 was a widespread lineage, found in a several Brazilian states and was characterized by two nucleotide substitution (ORF 6 and nucleoprotein); and clade 3 was predominant in Ceará state (CE) [12].

Minas Gerais (MG) ranks as the second most populous state and the fourth largest area in Brazil, as well as it represents the third position in the Gross Domestic Product (GDP) values. MG shares borders with the states of SP, Bahia (BA), Rio de Janeiro (RJ), Goiás (GO), and Mato Grosso do Sul (MS). The city of Belo Horizonte, the capital of MG, is the major urban and finance center in Latin America [28]. Due to its large population size, economic situation, and facilitated access to other economy important states, MG has represented a potential region of SARS-CoV-2 dissemination, and potentially contributed to aggravate the pandemic [29]. To date, about 2,230,000 cases and over 56,660 deaths

by COVID-19 have been reported in MG [30]. According to Xavier and colleagues, the majority of sequences analyzed from samples collected in MG up to April 2020 was classified as SARS-Cov-2 lineage B.1, which contains sequences from the United States of America (USA), Australia, China, and other countries [29]. Herein, we performed genome sequencing analysis of SARS-CoV-2 using 20 clinical samples of COVID-19 confirmed cases from 9 cities of MG to evaluate the molecular properties of circulating viral strains in the state from March to May 2020.

## 2. Methods

### 2.1. Ethics statement and samples

Samples used in this study were collected via nasopharyngeal swab in private medical diagnostic laboratories. Ethical approval for this study was obtained from the National Ethical Review Board with approval number CAAE 30127020.0.0000.0068.20 SARS-CoV-2 samples with RT-qPCR positive results, collected between March 25th and May 25th, 2020, from patients attending private laboratories of diagnosis from 9 cities of the state of Minas Gerais (MG), Brazil, were selected. Samples were processed for genome sequencing at the Institute of Tropical Medicine University of São Paulo (IMT-USP). Metadata included information on samples (date and municipality of collection, and cyclethreshold (Ct) of SARS-CoV-2 detection by RT-qPCR) and data from patients (gender and age) are presented in Table 1.

### 2.2. cDNA synthesis and virus multiplex PCR amplification

Viral RNA was used for cDNA transcription using Protoscript II First Strand cDNA synthesis Kit (New England Biolabs, UK) and random hexamers (Thermo Fisher Scientific, USA). Whole genome amplification was performed by multiplex PCR using SARS-CoV-2 primers described previously (<https://artic.network/ncov-2019>) and Q5 High-Fidelity DNA polymerase (New England Biolabs, UK) [31]. PCR conditions have been previously reported (<https://artic.network/ncov-2019>). PCR products were purified using the 1x AMPure XP beads (Beckman Coulter, United Kingdom) and quantified using fluorimeter with the Qubit dsDNA High Sensitivity assay on the Qubit 3.0 instrument (Life Technologies, USA).

### 2.3. Whole genome sequencing and genome assembly

About 1 ng/ $\mu$ L of DNA from each of the 20 samples selected to this study was used to proceed to the library preparation. Amplicons from each sample were normalized and pooled in an equimolar fashion and barcoded using the EXP-NBD104 (1–12) and EXPNBD114 (13–24) Native Barcoding Kits (Oxford Nanopore Technologies, UK), following a previously published protocol [31]. After barcoding ligation, libraries were loaded on a flow cell and sequenced by MinION for 8–24 h using SQK-LSK109 Kit (ONT, UK). To monitor sequencing in real-time and estimate the depth of coverage (target of 200-fold) across the genome for each barcoded sample (<https://artic.network/rampart>), RAMPART software from the ARTIC Network (<https://artic.network/ncov-2019>) was used. After generated reads, fast 5 files were base called, demultiplexed, and trimmed using Guppy software v2.2.7 (ONT, UK). Minimap2 v2.28.0 was used to obtain the consensus genomes by mapping the fast files to the reference genome of SARS-CoV-2 isolate Wuhan-Hu 1 (GenBank Accession Number MN908947) and SAM tools were used to converted these files in a sorted BAM file [32]. The quality test and length filtering were performed for each barcode using guppyplex (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>). Bio-Edit was used to build a multiple sequence alignment of the resulting dataset [33,34].

**Table 1**

Epidemiological Information and Lineages of SARS-CoV-2 identified on samples investigated in the study, and date from patients.

CADDE ID	Sample	Ct value	Collection date	Age	Gender	State	Municipality	Lineage	Most common countries
MG 1	Swab	25,59	25/03/20	58	Male	MG	Extrema	B.1.1.28	Australia, United Kingdom
MG 4	Swab	9,96	30/03/20	80	Male	MG	Uberlândia	B.1.1.28	Australia, United Kingdom
MG 6	Swab	20,58	31/03/20	73	Male	MG	Uberaba	B.1	USA, Spain, United Kingdom
MG 7	Swab	17,90	01/04/20	46	Male	MG	CambuÍ	B.1.1.28	Australia, United Kingdom
MG 8	Swab	20,09	02/04/20	63	Female	MG	Pouso Alegre	B.1.1.28	Australia, United Kingdom
MG 9	Swab	18,62	02/04/20	32	Female	MG	CambuÍ	B.1.1.33	USA, United Kingdom
MG 10	Swab	19,04	02/04/20	79	Female	MG	CambuÍ	B.1.1.33	USA, United Kingdom
MG 13	Swab	19,99	07/04/20	82	Male	MG	Uberlândia	B.1.1.28	Australia, United Kingdom
MG 15	Swab	16,87	07/04/20	25	Female	MG	Formiga	B.1.1.28	Australia, United Kingdom
MG 18	Swab	18,98	08/04/20	62	Female	MG	Extrema	B.1.1.33	USA, United Kingdom
MG 19	Swab	15,78	09/04/20	75	Male	MG	Santos Dumont	B.1	USA, Spain, United Kingdom
MG 20	Swab	14,89	09/04/20	25	Female	MG	Pouso Alegre	B.1	USA, Spain, United Kingdom
MG 21	Swab	18,67	15/04/20	31	Male	MG	Juiz de Fora	B.1	USA, Spain, United Kingdom
MG 22	Swab	22,09	15/04/20	46	Male	MG	Juiz de Fora	B.1.1.33	USA, United Kingdom
MG 24	Swab	23,43	17/04/20	61	Male	MG	Barbacena	B.1.1.33	USA, United Kingdom
MG 34	Swab	15,55	17/05/20	38	Male	MG	Juiz de Fora	B.1.1.28	Australia, United Kingdom
MG 38	Swab	14,61	18/05/20	69	Male	MG	Juiz de Fora	B.1.1.33	USA, United Kingdom
MG 42	Swab	19,99	19/05/20	56	Male	MG	Juiz de Fora	B.1	USA, Spain, United Kingdom
MG 51	Swab	16,89	23/05/20	30	Female	MG	Juiz de Fora	B.1.1.28	Australia, United Kingdom
MG 54	Swab	17,70	25/05/20	37	Female	MG	Juiz de Fora	B.1.1.33	USA, United Kingdom

#### 2.4. Collation of SARS-CoV-2 global datasets

The dataset of this study was generated using 20 sequences of SARS-CoV-2 genomes from samples collected in 9 MG cities. Additionally, 31 sequences from other regions of the state, available at GISAID, were included to the dataset [35] (<https://www.gisaid.org>), resulting in 51 whole genome sequences from samples of 16 cities of MG. The dataset generated with sequences from MG represents approximately 1 sequence for every 157 cases (0.63%) notified up to May 27th, 2020. Juiz de Fora was the third city in MG with the highest number of SARS-CoV-2 notified cases, including 3.18% of all notified cases in this city during the period of samples collection, representing about 1 sequence for every 31 cases [36].

#### 2.5. Phylogenetic analysis of SARS-CoV-2 in Minas Gerais

The complete dataset was generated with the 20 sequences from this study, as well as sequences from GISAID platform, totaling 1637 sequences of whole genomes of SARS-CoV-2 globally distributed. Wherein, 57.7% (945/1637) and 42.3% (692/1637) of the sequences represent isolates from Brazil and from other countries, respectively. Among sequences from Brazil, 51 were generated from samples collected in MG, being 20 of those new sequences generated in this study. Sequences used for analysis had the genome coverage >75% in average. MAFFT was used to build a multiple sequence alignment of the dataset. The accuracy of the observed substitutions and frameshift of sequences was further validated by manually and carefully verifying the aligned files. Sequences edition and phylogenetic reconstruction were performed using AliView [34]. A maximum likelihood phylogenetic tree was estimated using Hasegawa-Kishino Yano nucleotide substitution model with a gamma distributed rate variation among-site [37] in IQTree v.2.26 [38]. Finally, SARS-CoV-2 lineages were identified using Phylogenetic Assignment of Named Global Outbreak Lineages tool (<https://github.com/cov-lineages/pangolin>).

#### 2.6. Nucleotide substitutions and protein sequence prediction

Fasta files of sequences generated in this study were aligned to the reference genome of SARS-CoV-2, isolate Wuhan-Hu 1 (GenBank Accession Number NC\_045512.2), and the nucleotide substitutions were obtained according to reference sequence. The online platform PROVEAN (Protein Variation Effect Analyzer) was used to provide a prediction for a protein sequence [39]. The aligned sequences generated in this study referent to the coding regions of each protein were translated

and any genetic differentiation identified was analyzed in PROVEAN protein tool. As input, this tool accepts a protein sequence and amino acid variations and performs a BLAST search to identify homologous sequences and generates scores. Variants with a score equal to or below -2.5 are considered deleterious and variants with a score above -2.5 are considered neutral [39].

### 3. Results

62 samples from clinically suspected cases of COVID-19, collected at the private laboratory Diagnosis of Brazil (DB) in 9 cities of Minas Gerais state (MG) were screened for the detection of SARS-CoV-2 by RT-PCR. 20 samples with RT-qPCR positive results for SARS-CoV-2 (cycle threshold (Ct) values ranging from 9.96 to 25.59 - average of 18.41) (Table 1), were selected for this study. These samples were collected from 25th March to 25th May 2020, in Juiz de fora (35%), Uberlândia (10%), Uberaba (5%), Santos Dumont (5%), Pouso Alegre (10%), Formiga (5%), Extrema (10%), Cambuí (15%), and Barbacena (5%) (Fig. 1). As shown in Fig. 1, samples analyzed in this work were collected in cities from the southernmost region of MG, an area that shares borders with São Paulo (SP). Due to the geographical proximity, and the fact that SP state is a great economic pole, the mobility of people between these locations is regular. These samples were collected from female (8/20, 40%) and male patients (12/20, 60%) (Table 1). The average age of patients was 53.4 years old.

The 20 selected samples (DNA  $\geq 1$  ng/ $\mu$ L) were used to perform whole genome sequencing using a combination of multiplex PCR amplification (<https://artic.network/ncov-2019>) and Nanopore sequencing. We obtained an average reference coverage of 85.10% related to the reference genome NC\_045512.2. The 20 whole genome sequences of SARS-CoV-2 from samples collected in 9 MG cities generated in this study and 31 sequences from other regions of the state, available at GISAID (<https://www.gisaid.org>), were aligned to compose our dataset resulting in 51 whole genome sequences from samples of 16 cities of MG. Since the incompleteness of the sequences is inevitable, the accuracy of the observed substitutions and frameshift of sequences was validated by manually and carefully verifying the aligned files prior to perform phylogenetic analysis and investigate nucleotide substitutions, as well as analyze protein sequence prediction.

Phylogenetic trees were reconstructed from the 20 whole genome sequences generated in this study and additional 1617 complete genome sequences deposited on GISAID from March 23rd to 25th May 2020 (Fig. 2). Sequences clustered according to lineages A and B, representing the Wuhan/WH04/2020 (EPI\_ISL\_406801) and Wuhan-Hu-1

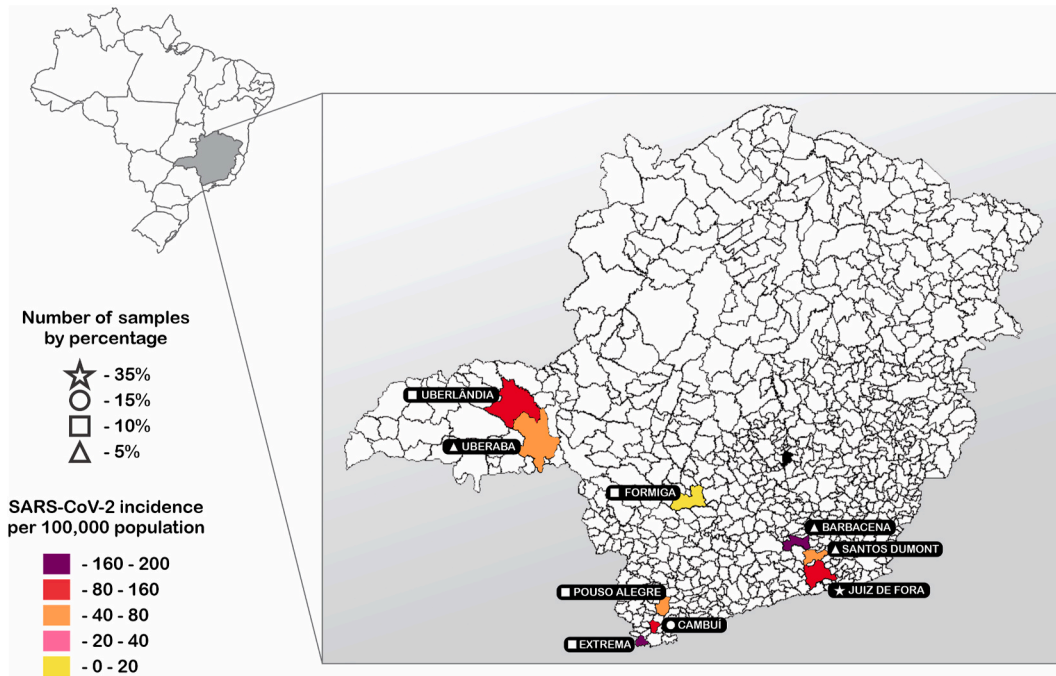


Fig. 1. Area under investigation. Map of the MG state showing the percentage of new SARS-CoV-2 sequences by cities and the incidence per 100,000 population.

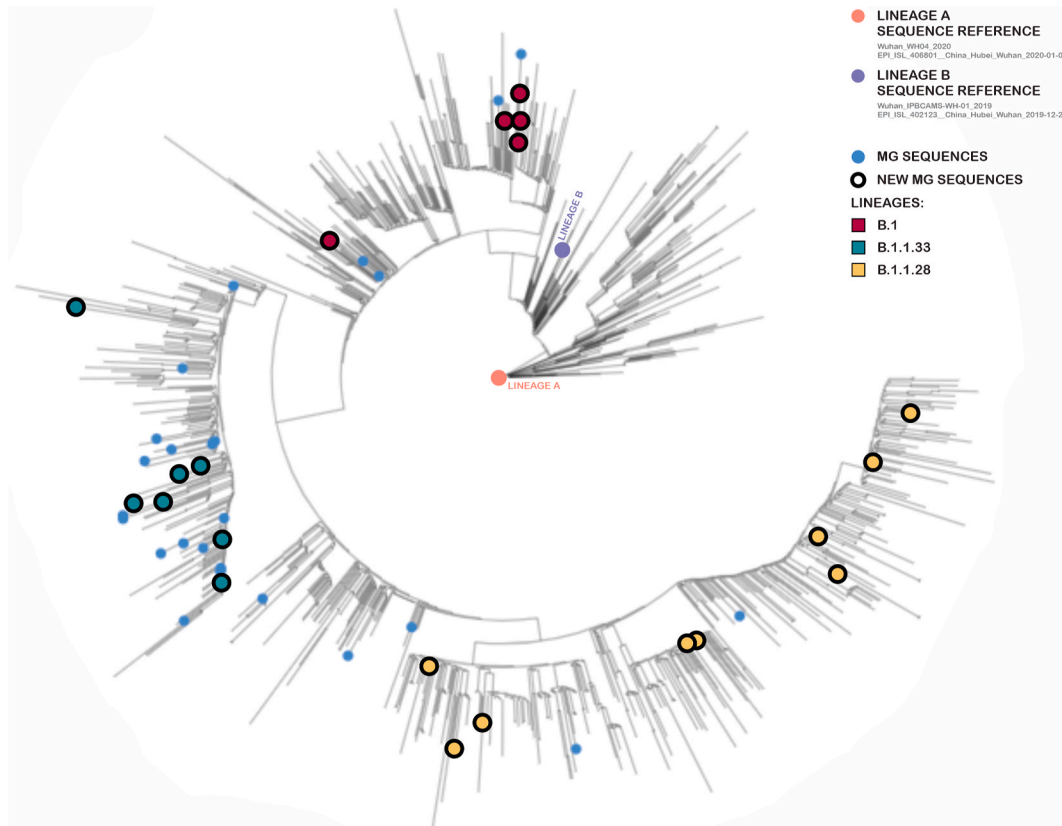


Fig. 2. Phylogenetic tree of the predominant lineages in MG. All samples from MG belong to lineage B, varying between B.1, B.1.1.33 and B.1.1.28. Blue markers represent samples from Minas Gerais; The black circles mark the SARS-CoV-2 sequences generated in this study; The circle along the tree represents the lineages. Orange marker represents the reference sequence of the lineage A (Wuhan/WH04/2020 - EPI\_ISL\_406801) and Purple marker represents the reference sequence of the lineage B (Wuhan-Hu-1 - EPI\_ISL\_402123). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(EPI\_ISL\_402123), respectively, as the recently proposed SARS-CoV-2 lineage nomenclature (Rambaut et al., 2020). Our phylogenetic analysis revealed that sequences from MG grouped to the lineages B.1, B.1.1.28 and B.1.1.33 (Fig. 2). Pangolin analysis also demonstrated the similarity between the sequences generated in this work and those identified in countries as China, USA, Australia, Portugal, United Kingdom and Brazil (Table 1).

Phylogenetic analysis also demonstrated that the sequences generated in this study from samples collected in Cambuí (1), Extrema (1), Formiga (1), Juiz de Fora (3), Pouso Alegre (2) and Barbacena (1) were grouped in clusters with isolates from SP (Fig. S1), suggesting a route of dissemination of the virus between these two states. Additionally, other sequences grouped with isolates from different Brazilian states, as Santa Catarina (SC; 1), Pará (PA; 2) and Rio Grande do Sul (RS; 1), or with another country (France; 1), suggesting independent introductions of the virus in the country. Moreover, sequences from Juiz de Fora (ID: MG 21) and Santos Dumont (ID: MG 19) grouped in the same monophyletic group, which is sustained by the proximity of these cities (approximately 48 km) (Fig. 1). A sequence from Juiz de Fora, generated in this study, also grouped with GISAID dataset sequences from the same locality, suggesting a local circulation of SARS-CoV-2.

The topology of the tree demonstrated that the viral isolates from the 20 generated sequences grouped in 3 main monophyletic groups (Fig. 3). Sequences from Juiz de Fora (ID: MG 42 and MG 21), Uberaba (ID: MG 6) and Santos Dumont (ID: MG 19) (Fig. 4A) were characterized by a nucleotide substitution in the spike protein (A23403G) (Fig. 4B). The cluster with sequences from Juiz de Fora (ID: MG 22 and MG 54), Cambuí (ID: MG 10) and Extrema (ID: MG 18) (Fig. 5A) presented two nucleotide substitutions: one in the ORF6 (T27299C) (Fig. 5B) and another in the nucleoprotein (T29148C) (Fig. 5B). Sequences from Cambuí (ID: MG 9 and MG 10) and Juiz de Fora (ID: MG 22, MG 38, MG 51 and MG 54) (Fig. 6A) revealed three nucleotide substitutions in

nucleocapsid phosphoprotein region (G28881A, G28882A and G28883C) (Fig. 6B). As described in Table 2, some sequences showed nucleotide substitution in viral non-structural proteins (nsps) 2, 3, 4, 7, 8, 12, and in the structural protein Spike (Table 2). Analysis performed using PROVEAN (Protein Variation Effect Analyzer) platform revealed that nucleotide substitutions in spike protein (A23403G) (Fig. 4B), nucleoprotein (T29148C) (Fig. 5B), nsp 3 (C6726T) (Table 2), nsp 8 (C12651T) (Table 2), and nucleocapsid phosphoprotein (G28881A, G28882A and G28883C) (Fig. 6B) are characterized as neutrals. Alternatively, nucleotide substitutions in ORF6 (T27299C) (Fig. 5B), nsp 2 (C920T) (Table 2), and nsp 7 (C12053T) (Table 2) are deleterious.

#### 4. Discussion

This study presents the information on the SARS-CoV-2 variants circulating in nine cities of the southernmost region of Minas Gerais (MG) from March to May 2020. Until 25<sup>th</sup> May 2020, the state reported 6962 cases of COVID-19. Due to its proximity to São Paulo city, and because of the intense industrialization that some Brazilian cities have recently undergone, such as Pouso Alegre and Extrema, this specific MG region has a representative role to understand the progress of SARS-CoV-2 infections in Brazil. Currently, there are very few sequences from this region deposited on GISAID database, compromising the understanding of the pandemic effects in this region of interest. The data presented here were obtained through sequencing of 20 samples of SARS-CoV-2 infections confirmed by RT-qPCR along with others 1617 sequences previously deposited on GISAID [35].

Molecular properties of circulating viral strains have been investigated through genetic analyses and surveillance. SARS-CoV-2 lineages A and B are characterized by Wuhan/WH04/2020 and Wuhan-Hu-1 sequences, respectively. Additionally, lineage B is divided in sublineages [40,41]. The 20 whole genome sequences generated in this study, and

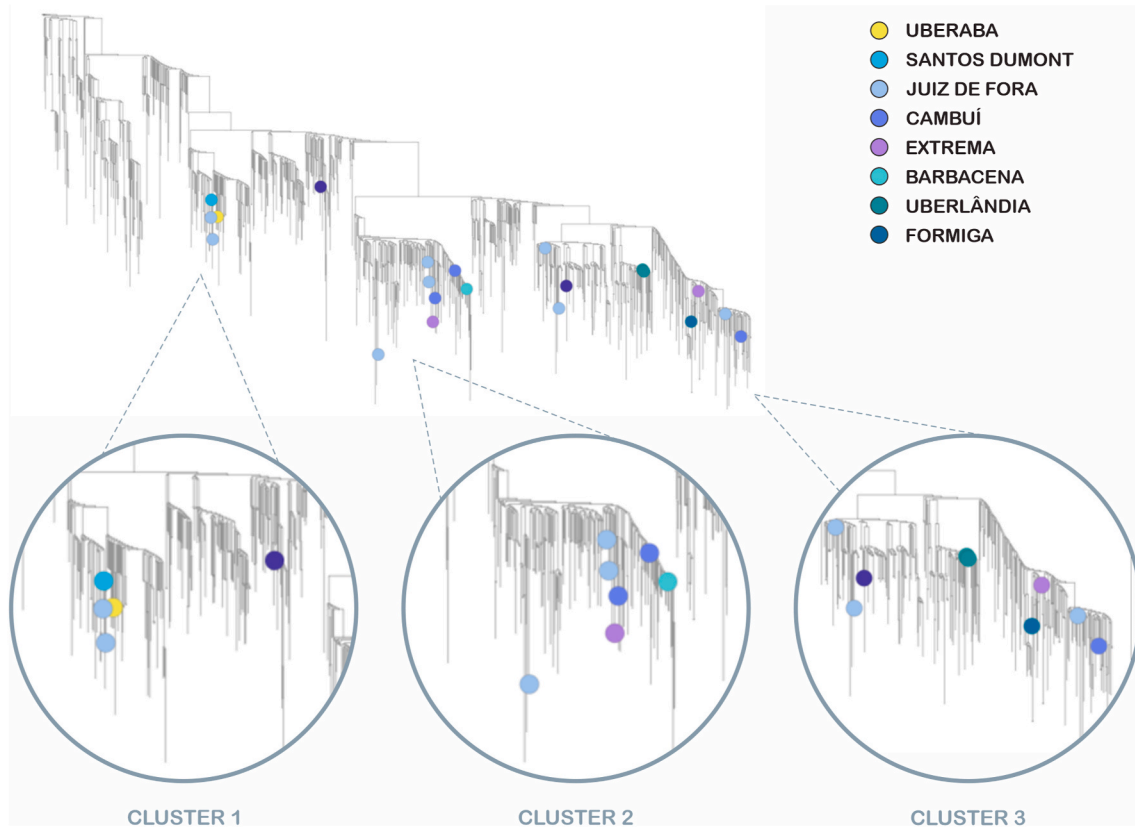


Fig. 3. Phylogenetic tree. The phylogenetic tree demonstrates sequences grouped into monophyletic groups. The colored circles represent the sequences generated in this study and their respective cities.

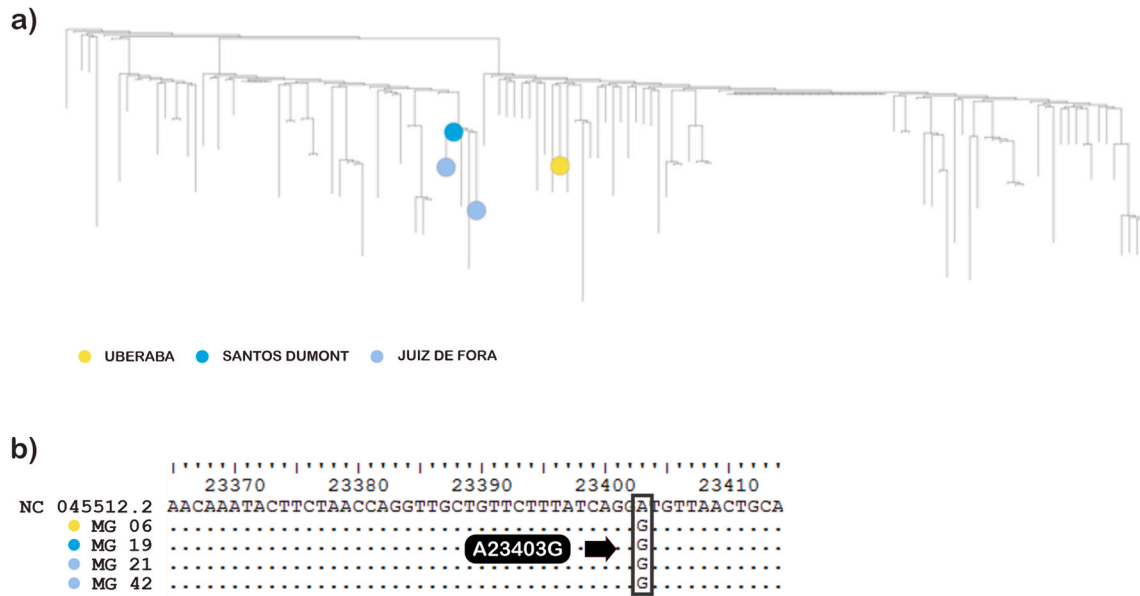


Fig. 4. Nucleotide substitutions. a) Phylogenetic tree representing the sequences from Uberaba, Juiz de Fora and Santos Dumont grouped in a clade. b) alignment of the sequences grouped with the reference sequence, identifying the nucleotide substitution A23403G. The colored circles represent the samples analyzed in this study and their respective cities; The nucleotide substitution regions are marked by the black rectangle; Nucleotide substitutions are indicated by the black text box.

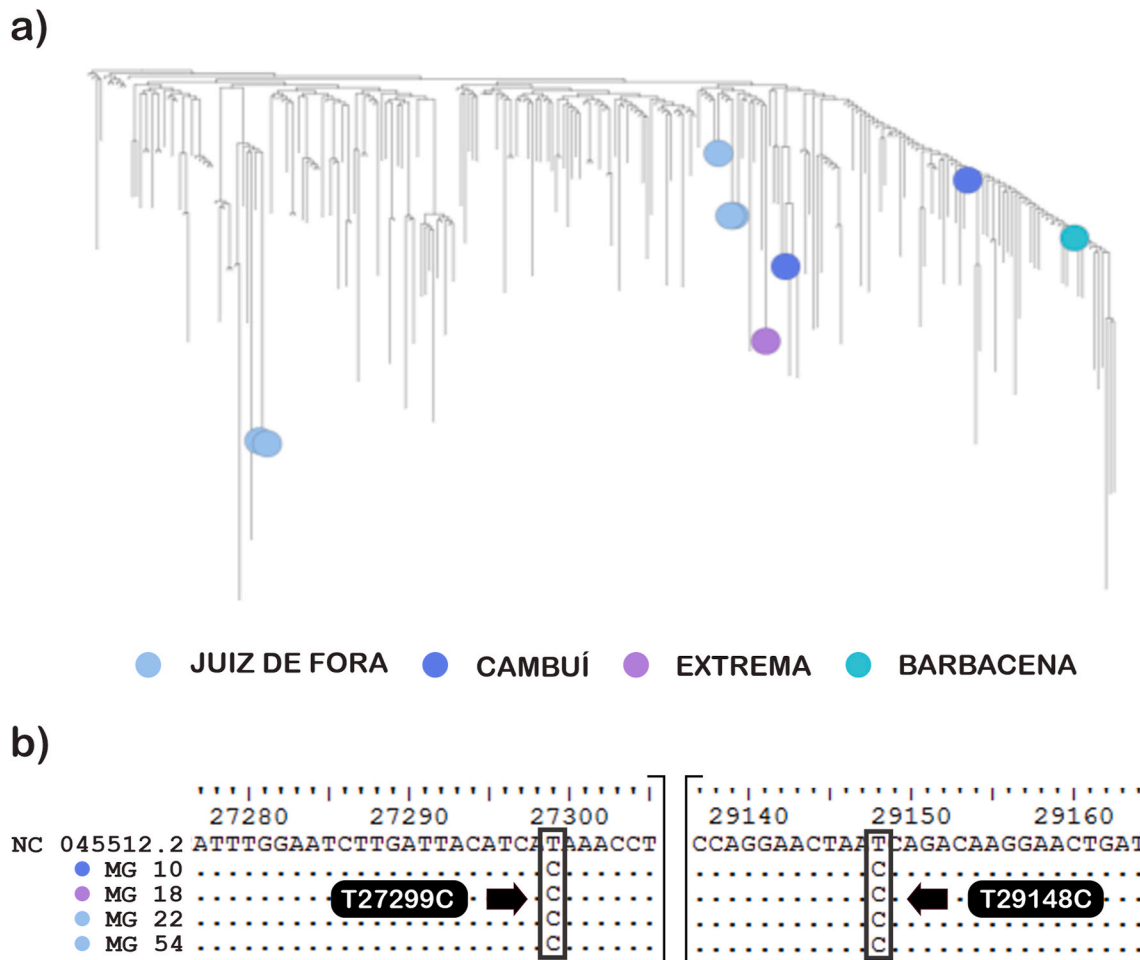
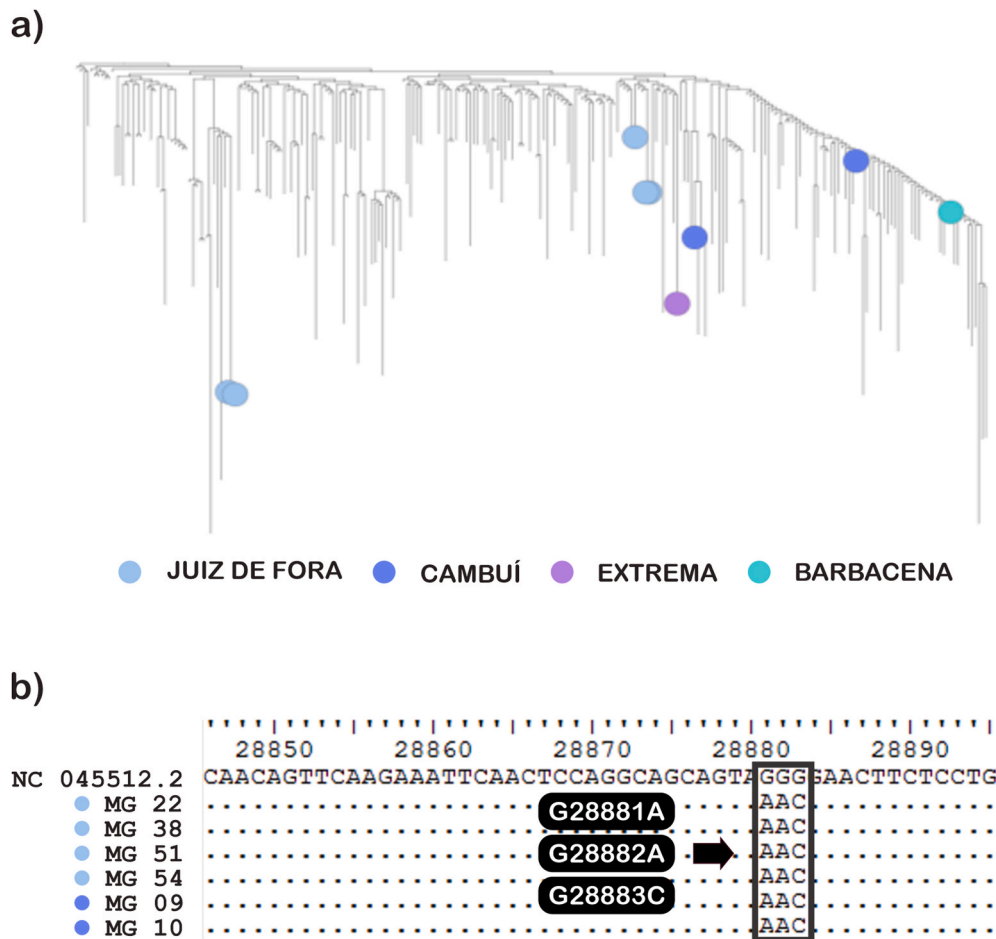


Fig. 5. a) Phylogenetic tree representing the sequences from Juiz de Fora, Cambuí, Extrema and Barbacena grouped in a clade. b) alignment of the sequences grouped with the reference sequence, identifying the nucleotide substitutions T27299C and T29148C. The colored circles represent the samples analyzed in this study and their respective cities; The nucleotide substitution regions are marked by the black rectangle; Nucleotide substitutions are indicated by the black text box.



**Fig. 6.** a) Phylogenetic tree representing the sequences from Juiz de Fora, Cambuí, Extrema and Barbacena grouped in clade. b) alignment of the sequences grouped with the reference sequence, identifying the nucleotide substitutions G28881A, G28882A, and G28883C. The colored circles represent the samples analyzed in this study and their respective cities; The nucleotide substitution regions are marked by the black rectangle; Nucleotide substitutions are indicated by the black text box.

**Table 2**

Characteristics of the nucleotide substitutions predicted by PROVEAN analysis in specific viral genome regions.

Nucleotide Substitution	Product	Samples ID	PROVEAN results
C920T	nsp2	MG 6	Deleterious
C1059T	nsp2	MG 6	
T5804C	nsp3	MG 42	
C6286T	nsp3	MG 8	
C6726T	nsp3	MG 54	Neutral
C8047T	nsp3	MG 20	
C8266T	nsp3	MG 38	
C9967T	nsp4	MG 18	
C12053T	nsp7	MG 7, MG 8, MG 51	Deleterious
C12651T	nsp8	MG 10	Neutral
G14028T	nsp12	MG 8	
A14271G	nsp12	MG 6	
A23403G	S protein	ALL SAMPLES	Neutral
C23422T	S protein	MG 20	
C23683T	S protein	MG 10	
T27299C	ORF 6	MG 10, MG 18	Deleterious
		MG 22, MG 54	
T29148C	Nucleoprotein	MG 10, MG 18, MG 22, MG 54	
G28881A	Nucleocapsid phosphoprotein	MG 22, MG 38, MG 52, MG 54	Neutral
G28882A	Nucleocapsid phosphoprotein	MG 22, MG 38, MG 52, MG 54	Neutral
G28883C	Nucleocapsid phosphoprotein	MG 22, MG 38, MG 52, MG 54	Neutral

additional 1617 from GISAID, composed the dataset of this study. The reconstruction of the phylogenetic tree showed that the 20 generated sequences grouped to the lineage B.1, B.1.1.28, and B.1.1.33 clusters. Our data showed the prevalence of SARS-CoV-2 lineages B.1.1.28 and B.1.1.33 circulation in Minas Gerais state during the period of analysis. The B.1.1.28 strain also circulated in countries such as Australia and United Kingdom, and B.1.1.33 also circulated in the United States [29]. According to Santos and colleagues, this means that, at some point of the pandemic, international introductions of SARS-CoV-2 occurred in Brazil, as well as in Minas Gerais. As an important commercial and technological region, and national and international well-connected state, MG potentially contributed to the viral introduction and spread of SARS-CoV-2 in Brazil [13].

Our results demonstrated that generated sequences from 6 cities of MG (Cambuí, Extrema, Formiga, Juiz de fora, Pouso Alegre, and Barbacena) clustered with sequences of isolates from São Paulo state (SP). At the beginning of the pandemic in Brazil, COVID-19 cases were mainly reported in São Paulo [13], a state that shares border with MG southernmost region. Since some isolates from these two states were grouped in the same monophyletic groups, it suggests a possible transmission route between these municipalities. Our data also showed the clustering of the generated sequences with isolates from other Brazilian states, or other country, and with an isolate from the same location in MG. All these findings support either, the dissemination of the virus among nearby geographical areas or independent introductions from other locations.

Nucleotide substitutions in the genomic regions of the spike protein,



ORF6, nucleoprotein and nucleocapsid phosphoprotein were observed in our analysis. In early 2020, viral strains with substitution in the genomic region of the spike protein were shown to predominantly circulate in SP state [13]. However, substitutions of nucleotides in the ORF6 and nucleoprotein genomic regions were found to be spatially wide spread in Brazil [13]. It suggests that, due to flexibility of social isolations measures and easy access to SP state, the viral strains that most circulated in SP during this period were also detected in the MG cities here studied, and have potentially spread to others states. The substitutions in the nucleocapsid phosphoprotein were described by Koyama and collaborators, who identified the same substitution in 1573 samples globally distribute [42].

SARS-CoV-2 contains open reading frames (ORFs) that encode the four main structural proteins (Spike, Envelope, Nucleocapsid and Membrane) [4]. The Spike protein interacts with cell host receptor controlling viral tissue tropism [43]. For SARS-CoV-2, Spike protein attaches to the host receptor Angiotensin-converting enzyme 2 (ACE-2) [44]. This protein is essential in the early stages of SARS-CoV-2 infection. Therefore, is important to track and investigate variations in the genetic sequences that encode the Spike protein. The nucleotide substitution A23403G in the spike genomic region, that generates the mutation D614G drew increased attention for being detected in genome sequences of SARS-CoV-2 from samples collected around the world [45], being the variant carrying the D614G-Spike protein the most prevalent in the global pandemic [46,47]. As stated by Plante and coworkers, studies in hamsters infected with the G614 variant showed that even though viral titers in the lungs did not increase, they were higher in nasal washes and trachea samples [45]. Clinical evidences demonstrated that D614G mutation enhances viral loads in the upper respiratory tract of COVID-19 patients, resulting in increased viral titers [45]. However, according to their data, sera from these hamsters revealed modestly higher neutralization titers against G614 virus than against D614, highlighting the need of further therapeutic antibodies studies with this viral variation [45]. In this context, Weissman and colleagues, using a pseudotyped virus system, evaluated the G614 mutation and suggested that this substitution increased the epitope exposure, resulting in an enhanced vulnerability to neutralization [39]. Additionally, Zhang and coworkers reported that the G614 spike presents an interaction with cellular receptor that modulates structural rearrangements for membrane fusion, and suggested an improved immunogenic site to neutralization [48]. Alternatively, some studies using animal models compared the neutralizing activity against S(D614) and S(G614) for several candidate vaccines [49–52], and concluded that single-residue mutations hardly change viral sensitivity to neutralization. The author suggest that some alterations in this sense only occur whether the mutation significantly alters S protein conformation [52]. Based on these findings, the mutation D614G possible does not interfere with vaccine efficiency.

Substitutions described in this study were either neutrals or deleterious. Neutral mutations occur when substitutions in the nucleotide or amino acid sequences do not cause loss or alteration in the protein function [53]. In the present work, neutral mutations were found in four different genomic regions (nsp3, nsp8, spike protein and nucleocapsid phosphoprotein). Substitutions in the spike protein have been reported in strains of samples from SP state [13], and demonstrated to be important for viral fitness [45]. Deleterious mutations can either, introduce codons through small frameshift deletions or insertions, and cause nonsense or splice junction alterations, or represent large deletions or duplications. Also, some mutations can compromise the gene function [54]. We observed deleterious mutations in three different genome regions: ORF6, nsp2 and nsp7. Despite being deleterious, the mutations in ORF6 were found sparsely distributed in Brazil [13]. This suggests that these substitutions may be beneficial or do not significantly interfere with the virus infection.

Through beneficial mutations, some SARS-CoV-2 variants become predominant around the globe. These variants are classified as variants of concern (VOC) [55]. To date, five variants were classified as VOC:

Alpha, Beta, Gamma, Delta and Ômicron [55,56]. The VOCs have been associated to higher transmissibility in all age groups, to the severity of the disease, and to increased number of covid-19 cases [57,58]. Genomic surveillance is important to monitor virus evolution and its consequences, in addition to prevent the emergence and spread of new SARS-CoV-2 variants. Although genome sequencing is the gold standard method to detect mutations, it is an expensive and time-consuming technique, making its implementation difficult for all countries. Brazil is an example, from the beginning of the pandemic to the present date, approximately 147,955 Brazilian SARS-CoV-2 genomes have been deposited in the GISAID database [59]. Considering that over 31,3 million cases of COVID-19 have been confirmed, Brazil has a low genomic data performance. This reality makes it even more important to take full advantage of all the data generated and bioinformatics analysis play these important role [55,56]. Through bioinformatic analysis, Wright and collaborators studied mutations that are potential threat based on the impact in human immune response or changes in virus biology caused by phenotypic alteration, including that boosted by vaccines or antiviral drugs [56]. For a better understanding of clinical effects caused by genetic changes undergone by the virus, it is important to correlate genomic and clinical data, being this interface made through bioinformatic analysis [55]. Tracking these mutations has become so important in a pandemic context that researchers have created websites to analyze amino acid replacement and predict the impact in neutralizing activity of monoclonal antibodies (mAbs), convalescent sera, and vaccines [56].

In summary, this work showed the predominance of a circulating lineage in the regions investigated, several nucleotide substitutions in important regions of virus genome, and the phylogenetic relationship between the newly generated sequences and sequences from the global dataset. At the time of the analysis, few substitutions were initially reported in other Brazilian states, which suggested that, over time and with flexibilization of social distance measures, SARS-CoV-2 has spread out to other regions of the country. Here, samples analyzed covered up to 60 days at the beginning of the pandemic in the MG state. Therefore, these findings provide very relevant information to a better understanding of the impacts of the genetic variability of SARS-CoV-2 on the ongoing pandemic, or even on future outbreaks, and reinforce the need for genomic surveillance in the spread of emerging viral pathogens.

## Repositories

GenBank: SAMN18521634, SAMN18521635, SAMN18521636, SAMN18521637, SAMN18521638, SAMN18521639, SAMN18521640, SAMN18521641, SAMN18521642, SAMN18521643, SAMN18521644, SAMN18521645, SAMN18521646, SAMN18521647, SAMN18521648, SAMN18521649, SAMN18521650, SAMN18521651, SAMN18521652, SAMN18521653.

GISAID: EPI\_ISL\_672672, EPI\_ISL\_904028, EPI\_ISL\_904031, EPI\_ISL\_672673, EPI\_ISL\_904023, EPI\_ISL\_904018, EPI\_ISL\_904020, EPI\_ISL\_904030, EPI\_ISL\_904034, EPI\_ISL\_904021, EPI\_ISL\_904022, EPI\_ISL\_904032, EPI\_ISL\_904029, EPI\_ISL\_904026, EPI\_ISL\_672674, EPI\_ISL\_904033, EPI\_ISL\_904019, EPI\_ISL\_904027, EPI\_ISL\_904024, EPI\_ISL\_904025.

## Ethical approval

Ethical approval for this study was obtained from the National Ethical Review Board with approval number CAAE 30127020.0.0000.0068.

## Funding information

Funding was provided by the Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0). FAPESP further supports IMC (2018/

17176-8 and 2019/12000-1), JGJ (2018/17176-8 and 2019/12000-1) ECR (88887595690/2020-00). DSC is supported by Claredon Fund and by the Department of Zoology, University of Oxford. Coordination for the Improvement of Higher Education Personnel (CAPES) – Brazil – Prevention and Combat of Outbreaks, Endemics, Epidemics and Pandemics Finance Code #88881.506794/2020-01 and CAPES – Finance code 001. VRG and GMF received the PhD scholarship (# 88887.505971/2020-00 and # 88887.571465/2020-00) from CAPES.

### Author statement

Giulia Magalhães Ferreira is Bachelor of Biomedical Science, with experience in Virology field. Has a master's degree in Immunology and Parasitology. Currently, she is a PhD student in Immunology and Parasitology by Federal University of Uberlândia-UFU, under the supervision of Prof. A. C. G. Jardim (Federal University of Uberlândia – UFU, Brazil). Her research interests now involve the genomic surveillance of SARS-CoV-2 variants.

### Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Acknowledgements

We are also grateful to authors from the originating laboratories, responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID Initiative, on which this research is based ([Supplementary Material](#)).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biologicals.2022.08.001>.

### References

- Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579:265–9. <https://doi.org/10.1038/s41586-020-2008-3>.
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- Alexander E. Gorbalenya. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;5:536–44. <https://doi.org/10.1038/s41564-020-0695-z>.
- Neuman BW, Adair BD, Yoshioka C, Quispe JD, Orca G, Kuhn P, et al. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J Virol* 2006;80:7918–28. <https://doi.org/10.1128/JVI.00645-06>.
- Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5:1403–7. <https://doi.org/10.1038/s41564-020-0770-5>.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33. <https://doi.org/10.1056/NEJMoa2001017>.
- WHO. Novel coronavirus (2019-nCoV) situation report - 1, 21 January 2020. 2020.
- WHO. Coronavirus disease (COVID-19) situation report - 103, 02 May 2020. 2020.
- Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* 2020;26:506–10. <https://doi.org/10.1038/s41591-020-0822-7>.
- Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 2020;20:669–77. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7).
- Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship. *Euro Surveill* February 2020;25. <https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256>. 2020.
- Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 2020;80(369):1255–60. <https://doi.org/10.1126/science.abd2161>.
- Santos I de A, Grosche VR, Sabino-Silva R, Jardim ACG. Antivirals against Coronaviruses: candidate drugs for the SARS-CoV-2 treatment? *Front Microbiol* 2020;11:1818. <https://doi.org/10.3389/FMICB.2020.01818>.
- Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N Engl J Med* 2020;382:2081–90. <https://doi.org/10.1056/NEJMoa2008457>.
- Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship. In: *Eurosurveillance* 2020. Yokohama: Japan; 2020. p. 25. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180>.
- Damasceno DHP, Amaral AA, Silva CA, Simões e Silva AC. The impact of vaccination worldwide on SARS-CoV-2 infection: a Review on vaccine mechanisms, results of clinical trials, vaccinal coverage and interactions with novel variants. *Curr Med Chem* 2022;29:2673–90. <https://doi.org/10.2174/0929867328666210902094254>.
- Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, Hasell J, Macdonald B, Dattani S, Roser M, Bobbie MacDonald DB and MR. Coronavirus pandemic (COVID-19). *Our World Data*. 2020.
- Mohamed K, Rzymiski P, Islam MS, Makuku R, Mushtaq A, Khan A, et al. COVID-19 vaccinations: the unknowns, challenges, and hopes. *J Med Virol* 2022;94:1336–49. <https://doi.org/10.1002/jmv.27487>.
- Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, Hasell J, Macdonald B, Dattani S, Roser M, Bobbie MacDonald DB and MR. Coronavirus pandemic (COVID-19). [https://ourworldindata.org/covid-vaccinations?country=OWID\\_WRL](https://ourworldindata.org/covid-vaccinations?country=OWID_WRL). [Accessed 31 May 2022].
- COVID-19 treatment guidelines (NIH US). n.d. <https://www.covid19treatmentguidelines.nih.gov/therapies/antiviral-therapy/summary-recommendations/#:~:text=Remdesivir is the only drug,the treatment of COVID-19.> [Accessed 14 June 2022].
- Food and Drug Administration. Fact sheet for healthcare providers: emergency use authorization for molnupiravir. n.d. <https://www.fda.gov/media/155054/download>. [Accessed 14 June 2020].
- Food and Drug Administration. Fact sheet for healthcare providers: emergency use authorization for Paxlovid. n.d. <https://www.fda.gov/media/155050/download>. [Accessed 14 June 2022].
- Anvisa [n.d.].
- Anvisa. n.d.(accessed June 14, 2022), <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2022/anvisa-aprova-uso-emergencial-do-medicamento-molnupiravir-para-covid-19#:~:text=O molnupiravir é um medicamento,casos de hospitalização e mortes>.
- Siddiqi HK, Mehra MR. COVID-19 illness in native and immunosuppressed states: a clinical–therapeutic staging proposal. *J Heart Lung Transplant* 2020;39:405–7. <https://doi.org/10.1016/j.healun.2020.03.012>.
- No Title n.d. <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003>.
- Jesus JG de, Sacchi C, Candido D da S, Claro IM, Sales PCS, Manuli ER, et al. Importation and early local transmission of COVID-19 in Brazil, 2020, vol. 62. *Rev Inst Med Trop Sao Paulo*; 2020. <https://doi.org/10.1590/s1678-99462020062030>.
- Instituto Brasileiro de Geografia e Estatística - IBGE. Cidades e Estados: Minas Gerais [n.d.].
- Xavier J, Giovanetti M, Adelino T, Fonseca V, Barbosa da Costa AV, Ribeiro AA, et al. The ongoing COVID-19 epidemic in Minas Gerais, Brazil: insights from epidemiological data and SARS-CoV-2 whole genome sequencing. *Emerg Microb Infect* 2020;9:1824–34. <https://doi.org/10.1080/22221751.2020.1803146>.
- Brasil M da, SS de V em S. *Boletim epidemiológico - Minas Gerais*. 2020.
- Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* 2017;12:1261–76. <https://doi.org/10.1038/nprot.2017.066>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
- Phillips JL, Gnanakaran S. BioEdit: an important software for molecular biology. *Proteins: Struct, Funct, Bioinf* 2015;83:46–65. <https://doi.org/10.1002/prot.24632>.
- Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. 2014. p. 131–46. <https://doi.org/10.1007/978-1-62703-646-7-8>.
- Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *Euro Surveill* 2017;22. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
- Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Boletim Epidemiológico de Minas Gerais [n.d.].
- Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 1994. <https://doi.org/10.1007/BF00160154>.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020. <https://doi.org/10.1093/molbev/msaa015>.
- Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015. <https://doi.org/10.1093/bioinformatics/btv195>.

- [40] Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020;80. <https://doi.org/10.1126/science.abb4218>.
- [41] Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* 2020;80. <https://doi.org/10.1126/science.abb6105>.
- [42] Koyama T, Platt D, Parida L. Variant analysis of SARS-cov-2 genomes. *Bull World Health Organ* 2020. <https://doi.org/10.2471/BLT.20.253591>.
- [43] Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL, et al. Pre-fusion structure of a human coronavirus spike protein. *Nature* 2016. <https://doi.org/10.1038/nature17200>.
- [44] Hussain M, Jabeen N, Raza F, Shabbir S, Baig AA, Amanullah A, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J Med Virol* 2020. <https://doi.org/10.1002/jmv.25832>.
- [45] Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 2020. <https://doi.org/10.1038/s41586-020-2895-3>.
- [46] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182:812–27. <https://doi.org/10.1016/j.cell.2020.06.043>. e19.
- [47] Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M, Dinnon KH, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* 2020;80(370):1464–8. <https://doi.org/10.1126/science.abe8499>.
- [48] Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* 2021;80. <https://doi.org/10.1126/science.abf2303>.
- [49] Corbett KS, Edwards DK, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* 2020;586:567–71. <https://doi.org/10.1038/s41586-020-2622-0>.
- [50] McAuley AJ, Kuiper MJ, Durr PA, Bruce MP, Barr J, Todd S, et al. Experimental and in silico evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. *Npj Vaccines* 2020;5:96. <https://doi.org/10.1038/s41541-020-00246-8>.
- [51] Ami Patel, Jewell Walters, Emma L. Reuschel, Katherine Schultheis, Elizabeth Parzych, Ebony N. Gary, Igor Maricic, Mansi Purwar, Zeena Eblimit, Susanne N. Walker, Diana Guimet, Pratik Bhojnarwal, Arthur Doan, Ziyang Xu, Dustin Elwood, Sophia M. Reeder, KEB. Intradermal-delivered DNA vaccine provides anamnestic protection in a rhesus macaque SARS-CoV-2 challenge model n.d. <https://doi.org/10.1101/2020.07.28.225649>. 2007.2028.225649.
- [52] Jackson CB, Zhang L, Farzan M, Choe H. Functional importance of the D614G mutation in the SARS-CoV-2 spike protein. *Biochem Biophys Res Commun* 2021; 538:108–15. <https://doi.org/10.1016/j.bbrc.2020.11.026>.
- [53] Kishimoto T, Ying BW, Tsuru S, Iijima L, Suzuki S, Hashimoto T, et al. Molecular clock of neutral mutations in a fitness-increasing evolutionary process. *PLoS Genet* 2015. <https://doi.org/10.1371/journal.pgen.1005392>.
- [54] Borg Å, Haile RW, Malone KE, Capanu M, Diep A, Törngren T, et al. Characterization of BRCA1 and BRCA2 deleterious mutations and variants of unknown clinical significance in unilateral and bilateral breast cancer: the WECARE study. *Hum Mutat* 2010. <https://doi.org/10.1002/humu.21202>.
- [55] Alam I, Radovanovic A, Incitti R, Kamau AA, Alarawi M, Azhar EI, et al. CovMT: an interactive SARS-CoV-2 mutation tracker, with a focus on critical variant n.d. [https://doi.org/10.1016/S1473-3099\(21\)00078-5](https://doi.org/10.1016/S1473-3099(21)00078-5).
- [56] Wright DW, Harvey WT, Hughes J, Cox M, Peacock TP, Colquhoun R, et al. Tracking SARS-CoV-2 mutations & variants through the COG-UK-mutation explorer. *Virus Evol* 2022. <https://doi.org/10.1093/ve/veac023>.
- [57] Wang P, Wang M, Yu J, Cerutti G, Nair MS, Huang Y, et al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *BioRxiv Prepr Serv Biol* 2021: 2021. <https://doi.org/10.1101/2021.03.01.433466>. 03.01.433466.
- [58] Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021;80. <https://doi.org/10.1126/science.abg3055>.
- [59] GISAID - Initiative [n.d].