



Seroprevalence, Prevalence, and Genomic Surveillance: Monitoring the Initial Phases of the SARS-CoV-2 Pandemic in Betim, Brazil

Ana Valesca Fernandes Gilson Silva^{1*†}, Diego Menezes^{2,3†},
Filiipe Romero Rebello Moreira^{4†}, Octávio Alcântara Torres¹,
Paula Luize Camargos Fonseca^{2,3}, Rennan Garcias Moreira⁵, Hugo José Alves^{2,3},
Vivian Ribeiro Alves¹, Tânia Maria de Resende Amaral¹, Adriano Neves Coelho¹,
Júlia Maria Saraiva Duarte³, Augusto Viana da Rocha¹,
Luiz Gonzaga Paula de Almeida⁶, João Locke Ferreira de Araújo^{2,3},
Hilton Soares de Oliveira¹, Nova Jersey Cláudio de Oliveira¹, Camila Zolini⁴,
Jôsy Hubner de Sousa⁷, Elizângela Gonçalves de Souza¹, Rafael Marques de Souza^{2,3},
Luciana de Lima Ferreira^{2,3}, Alexandra Lehmkuhl Gerber⁶,
Ana Paula de Campos Guimarães⁶, Paulo Henrique Silva Maia¹,
Fernanda Martins Marim^{2,3}, Lucyene Migueta⁸, Cristiane Campos Monteiro¹,
Tuffi Saliba Neto¹, Fabricia Soares Freire Pugêdo¹, Daniel Costa Queiroz^{2,3},
Damares Nígia Alborgueti Cuzzuol Queiroz¹, Luciana Cunha Resende-Moreira⁹,
Franciele Martins Santos⁷, Erika Fernanda Carlos Souza¹, Carolina Moreira Voloch⁴,
Ana Tereza Vasconcelos⁶, Renato Santana de Aguiar^{2,3,10*} and Renan Pedra de Souza^{2,3*}

OPEN ACCESS

Edited by:

M. Pilar Francino,
Fundación para el Fomento de la
Investigación Sanitaria y Biomédica
de la Comunitat Valenciana (FISABIO),
Spain

Reviewed by:

Adriano de Bernardi Schneider,
University of California, Santa Cruz,
United States
Juan Angel Patino Galindo,
Columbia University, United States

*Correspondence:

Ana Valesca Fernandes Gilson
Silva
anavalescafemandes@hotmail.com
Renato Santana de Aguiar
santanarnt@gmail.com
Renan Pedra de Souza
renanpedra@gmail.com

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Infectious Agents and Disease,
a section of the journal
Frontiers in Microbiology

Received: 22 October 2021

Accepted: 07 January 2022

Published: 07 February 2022

¹ Escola de Saúde Pública de Betim, Betim, Brazil, ² Programa de Pós Graduação em Genética, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ³ Laboratório de Biologia Integrativa, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁴ Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, ⁵ Centro de Laboratórios Multiusuários, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁶ Laboratório Nacional de Computação Científica, Petrópolis, Brazil, ⁷ Programa de Pós-graduação em Biologia Celular, Departamento de Morfologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁸ Departamento de Patologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁹ Departamento de Botânica, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ¹⁰ Instituto D'Or de Pesquisa e Ensino (IDOR), Rio de Janeiro, Brazil

The COVID-19 pandemic has created an unprecedented need for epidemiological monitoring using diverse strategies. We conducted a project combining prevalence, seroprevalence, and genomic surveillance approaches to describe the initial pandemic stages in Betim City, Brazil. We collected 3239 subjects in a population-based age-, sex- and neighborhood-stratified, household, prospective; cross-sectional study divided into three surveys 21 days apart sampling the same geographical area. In the first survey, overall prevalence (participants positive in serological or molecular tests) reached 0.46% (90% CI 0.12–0.80%), followed by 2.69% (90% CI 1.88–3.49%) in the second survey and 6.67% (90% CI 5.42–7.92%) in the third. The underreporting reached 11, 19.6, and 20.4 times in each survey. We observed increased odds to test positive in females compared to males (OR 1.88 95% CI 1.25–2.82), while the single best predictor for positivity was ageusia/anosmia (OR 8.12, 95% CI 4.72–13.98). Thirty-five SARS-CoV-2

genomes were sequenced, of which 18 were classified as lineage B.1.1.28, while 17 were B.1.1.33. Multiple independent viral introductions were observed. Integration of multiple epidemiological strategies was able to adequately describe COVID-19 dispersion in the city. Presented results have helped local government authorities to guide pandemic management.

Keywords: COVID-19, molecular epidemiology, epidemiology, whole genome sequencing, SARS-CoV-2 variant

INTRODUCTION

Since its emergence in December 2019, the new human coronavirus has had a tremendous impact on humanity due to the pandemic nature of its infection, called COVID-19 (Zhou et al., 2020). The SARS-CoV-2 pathogen was described on January 24, 2020. In Brazil, the first case of COVID-19 was reported on February 26, 2020, in the city of São Paulo (Araujo et al., 2020). The virus spread rapidly, and the country had the highest number of cases and deaths in Latin America, experiencing its first peak wave in late July 2020. Although most patients were identified in the most prominent Brazilian cities, São Paulo and Rio de Janeiro, dispersion to other municipalities was quickly reported. Betim, a town located in the Minas Gerais State in Brazil with an estimated population of 439,340 in 2019, had its first reported SARS-CoV-2 case on March 23, 2020, in two patients returning from Europe. Two months later, on May 23, 2020, only 73 confirmed cases had been reported, although 4380 suspected cases were identified in public databases indicating limited testing availability.

Brazilian public healthcare system has prioritized testing subjects with symptoms due to scarce diagnostic tests, particularly in the early days of the pandemic. Since data suggest that symptomatic cases represent a fraction of persons infected with SARS-CoV-2, official statistics were expected to be underestimated (Wu et al., 2020). Several aspects may influence COVID-19 symptom presentation (Araújo et al., 2021; Rossi et al., 2021). Epidemiological surveillance using prevalence studies is needed to evaluate the true extent of SARS-CoV-2 dispersion, significantly extending testing to asymptomatic subjects. Combining serological and molecular tests may be a more robust strategy to uncover viral diffusion in a territory, avoiding each test's kinetic detection limitations. Valid prevalence and seroprevalence estimates for a population rely on two major factors: (i) a representative population sample and (ii) accurate diagnostic testing (Byambasuren et al., 2021).

While the epidemiological investigation is essential for controlling COVID-19, genomic surveillance is crucial. Robust SARS-CoV-2 variant monitoring can track viral evolution, detect new variants, describe patterns and clusters of transmission, outbreak tracking, among others. Therefore, it can provide actionable information on implementing a more targeted public health strategy that addresses local priorities through stakeholder engagement and mitigation efforts (Robishaw et al., 2021). We conducted a study combining seroprevalence, prevalence, and genomic surveillance approaches to understand the SARS-CoV-2 epidemic spread in Betim city.

MATERIALS AND METHODS

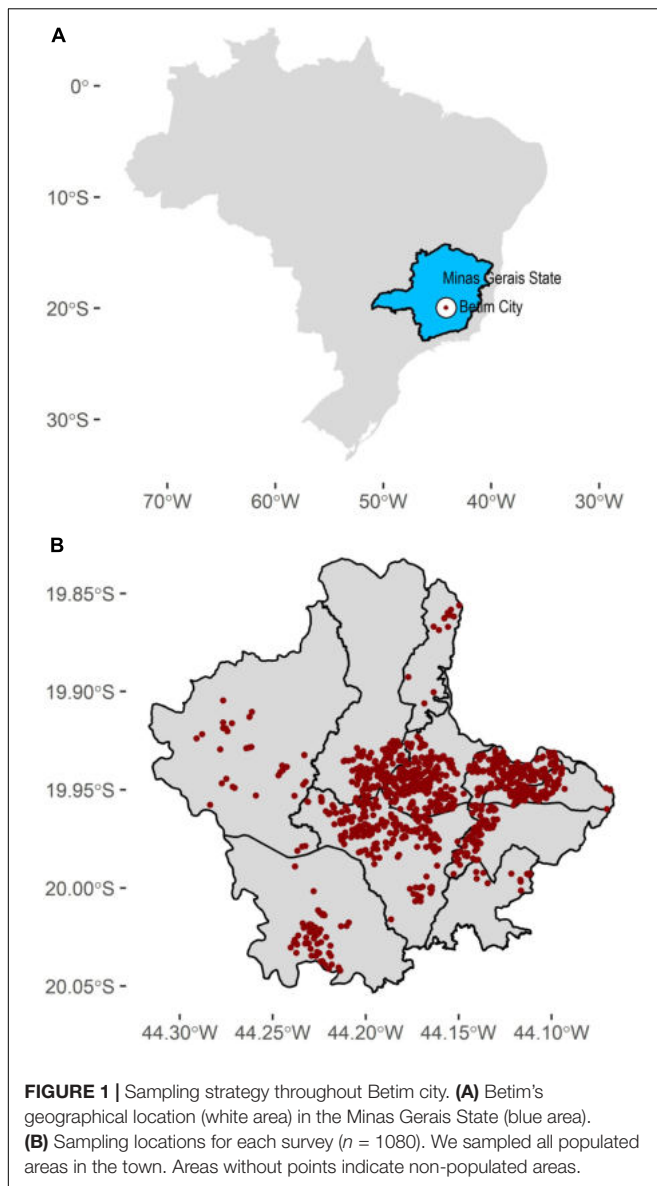
Seroprevalence and Prevalence

The Research Ethics Committee approved the present experiment under protocol CAAE 31459220.2.0000.5651. We conducted a population-based age-, sex- and neighborhood-stratified, household, prospective; cross-sectional study repeated every 21 days in the same geographic area to determine the extent of SARS-CoV-2 transmission in Betim, Minas Gerais, Brazil (Figures 1A,B). All populated areas in the city were sampled. Three surveys were held: June 3–5, June 23–25, and July 13–15, 2020. The sample size ($n = 1,080$ each survey) was estimated considering dichotomous outcome (positive or negative), the population of 439,340 inhabitants, the confidence level of 90%, the maximum margin of error of 2.5%, and lack of *a priori* information on the prevalence of SARS-CoV-2 in the municipality's population (the latter represented by $p = q = 0.5$) and using the equation below:

$$n = \frac{Z_{\frac{\alpha}{2}}^2 \cdot \hat{p} \cdot \hat{q} \cdot N}{E^2 \cdot (N - 1) + Z_{\frac{\alpha}{2}}^2 \cdot \hat{p} \cdot \hat{q}}$$

Random sampling was employed to ensure representativeness of the population, stratified by sex, age (0–5; 6–19; 20–39; 40–59, and 60 years or older) and city neighborhoods (Centro, Alterosas, Imbiruçu, Norte, Teresópolis, PTB, Citrolândia, Vianópolis, Icaivera, and Petrovale). Every census tract (population stratum created by Governmental agencies) was sampled with at least one address. In case of refusal or closed households, the closest home was selected. Thirty-six teams (one driver, one nurse, and one community health worker) worked on active sampling subjects in 1080 addresses during 3 days. Clinical and epidemiological data were obtained using a questionnaire during interviews with participants or their legal guardians who signed the Informed Consent. Biological samples were collected using a nasal swab to conduct RT-PCR and capillary blood obtained by fingerstick for the serological test.

RT-PCR to detect SARS-CoV-2 RNA was initially conducted in pools of 10 samples (Lohse et al., 2020). Whenever pools were positive, individual samples were examined. Molecular diagnosis was established according to the CDC 2019–Novel Coronavirus Real-Time RT-PCR Diagnostic Panel (N1, N2, and RNP primers). Serological tests were conducted using the SARS-CoV-2 Antibody Test (Guangzhou Wondfo Biotech Co., Ltd. Guangzhou, Guangdong, China) that detects IgM/IgG



antibodies. The same test was used in a previous study in Brazil (Hallal et al., 2020). Reported sensitivity is 86.43% (95% CI: 82.41~89.58%) and specificity 99.57% (95% CI: 97.63~99.92%). We have validated antibody tests using serum samples from subjects who were SARS-CoV-2 positive confirmed with RT-PCR.

Associations of each variable of interest with surveys and positive status were assessed using chi-square tests. Odds ratios were estimated using logistic regression with the *glm* function. Spatial geostatistical modeling and prediction were carried out using the *gstat* and *predict* functions from the *gstat* package. All analyses were carried out in R software (version 4.1.1).

Genomic Surveillance

Whole viral genome amplification and DNA library preparation was carried out as described elsewhere (Moreira et al., 2021). Briefly, QIAseq SARS-CoV-2 Primer Panel—QIAGEN kit was

used to amplify positive samples, following manufacturer instructions. In total, 39 of the 84 detectable samples were eligible for library preparation based on their CTs ≤ 30 . Library concentration was measured using the QIAseq Library Quant Assay—QIAGEN kit, and the fragment integrity and size were evaluated using Bioanalyzer (Agilent Technologies, Waldbronn, DE). Sequencing was carried out on a MiSeq (Illumina, San Diego, CA, United States).

The raw data generated were filtered by Trimmomatic v0.39 (Bolger et al., 2014), which trimmed low-quality bases (Phred score < 30) and removed short reads (< 50 nucleotides) as well as adapters and primer sequences. Reads were then mapped against the SARS-CoV-2 reference genome (accession number: NC_045512.2) with Bowtie2 (Langmead et al., 2009). The resulting BAM files were manipulated with SAMtools, BCFtools (Li et al., 2009), and BEDtools (Quinlan and Hall, 2010) to generate consensus genome sequences. Bases with less than 10x sequencing depth were masked. In total, 35 of the 39 genome sequences presented coverage greater than 79% and average sequencing depth greater than 200x (**Supplementary Table 1**). The 35 consensus genome sequences were submitted to the PANGOLIN 2.0 lineage classification tool (database version February 2, 2021) (Rambaut et al., 2020).

To confirm the PANGOLIN identification and further contextualize the diversity of lineages circulating in Betim, we performed a set of phylogenetic analyses. First, a global dataset was assembled from a subset of high-quality data available on GISAID and the newly generated genomes ($n = 3,814$). This dataset contained all Brazilian sequences and one per week for each country, as available on GISAID until January 12, 2021. These sequences were aligned with MAFFT v7.475 (Katoh and Standley, 2013), and a maximum likelihood tree was inferred on IQ-Tree 2 (Minh et al., 2020), under the GTR+F+I+G4 model (Tavaré, 1986; Yang, 1994). Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT) was used to assess branches' statistical support (Guindon et al., 2010).

Two subsets of the previous dataset were assembled to explore the temporal dynamics of introduction and circulation of SARS-CoV-2 in Betim, comprehending sequences belonging to lineages B.1.1.28 ($n = 258$) and B.1.1.33 ($n = 284$). The parameterization of the phylogeographic model was set to be primarily informative concerning introductions of SARS-CoV-2 in Betim. Therefore, we set the model with six discrete categories: Betim City, Minas Gerais State, Rio de Janeiro State, São Paulo State, other Brazilian States, and foreign sequences. These locations were represented by 18, 2, 22, 71, 79, and 66 sequences in dataset B.1.1.28 while B.1.1.33 dataset composition was 17, 20, 53, 52, 73, and 69 sequences from each region, respectively.

Maximum likelihood trees were inferred from these datasets, and their temporal signal was evaluated with *tempest* v1.5.3 (Rambaut et al., 2016). Time scaled phylogenies were then inferred from these datasets with BEAST v1.10.4 (Suchard et al., 2018), using: (i) the HKY+I+G4 nucleotide substitution model (Yang, 1994), (ii) the strict molecular clock model, (iii) the non-parametric coalescent skygrid tree prior (Gill et al., 2013) and (iv) a symmetric discrete phylogeographic model (Lemey et al., 2009). A normal prior distribution (mean = 1.13×10^{-3} ; std = $5.1 \times$

10^{-4}) on clock rate was assumed, based on a previous estimate (Candido et al., 2020). The cutoff values of the skygrid tree prior were set based on the previously estimated dates for the emergence of each lineage (Candido et al., 2020). The number of grids of the tree priors was set to match the approximate number of weeks comprehended between the estimated dates for lineages' emergence and the dates of the most recently sampled sequences (41 weeks, both datasets). Two and three independent chains of 200 million generations sampling every 10,000 states were performed for datasets B.1.1.33 and B.1.1.28, respectively. Tracer v1.7.1 (Rambaut et al., 2018) was used to verify mixing and convergence of chains (effective sample size > 200 for all parameters), which were then combined with logcombiner v1.10.4 after 10% burning removal. Maximum clade credibility trees were generated with treeannotator v1.10.4. All logs and trees are available in https://github.com/LBI-lab/SARS-CoV-2_phylogenies.git.

RESULTS

Seroprevalence and Prevalence

Evaluation of clinical and epidemiological data showed no significant difference for the presence of any prior health condition across surveys (pneumopathy, chronic neurological disease, pregnant, postpartum, chronic cardiovascular disease, chronic kidney disease, obesity, asthma, immunodepression, chronic liver disease, diabetes, hypertension, transplanted, cancer or any comorbidity) indicating proper sampling was conducted since there was no reason to find significant differences in the period (Table 1). Four symptoms (cough, sore throat, myalgia, and rhinorrhea) and contact with a symptomatic person increased while international travel decreased. Prevalence and seroprevalence increased across surveys.

Sampling was conducted in the early stages of the pandemic, as seen in the number of absolute reported cases (Figure 2A). Cumulative confirmed cases were underestimated (Figure 2B). In the first survey, overall prevalence (participants positive in serological or molecular tests) reached 0.46% (90% CI 0.12–0.80%), followed by 2.69% (90% CI 1.88–3.49%) in the second survey and 6.67% (90% CI 5.42–7.92%) in the third. The underreporting was obtained by the difference between survey prevalence and official data, and its magnitude reached 11, 19.6, and 20.4 times (Figure 2B). Overall prevalence increase was observed across most administrative regions (Figures 2C,D). Active transmission areas (RT-PCR positive participants) were observed increasing across time (Figures 3A–C). By the third survey, almost all populated city areas were likely to have viral circulation (Figure 3C).

We have also evaluated whether clinical and epidemiological variables were associated with molecular or serological test positivity (Table 2). Several significant results were observed, mostly with reported symptoms (fever, cough, sore throat, dyspnoea, myalgia, rhinorrhea, respiratory discomfort, nausea/vomit, headache, prostration, ageusia/ anosmia). We also observed increased odds to test positive in females compared to males (OR 1.88 95% CI 1.25–2.82) and clear enrichment

of positive cases in certain city regions (e.g., Imbiruçu and Terezópolis). Surprisingly, people with obesity were more likely to be positive (OR 3.33, 95% CI 1.68–6.59). The single best predictor for positivity was ageusia/anosmia (OR 8.12, 95% CI 4.72–13.98). Non-significant associations were also found (Supplementary Table 2).

Genomic Viral Surveillance

In total, 35 novel SARS-CoV-2 genome sequences were obtained (GISAID EPI_ISL_5416087–5416121). The sequences were classified by PANGOLIN 2.0 to assess the genetic diversity of SARS-CoV-2 circulating in Betim. 18 of the 35 genomes were classified as lineage B.1.1.28, while 17 were B.1.1.33 (Probability = 1.0). Further, a maximum likelihood tree was inferred from the global dataset GISAID (Shu and McCauley, 2017). No difference in the dispersion pattern was observed across lineages.

The analysis supported these results, revealing sequences from the Betim cluster within several clades of these lineages confirming the circulation of (B.1.1.28 and B.1.1.33 during the first wave of COVID-19 pandemics in the city (Figure 4). The spread of Betim sequences across the tree suggests multiple independent introductions occurred in the town. Further, eight clades majorly composed by Betim sequences were inferred with variable degrees of statistical support (median SH-aLRT = 82.75, range: 0–100), suggesting the occurrence of local transmission in the city after initial introduction events. In addition to these clusters, nine introductions supported by single sequences have also been detected. Most Betim sequences or clusters are closely related to sequences from Rio de Janeiro and São Paulo, two neighboring States connected by highways to Minas Gerais. To formally assess the dynamics of introduction and spread of SARS-CoV-2 in Betim, separated datasets for lineages B.1.1.28 and B.1.1.33 were evaluated. Regression between sampling times and genetic distances revealed both datasets had moderate temporal signal (B.1.1.28: $R^2 = 0.49$; B.1.1.33: $R^2 = 0.58$), justifying molecular clock analysis.

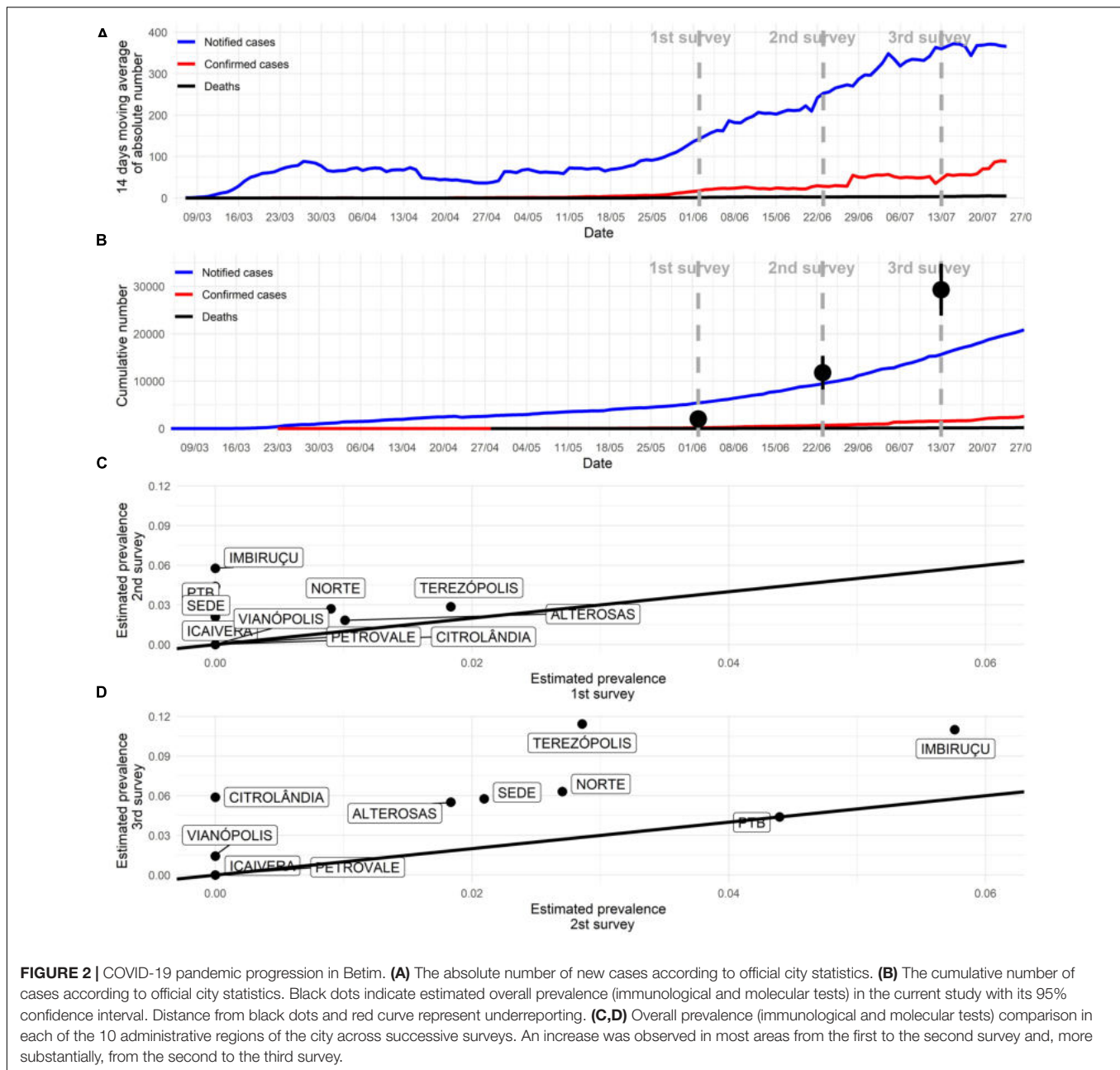
The time-scaled phylogeographic analysis performed with dataset B.1.1.28 suggests this lineage emerged on February 22, 2020, in São Paulo (95% highest posterior density, HPD: February 11, 2020–March 05, 2020; geographic model posterior probability, PP = 0.91), later spreading to other Brazilian states (Figure 5A). The phylogeny reveals that two introduction events, dated between April 19, 2020 (95% HPD: April 17, 2020–May 11, 2020) and April 22, 2020 (95% HPD: April 20, 2020–May 27, 2020), led to the emergence of Betim clusters (harboring between two and six sequences). Additionally, four introductions related to single sequences have been detected. The phylogeographic model suggests that three introductions occurred from another Brazilian region to Betim, in addition to other single events from RJ, another one from SP, and another from foreign sequences. All events presented high statistical support (PP > 92% for all events).

The phylogeographic reconstruction performed for dataset B.1.1.33 infers the origin of this lineage on February 06, 2020, in Rio de Janeiro (95% HPD: January 14, 2020–February 25, 2020, PP = 0.78). The model supports the occurrence of many

TABLE 1 | Clinical and epidemiological data obtained from participants.

Variable	Level	Overall n (%)	First survey n (%)	Second survey n (%)	Third survey n (%)	p-value
Administrative Regions	Alterosas	634 (19.6%)	198 (18.4%)	218 (20.2%)	218 (20.2%)	0.9584
	Citrolândia	219 (6.8%)	83 (7.7%)	68 (6.3%)	68 (6.3%)	
	Icaivera	62 (1.9%)	20 (1.9%)	21 (1.9%)	21 (1.9%)	
	Imbiruçu	565 (17.4%)	183 (17.0%)	191 (17.7%)	191 (17.7%)	
	Norte	333 (10.3%)	111 (10.3%)	111 (10.3%)	111 (10.3%)	
	Petrovale	41 (1.3%)	13 (1.2%)	14 (1.3%)	14 (1.3%)	
	PTB	290 (9.0%)	108 (10.0%)	91 (8.4%)	91 (8.4%)	
	Sede	583 (18.0%)	201 (18.6%)	191 (17.7%)	191 (17.7%)	
	Terezópolis	319 (9.8%)	109 (10.1%)	105 (9.7%)	105 (9.7%)	
	Vianópolis	193 (6.0%)	53 (4.9%)	70 (6.5%)	70 (6.5%)	
Sex	Female	1628 (50.3%)	548 (50.8%)	536 (49.6%)	544 (50.4%)	0.8619
Age range	0–5	217 (6.7%)	71 (6.6%)	73 (6.8%)	73 (6.8%)	1.0000
	6–19	650 (20.1%)	218 (20.2%)	217 (20.1%)	215 (19.9%)	
	20–39	1067 (32.9%)	354 (32.8%)	355 (32.9%)	358 (33.1%)	
	40–59	871 (26.9%)	291 (27.0%)	289 (26.8%)	291 (26.9%)	
	Above 60	434 (13.4%)	145 (13.4%)	146 (13.5%)	143 (13.2%)	
Pneumopathy	Yes	30 (0.9%)	7 (0.6%)	13 (1.2%)	10 (0.9%)	0.4042
Chronic neurological disease	Yes	39 (1.2%)	16 (1.5%)	10 (0.9%)	13 (1.2%)	0.4948
Pregnant	Yes	28 (0.9%)	10 (0.9%)	11 (1.0%)	7 (0.6%)	0.6257
Postpartum	Yes	9 (0.3%)	2 (0.2%)	3 (0.3%)	4 (0.4%)	0.7165
Chronic cardiovascular disease	Yes	96 (3.0%)	34 (3.2%)	39 (3.6%)	23 (2.1%)	0.1154
Chronic kidney disease	Yes	50 (1.5%)	24 (2.2%)	12 (1.1%)	14 (1.3%)	0.0799
Obesity	Yes	105 (3.2%)	33 (3.1%)	37 (3.4%)	35 (3.2%)	0.8903
Asthma	Yes	173 (5.3%)	65 (6.0%)	58 (5.4%)	50 (4.6%)	0.3537
Immunodepression	Yes	22 (0.7%)	9 (0.8%)	5 (0.5%)	8 (0.7%)	0.5507
Chronic liver disease	Yes	15 (0.5%)	4 (0.4%)	7 (0.6%)	4 (0.4%)	0.5478
Diabetes	Yes	228 (7.0%)	78 (7.2%)	74 (6.9%)	76 (7.0%)	0.9430
Hypertension	Yes	563 (17.4%)	190 (17.6%)	186 (17.2%)	187 (17.3%)	0.9698
Transplanted	Yes	4 (0.1%)	2 (0.2%)	1 (0.1%)	1 (0.1%)	0.7780
Cancer	Yes	23 (0.7%)	10 (0.9%)	8 (0.7%)	5 (0.5%)	0.4342
Any comorbidity	Yes	955 (29.5%)	327 (30.3%)	320 (29.6%)	308 (28.5%)	0.6552
Fever	Yes	224 (6.9%)	66 (6.1%)	70 (6.5%)	88 (8.1%)	0.1398
Cough	Yes	648 (20.0%)	185 (17.1%)	213 (19.7%)	250 (23.1%)	0.0022
Sore throat	Yes	397 (12.3%)	112 (10.4%)	125 (11.6%)	160 (14.8%)	0.0051
Dyspnoea	Yes	141 (4.4%)	49 (4.5%)	46 (4.3%)	46 (4.3%)	0.9336
Myalgia	Yes	284 (8.8%)	74 (6.9%)	99 (9.2%)	111 (10.3%)	0.0165
Rhinorrhoea	Yes	717 (22.1%)	205 (19.0%)	240 (22.2%)	272 (25.2%)	0.0025
Respiratory discomfort	Yes	188 (5.8%)	63 (5.8%)	58 (5.4%)	67 (6.2%)	0.7084
Nausea/vomit	Yes	120 (3.7%)	37 (3.4%)	39 (3.6%)	44 (4.1%)	0.7156
Headache	Yes	790 (24.4%)	244 (22.6%)	259 (24.0%)	287 (26.6%)	0.0936
Prostration	Yes	188 (5.8%)	60 (5.6%)	51 (4.7%)	77 (7.1%)	0.0523
Diarrhea	Yes	211 (6.5%)	59 (5.5%)	76 (7.0%)	76 (7.0%)	0.2336
Conjunctivitis	Yes	32 (1.0%)	13 (1.2%)	11 (1.0%)	8 (0.7%)	0.5478
Ageusia/anosmia	Yes	101 (3.1%)	30 (2.8%)	30 (2.8%)	41 (3.8%)	0.2914
Loss of voice	Yes	56 (1.7%)	18 (1.7%)	13 (1.2%)	25 (2.3%)	0.1381
Sought health assistance	Hospital	138 (4.3%)	41 (3.8%)	41 (3.8%)	56 (5.2%)	0.1492
	Basic Health Unit	129 (4.0%)	42 (3.9%)	41 (3.8%)	46 (4.3%)	
	Emergency Care Unit	127 (3.9%)	38 (3.5%)	35 (3.2%)	54 (5.0%)	
	None	2845 (87.8%)	958 (88.8%)	963 (89.2%)	924 (85.6%)	
Admitted to a health institution	Yes	38 (1.2%)	11 (1.0%)	12 (1.1%)	15 (1.4%)	0.7085
International travel	Yes	14 (0.4%)	10 (0.9%)	4 (0.4%)	0 (0.0%)	0.0043
Household contact with symptomatic person	Yes	640 (19.8%)	157 (14.6%)	193 (17.9%)	290 (26.9%)	<0.0001
Serological test	Reactive	39 (1.2%)	3 (0.3%)	8 (0.7%)	28 (2.6%)	<0.0001
	Non-reactive	3200 (98.8%)	1076 (99.7%)	1072 (99.3%)	1052 (97.4%)	
PCR test	Detected	84 (2.6%)	2 (0.2%)	22 (2.0%)	60 (5.6%)	<0.0001
	Undetected	3112 (96.1%)	1035 (95.9%)	1057 (98.0%)	1020 (94.4%)	
	Indeterminate	42 (1.3%)	42 (3.9%)	0 (0.0%)	0 (0.0%)	
Prevalence	Serological reactive and/or PCR detected	106 (3.3%)	5 (0.5%)	29 (2.7%)	72 (6.7%)	<0.0001

Bolded p-values indicate $p < 0.05$.

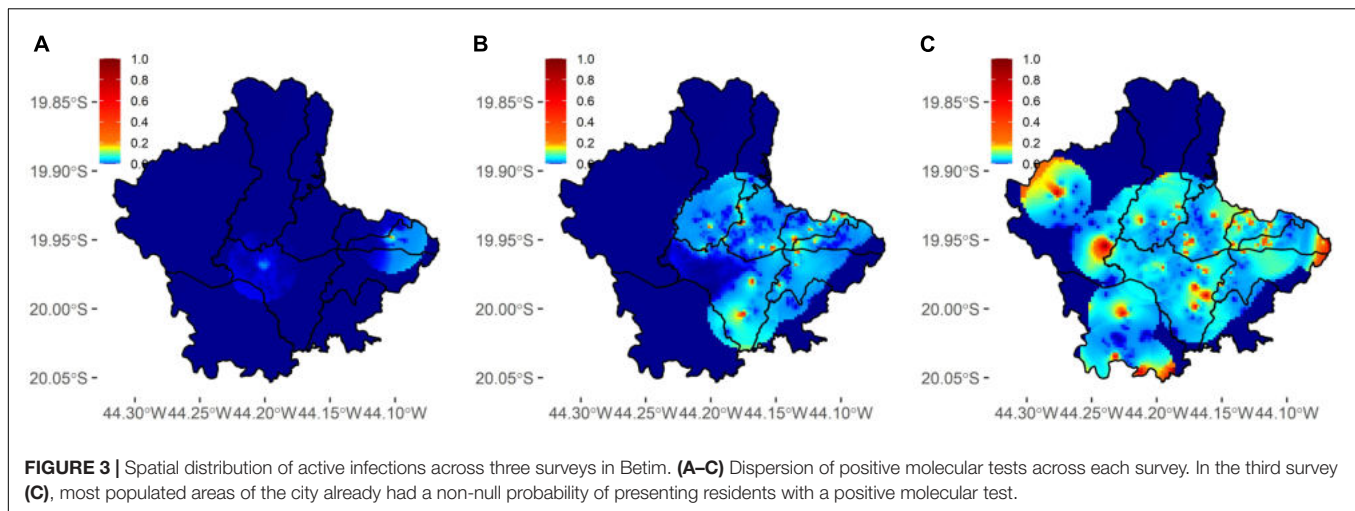


Betim clusters. One cluster comprises four sequences, dating to May 27, 2020 (95% HPD: May 01, 2020–June 03, 2020) grouped with other sequences from other Brazilian regions and foreign. The model has also estimated eight introductions supported by single sequences. According to our phylogeny, the B.1.1.33 introductions came from different locations, such as the states of Rio de Janeiro, São Paulo, Minas Gerais, other Brazilian regions, and foreign sequences (PP > 0.81 for all events) (**Figure 5B**). The patterns reconstructed by both phylogeographic inferences are consistent, indicating the emergence of lineages B.1.1.28 and B.1.1.33 was followed by multiple importation events to diverse regions within the country, likely driven by human mobility. Additionally, evolutionary rate estimates also differed between

datasets (B.1.1.28: 8.6372×10^{-4} , 95% HPD: 7.8379×10^{-4} – 9.4559×10^{-4} ; B.1.1.33: 6.8743×10^{-4} , 95% HPD 6.1784×10^{-4} – 7.5446×10^{-4}).

DISCUSSION

Betim is a medium-sized Brazilian city (439,340 inhabitants, 343 thousand square kilometers) crossed by national roads connecting major Brazilian cities and serving as a local hub for the Brazilian Public Health System. Understanding its pandemic dynamic may provide relevant information for municipalities with similar features. Here, we estimated the overall prevalence



of active infections, seroprevalence and conducted genomic surveillance before the first pandemic wave in Betim. Brazilian molecular diagnostic capacity was insufficient in the first months of the pandemic (Grotto et al., 2020). Therefore, COVID-19 cases may have been included in the official statistics as severe acute respiratory infection cases with unknown etiology. Data until May 2020 indicated a positive association between higher per-capita income and molecular COVID-19 diagnosis, while the severe acute respiratory infection cases with unknown etiology were associated with lower per-capita income, suggesting a possible diagnosis bias related to economic status (de Souza et al., 2020). Inadequate diagnosis availability may lead to underreporting (Kupek, 2021). Our data estimated underreporting rates up to 20 times.

No studies have been conducted in Brazil evaluating active infection prevalence using adequate sampling. Our study design was inspired by previous research conducted in Santa Clara, United States, using pooled samples (Hogan et al., 2020). Pooled PCR tests were initially suggested to be used in asymptomatic people (Lohse et al., 2020) and later were recommended for surveillance studies in populations with low infection prevalence (Mutesa et al., 2021). Seroprevalence studies were conducted during the first wave in Brazil that peaked in July 2020. Two of the highest city seroprevalences reported during the period were Boa Vista (25.4% in June 2020) (Hallal et al., 2020) and São Luiz (40.4% between the end of July and August 2020) (da Silva et al., 2020), both in the northern area of the country. A nationwide survey carried out in May and June 2020 presented seroprevalence lower than 2% during both surveys in all sampled cities neighboring Betim (less than 200 km), corroborating our findings (Hallal et al., 2020). Furthermore, seroprevalences higher than 10% were solely found in towns in the North Region (Hallal et al., 2020). In December 2020, Manaus, the largest city in the North Region, experienced a resurgence of COVID-19 (Sabino et al., 2021) despite high seroprevalence (Buss et al., 2021), likely due to the gamma variant (Faria et al., 2021).

Previous seroprevalence studies have indicated ethnic and socioeconomic bias for SARS-CoV-2 infection in Brazil since

the pandemic's beginning (Amorim Filho et al., 2020; Horta et al., 2020). Results from Rio de Janeiro in April 2020 indicated that younger blood donors with lower education levels were more likely to test positive for SARS-CoV-2 antibodies (Amorim Filho et al., 2020). A nationwide study revealed that the poorest quintile was 2.16 times more likely to test positive with the lowest risks among white, educated, and wealthy individuals (Horta et al., 2020). Likewise, we found one of the highest prevalences in the poorest neighborhood, Terezópolis, that include the largest slum of the city where more than 23,000 people live.

Further modeling results showed higher infection rates among young adults, lower socioeconomic status, and people without healthcare access in the less developed North and Northeast areas until August 2020 (de Lima et al., 2021). Betim also presents most of its inhabitants with less than 59 years (90.7%), but no age effect was observed in the infectivity rates. Increased female infection odds were observed, although previous reports indicated a gender predisposition toward death in some Brazilian regions with higher male risk (Baqui et al., 2020). One possible explanation could be that 70% of the global health workforce are women (Lotta et al., 2021) and a gender bias of pandemic perception and attitude (Galasso et al., 2020).

COVID-19 diffusion presents strong socio-spatial determinants. Relocation diffusion from more- to less-developed regions and hierarchical diffusion from countries with higher population and density were relevant since early 2020 (Sigler et al., 2021). Data indicated a similar pattern in the São Paulo State with contiguous diffusion from the capital metropolitan area and hierarchical with long-distance spread through major highways that connects São Paulo city with cities of regional relevance (Fortaleza et al., 2021). Modeling results revealed that São Paulo city may have accounted for more than 85% of the initial case spread in the entire country (Nicolelis et al., 2021). Betim is directly connected to São Paulo city by a main national highway which may have contributed to COVID-19 diffusion.

Genomic surveillance is a powerful tool to elucidate viral dispersion patterns. The first sequencing work conducted in

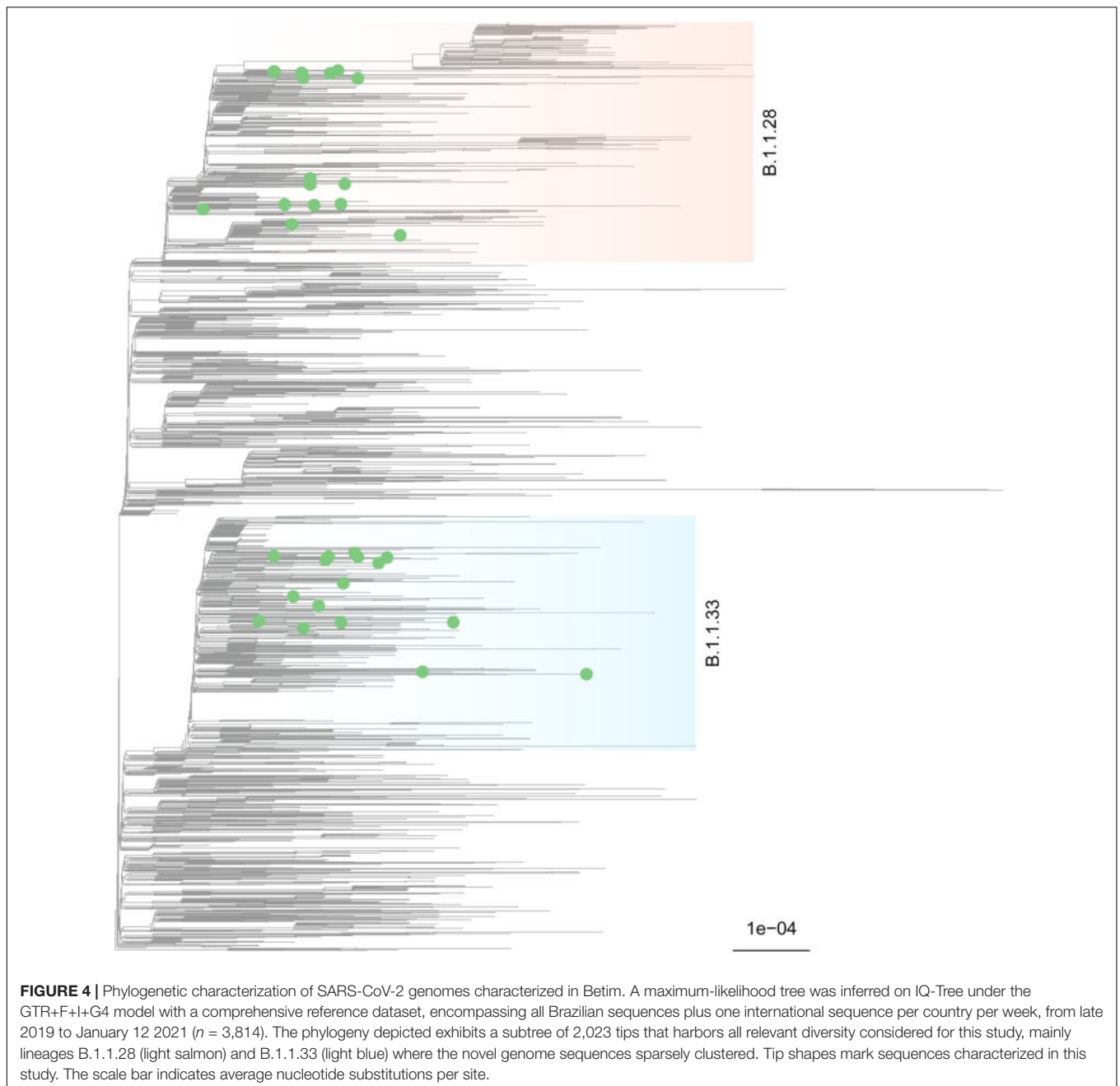
TABLE 2 | Significant clinical and epidemiological data associations with a positive test (serological or molecular).

Variable	Level	Positive	Negative	p-value
Survey	First	5 (4.7%)	1074 (34.3%)	<0.0001
	Second	29 (27.4%)	1051 (33.5%)	
	Third	72 (67.9%)	1008 (32.2%)	
Administrative regions	Alterosas	18 (17.0%)	616 (19.7%)	0.0024
	Citrolândia	4 (3.8%)	215 (6.9%)	
	Icaivera	0 (0.0%)	62 (2.0%)	
	Imbiruçu	32 (30.2%)	533 (17.0%)	
	Norte	11 (10.4%)	322 (10.3%)	
	Petrovale	0 (0.0%)	41 (1.3%)	
	PTB	8 (7.5%)	282 (9.0%)	
	Sede	15 (14.2%)	568 (18.1%)	
	Terezópolis	17 (16.0%)	302 (9.6%)	
	Vianópolis	1 (0.9%)	192 (6.1%)	
Sex	Female	69 (65.1%)	1559 (49.8%)	0.0026
Fever	No	88 (83.0%)	2927 (93.4%)	<0.0001
	Yes	18 (17.0%)	206 (6.6%)	
Cough	No	73 (68.9%)	2518 (80.4%)	0.0053
	Yes	33 (31.1%)	615 (19.6%)	
Sore throat	No	77 (72.6%)	2765 (88.3%)	<0.0001
	Yes	29 (27.4%)	368 (11.7%)	
Dyspnoea	No	96 (90.6%)	3002 (95.8%)	0.0180
	Yes	10 (9.4%)	131 (4.2%)	
Myalgia	No	72 (67.9%)	2883 (92.0%)	<0.0001
	Yes	34 (32.1%)	250 (8.0%)	
Rhinorrhoea	No	70 (66.0%)	2452 (78.3%)	0.0041
	Yes	36 (34.0%)	681 (21.7%)	
Respiratory discomfort	No	90 (84.9%)	2961 (94.5%)	<0.0001
	Yes	16 (15.1%)	172 (5.5%)	
Nausea/vomit	No	94 (88.7%)	3025 (96.6%)	<0.0001
	Yes	12 (11.3%)	108 (3.4%)	
Headache	No	50 (47.2%)	2399 (76.6%)	<0.0001
	Yes	56 (52.8%)	734 (23.4%)	
Prostration	No	83 (78.3%)	2968 (94.7%)	<0.0001
	Yes	23 (21.7%)	165 (5.3%)	
Ageusia/anosmia	No	87 (82.1%)	3051 (97.4%)	<0.0001
	Yes	19 (17.9%)	82 (2.6%)	
Obesity	No	96 (90.6%)	3038 (97.0%)	<0.0001
	Yes	10 (9.4%)	95 (3.0%)	
Sought health assistance	Hospital	8 (7.5%)	130 (4.1%)	0.0032
	None	81 (76.4%)	2764 (88.2%)	
	Basic Health Unit	8 (7.5%)	121 (3.9%)	
	Emergency Care Unit	9 (8.5%)	118 (3.8%)	
Household contact with symptomatic person	No	71 (67.0%)	2528 (80.7%)	0.0007
	Yes	35 (33.0%)	605 (19.3%)	

Non-significant associations are presented in **Supplementary Table 1**. Bolded p-values indicate $p < 0.05$.

Brazil evaluated the first six positive individuals and reported the same predominant lineages in Italy (Jesus et al., 2020). Later, a study with samples collected until late April 2020 from different country areas showed the dominance of clade B-derived lineages. At the national level, the respective frequency of these clades was seen in a 98.98%/1.02% ratio (Candido et al., 2020). In

Minas Gerais State, A lineages represented 2.5% of the infections, B.1 appeared in 92.5% of the samples, and B was responsible for 5% of the cases (Xavier et al., 2020). The exclusivity of lineages B.1.1.28 and B.1.1.33 circulating in Betim-MG from June to July 2020, given that multiple introductions from different country regions were demonstrated, is representative of the



extent of these lineages' dominance in the Brazilian scenario at the moment. Independent introductions also emphasize the importance of inter-state mobility barriers as a measure to control the epidemic.

Our study presents some limitations. First, the household survey is less likely to sample severe cases, thus underestimating symptomatic COVID-19. Second, all clinical data were self-reported, which may lead to reporting bias (Baker et al., 2004). Third, we could not sequence all PCR positive samples due to the low viral load and sequencing technology employed, but we do not expect biased frequency estimation since no differences in mean viral load was reported across B.1.1.28

and B.1.1.33 lineages. A different scenario, later in 2020, was observed after variants of concern detection that led to higher mean viral loads (e.g., P.1 or gamma variant) (Faria et al., 2021). Fourth, our study provides a limited picture of the local epidemic because of the short period across surveys, although it was the moment when the city had less reliable data regarding pandemic progression. Nevertheless, our study shows the potential to integrate different epidemiological inquiries (prevalence, seroprevalence, and genomic surveillance) to describe pandemic dispersion adequately. Moreover, our findings present original and relevant evidence that has helped local government authorities to guide pandemic management.

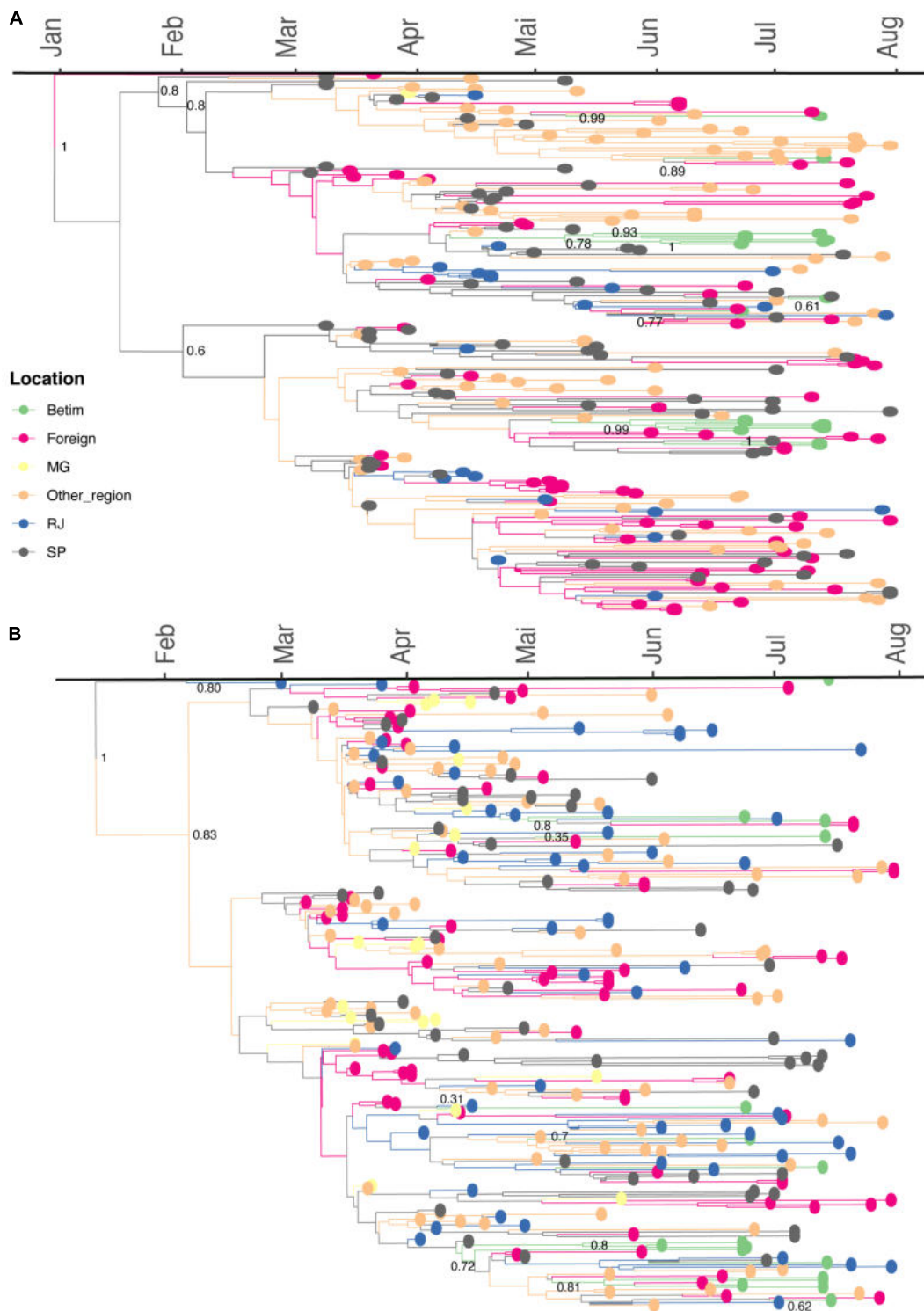


FIGURE 5 | Spread of B.1.1.28 and B.1.1.33 lineages in Betim city. **(A)** Time-resolved maximum clade credibility phylogeny from a dataset comprehending 240 publicly available B.1.1.28 sequences and the 18 genomes generated in this study. **(B)** Time-resolved maximum clade credibility phylogeny from a dataset including 267 publicly available B.1.1.33 sequences and the 17 genomes generated in this study. For both analyses, the HKY+I+G4 nucleotide substitution model was used. The diamond indicates sequences from Betim city obtained in this study. The trees inferred are available on https://github.com/LBI-lab/SARS-CoV-2_phylogenies.git.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Betim Ethics Research Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AVFGS, OT, VA, TA, AC, AR, HO, NO, EGS, PM, CM, TN, FP, EFS, RA, and RPS: study design. AVFGS, DM, FRM, OT, PF, RM, HA, VA, TA, AC, JMS, AR, LA, JA, HO, NO, CZ, JHS, EGS, RS, LF, AL, APG, PM, FMM, LM, CM, TN, FP, DCQ, DNQ, LR-M, FS, EFS, CV, AV, RA, and RPS: experimental procedures. AVFGS, DM, FRM, OT, PF, and RPS: manuscript draft. All authors revised and approved the manuscript.

FUNDING

We acknowledge support from the Fundo Municipal de Saúde de Betim, Rede Corona-ômica BR MCTI/FINEP

REFERENCES

- Amorim Filho, L., Szwarcwald, C. L., Mateos Sog, S., Leon, A. C. M. P., Medronho, R. A., Veloso, V. G., et al. (2020). Seroprevalence of anti-SARS-CoV-2 among blood donors in Rio de Janeiro, Brazil. *Rev. Saude Publica* 54:69. doi: 10.11606/s1518-8787.2020054002643
- Araujo, D. B., Machado, R. R. G., Amgarten, D. E., de, F., Malta, M., de Araujo, G. G., et al. (2020). SARS-CoV-2 isolation from the first reported patients in Brazil and establishment of a coordinated task network. *Mem. Inst. Oswaldo Cruz* 115:e200342. doi: 10.1590/0074-02760200342
- Araújo, J. L., Menezes, D., Saraiva-Duarte, J. M., Ferreira, L., Aguiar, R., and Souza, R. (2021). Systematic review of host genetic association with Covid-19 prognosis and susceptibility: What have we learned in 2020? *Rev. Med. Virol.* e2283. doi: 10.1002/rmv.2283
- Baker, M., Stabile, M., and Deri, C. (2004). What do self-reported, objective, measures of health measure? *J. Hum. Resour.* 39:1067. doi: 10.2307/3559039
- Baqui, P., Bica, I., Marra, V., Ercole, A., and van der Schaar, M. (2020). Ethnic and regional variations in hospital mortality from COVID-19 in Brazil: a cross-sectional observational study. *Lancet Glob. Health* 8, e1018–e1026. doi: 10.1016/S2214-109X(20)30285-0
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Buss, L. F., Prete, C. A., Abraham, C. M. M., Mendrone, A., Salomon, T., de Almeida-Neto, C., et al. (2021). Three-quarters attack rate of SARS-CoV-2 in the Brazilian amazon during a largely unmitigated epidemic. *Science* 371, 288–292. doi: 10.1126/science.abe9728
- Byambasuren, O., Dobler, C. C., Bell, K., Rojas, D. P., Clark, J., McLaws, M.-L., et al. (2021). Comparison of seroprevalence of SARS-CoV-2 infections with cumulative and imputed COVID-19 cases: Systematic review. *PLoS One* 16:e0248946. doi: 10.1371/journal.pone.0248946

affiliated to RedeVirus/MCTI (FINEP 01.20.0029.000462/20, CNPq 404096/2020-4), CNPq (AV: 303170/2017-4; RA: 312688/2017-2 and 439119/2018-9; RPS: 310627/2018-4), MEC/CAPES (14/2020–23072.211119/2020-10), FINEP (0494/20 01.20.0026.00 and UFMG-NB3 1139/20), FAPEMIG (RPS: APQ-00475-20) and FAPERJ (AV: E-26/202.903/20 and Corona-ômica-RJ E-26/210.179/2020; CV: 26/010.002278/2019; RA 202.922/2018).

ACKNOWLEDGMENTS

We want to thank nurses, community health workers, drivers, and management personnel who collaborated on this project. We also thank Guilherme Carvalho da Paixão for his support. We gratefully acknowledge the 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 Table 3**).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.799713/full#supplementary-material>

- Candido, D. S. I., Claro, M., de Jesus, J. G., Souza, W. M., Moreira, F. R. R. R., Dellicour, S., et al. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369, 1255–1260. doi: 10.1126/SCIENCE.ABD2161
- da Silva, A. A. M., Lima-Neto, L. G., Azevedo, C. M. P. E. S., Costa, L. M. M. D., Bragança, M. L. B. M., Barros Filho, A. K. D., et al. (2020). Population-based seroprevalence of SARS-CoV-2 and the herd immunity threshold in Maranhão. *Rev. Saude Publica* 54:131. doi: 10.11606/s1518-8787.2020054003278
- de Lima, E. E. C., Gayawan, E., Baptista, E. A., and Queiroz, B. L. (2021). Spatial pattern of COVID-19 deaths and infections in small areas of Brazil. *PLoS One* 16:e0246808. doi: 10.1371/journal.pone.0246808
- de Souza, W. M., Buss, L. F., da, D., Candido, S., Carrera, J.-P., Li, S., et al. (2020). Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat. Hum. Behav.* 4, 856–865. doi: 10.1038/s41562-020-0928-4
- Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. D. S., Mishra, S., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372, 815–821. doi: 10.1126/science.abh2644
- Fortaleza, C. M. C. B., Guimarães, R. B., Catão, R. C., Ferreira, C. P., Berg de Almeida, G., Nogueira Vilches, T., et al. (2021). The use of health geography modeling to understand early dispersion of COVID-19 in São Paulo, Brazil. *PLoS One* 16:e0245051. doi: 10.1371/journal.pone.0245051
- Galasso, V., Pons, V., Profeta, P., Becher, M., Brouard, S., and Foucault, M. (2020). Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries. *Proc. Natl. Acad. Sci. U.S.A.* 117, 27285–27291. doi: 10.1073/pnas.2012520117
- Gill, M. S., Lemey, P., Faria, N. R., Rambaut, A., Shapiro, B., and Suchard, M. A. (2013). Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* 30, 713–724.
- Grotto, R. M. T., Santos Lima, R., de Almeida, G. B., Ferreira, C. P., Guimarães, R. B., Pronunciate, M., et al. (2020). Increasing molecular diagnostic capacity and COVID-19 incidence in Brazil. *Epidemiol. Infect.* 148:e178. doi: 10.1017/S0950268820001818

- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hallal, P. C., Hartwig, F. P., Horta, B. L., Silveira, M. F., Struchiner, C. J., Vidaletti, L. P., et al. (2020). SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Glob. Health* 8, e1390–e1398. doi: 10.1016/S2214-109X(20)30387-9
- Hogan, C. A., Sahoo, M. K., and Pinsky, B. A. (2020). Sample Pooling as a strategy to detect community transmission of SARS-CoV-2. *JAMA* 323:1967. doi: 10.1001/jama.2020.5445
- Horta, B., Silveira, M., Barros, A., Barros, F., Hartwig, F., Dias, M., et al. (2020). Prevalence of antibodies against SARS-CoV-2 according to socioeconomic and ethnic status in a nationwide Brazilian survey. *Rev. Panam. Salud Pública* 44, 1–7. doi: 10.26633/RPSP.2020.135
- Jesus, J. G., Sacchi, C., Candido, D. D. S., Claro, I. M., Sales, F. C. S., Manuli, E. R., et al. (2020). Importation and early local transmission of COVID-19 in Brazil, 2020. *Rev. Inst. Med. Trop. Sao Paulo* 62, e30. doi: 10.1590/s1678-9946202062030
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kupek, E. (2021). How many more? Under-reporting of the COVID-19 deaths in Brazil in 2020. *Trop. Med. Int. Health* 26, 1019–1028. doi: 10.1111/tmi.13628
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Lemey, P., Rambaut, A., Drummond, A. J., and Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5:e1000520. doi: 10.1371/journal.pcbi.1000520
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lohse, S., Pfuhl, T., Berkó-Göttel, B., Rissland, J., Geißler, T., Gärtner, B., et al. (2020). Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *Lancet Infect. Dis.* 20, 1231–1232. doi: 10.1016/S1473-3099(20)30362-5
- Lotta, G., Fernandez, M., Pimenta, D., and Wenham, C. (2021). Gender, race, and health workers in the COVID-19 pandemic. *Lancet* 397:1264. doi: 10.1016/S0140-6736(21)00530-4
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Moreira, F. R. R., Bonfim, D. M., Zauli, D. A. G., Silva, J. P., Lima, A. B., Malta, F. S. V., et al. (2021). Epidemic spread of SARS-CoV-2 lineage B.1.1.7 in Brazil. *Viruses* 13: 984. doi: 10.3390/v13060984
- Mutesa, L., Ndishimye, P., Butera, Y., Souopgui, J., Uwineza, A., Rutayisire, R., et al. (2021). A pooled testing strategy for identifying SARS-CoV-2 at low prevalence. *Nature* 589, 276–280. doi: 10.1038/s41586-020-2885-5
- Nicoletti, M. A. L., Raimundo, R. L. G., Peixoto, P. S., and Andreazzi, C. S. (2021). The impact of super-spreader cities, highways, and intensive care availability in the early stages of the COVID-19 epidemic in Brazil. *Sci. Rep.* 11:13001. doi: 10.1038/s41598-021-92263-3
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., and Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst. Biol.* 67, 901–904. doi: 10.1093/sysbio/syy032
- Rambaut, A., Holmes, E. C., O’Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407. doi: 10.1038/s41564-020-0770-5
- Rambaut, A., Lam, T. T., Max Carvalho, L., and Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2:vev007. doi: 10.1093/ve/vev007
- Robshaw, J. D., Alter, S. M., Solano, J. J., Shih, R. D., DeMets, D. L., Maki, D. G., et al. (2021). Genomic surveillance to combat COVID-19: challenges and opportunities. *Lancet Microb.* 2, e481–e484. doi: 10.1016/S2666-5247(21)00121-X
- Rossi, ÁD., de Araújo, J. L. F., de Almeida, T. B., Ribeiro-Alves, M., de Almeida Velozo, C., de Almeida, J. M., et al. (2021). Association between ACE2 and TMPRSS2 nasopharyngeal expression and COVID-19 respiratory distress. *Sci. Rep.* 11:9658. doi: 10.1038/s41598-021-88944-8
- Sabino, E. C., Buss, L. F., Carvalho, M. P. S., Prete, C. A., Crispim, M. A. E., Fraiji, N. A., et al. (2021). Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* 397, 452–455. doi: 10.1016/S0140-6736(21)00183-5
- Shu, Y., and McCauley, J. (2017). GISAID: global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494
- Sigler, T., Mahmuda, S., Kimpton, A., Loginova, J., Wohland, P., Charles-Edwards, E., et al. (2021). The socio-spatial determinants of COVID-19 diffusion: the impact of globalisation, settlement characteristics and population. *Global Health* 17:56. doi: 10.1186/s12992-021-00707-2
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4:vey016. doi: 10.1093/ve/vey016
- Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Am. Math. Soc. Lect. Math. Life Sci.* 17, 57–86.
- Wu, S. L., Mertens, A. N., Crider, Y. S., Nguyen, A., Pokpongkiet, N. N., Djajadi, S., et al. (2020). Substantial underestimation of SARS-CoV-2 infection in the United States. *Nat. Commun.* 11:4507. doi: 10.1038/s41467-020-18272-4
- Xavier, J., Giovanetti, M., Adelino, T., Fonseca, V., Barbosa da Costa, A. V., Ribeiro, A. A., 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, 1824–1834. doi: 10.1080/22221751.2020.1803146
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* 39, 306–314. doi: 10.1007/BF00160154
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

Conflict of 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.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Citation: Silva AVFG, Menezes D, Moreira FRR, Torres OA, Fonseca PLC, Moreira RG, Alves HJ, Alves VR, Amaral TMdR, Coelho AN, Saraiva Duarte JM, da Rocha AV, de Almeida LGP, de Araújo JLE, de Oliveira HS, de Oliveira NJC, Zolini C, de Sousa JH, de Souza EG, de Souza RM, Ferreira LdL, Lehmkühl Gerber A, Guimarães APdC, Maia PHS, Marim FM, Miguita L, Monteiro CC, Neto TS, Pugêdo FSE, Queiroz DC, Queiroz DNAC, Resende-Moreira LC, Santos FM, Souza EFC, Voloch CM, Vasconcelos AT, de Aguiar RS and Souza RP (2022) Seroprevalence, Prevalence, and Genomic Surveillance: Monitoring the Initial Phases of the SARS-CoV-2 Pandemic in Betim, Brazil. *Front. Microbiol.* 13:799713. doi: 10.3389/fmicb.2022.799713*

Copyright © 2022 Silva, Menezes, Moreira, Torres, Fonseca, Moreira, Alves, Alves, Amaral, Coelho, Saraiva Duarte, da Rocha, de Almeida, de Araújo, de Oliveira, de Oliveira, Zolini, de Sousa, de Souza, de Souza, Ferreira, Lehmkühl Gerber, Guimarães, Maia, Marim, Miguita, Monteiro, Neto, Pugêdo, Queiroz, Queiroz, Resende-Moreira, Santos, Souza, Voloch, Vasconcelos, de Aguiar and Souza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.