SARS-CoV-2 intralineage variation and temporal patterns of COVID-19 risk factors in three cities of southeastern Brazil: Age, sex, and race

André Eterovic1 | Márcia Aparecida Sperança2 | Ivana Barros de Campos3 | Daniela Rodrigues Colpas3 | Itatiana Ferreira Rodart3 | Andréia Moreira dos Santos Carmo3 | Elaine Cristina de Mattos3 | Vilma dos Santos Menezes Gaiotto Daros3 | Maria Cecilia Cergole-Novella3

1Department Ciências Naturais e Humanas Center, Universidade Federal do ABC, Santo André, São Paulo, Brazil
2Department Ciências Naturais e Humanas Center, Universidade Federal do ABC, São Bernardo do Campo, São Paulo, Brazil
3Department Santo André Regional Center, Adolfo Lutz Institute, São Paulo, Brazil

Correspondence
Maria Cecilia Cergole-Novella, Instituto Adolfo Lutz, Santo André Regional Center, Avenue Ramiro Colleoni, Vila Dora, Santo André, SP 09040-160, Brazil. Email: maria.novella@ial.sp.gov.br

Abstract
The Santo André Regional Center from Adolfo Lutz Institute evaluated 91,537 samples by reverse transcription-polymerase chain reaction (RT-PCR) to detect severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from March 2020 to April 2021. The age, sex, and race of patients from three cities in southeastern Brazil, namely São Bernardo do Campo (SBC), Diadema, and Mauá were assessed in association to the rate of positive results using generalized linear models. Circulating lineages were obtained from GISAID and intralineage genetic variation was investigated employing Lasergene software. A declining number of reported cases around October to November 2020 separate two epidemic waves in the three cities. Mauá differed by the highest positive RT-PCR scores in January and February. GISAID classification of 38 SARS-CoV-2 complete genomic sequences showed the circulation of lineages P.1, B.1.1.28, P.2, B.1.1.332; P.1, P.2, B.1.1.28, B.1.1.33; and P.1, P.2 in SBC, Diadema and Mauá, respectively. Intralineage variation revealed a significant amino-acid substitution in the ORF3a encoding protein (A33S) present in four out of six (67%) P.1 Mauá isolates. As ORF3a encodes a nonselective Ca2+ permeable cation channel, supposed to interfere in airway homeostasis, specific mutations could increase its pathogenic effect resulting in a higher number of symptomatic individuals explaining why the second wave was more intense in Mauá city.

KEYWORDS
- coronavirus infection
- COVID-19
- epidemiology
- pandemic
- public health
- RT-PCR

1 | INTRODUCTION

In December 2019, a new coronavirus causing a pneumonia-like respiratory disease was discovered in Wuhan, China.1 In January 2020, the sequencing data were made available to the scientific community,2 and the new virus was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes a highly complex disease called coronavirus disease-2019 (COVID-19).3 SARS-CoV-2 can affect the lungs, with the possibility of progressing to a SARS death indicated in its name. It spreads
mainly through secretions droplets of the upper respiratory tract. The surveillance of respiratory viruses of public health relevance has a characteristic dynamic, due to their potential to cause epidemics. Studies that result in an improved understanding of COVID-19 may lead to more targeted and effective community-based and healthcare system-based interventions. The collection of COVID-19 data among specific population strata and their dissemination remains critically important to guide policy, health care, prevention, and intervention efforts.5,6

For diagnosis of COVID-19, nasopharyngeal secretion samples are obtained by using synthetic fiber swabs with a plastic rod, being the best time for collection between the first and the seventh day after the onset of symptoms.7 In hospitalized patients, the Brazilian Ministry of Health recommends the collection of bronchoalveolar lavage, tracheal or nasopharynx suction, or even oropharynx as the preferred swab sample. The collection process of these clinical samples is less prone to errors than those previously indicated swab methods, thus favoring results with greater accuracy for reverse transcription-polymerase chain reaction (RT-PCR) in real time.7 It is possible to detect the presence of viral RNA still in the initial stage of the disease, assisting in the diagnosis for better medical conduct, as well as in the necessary public health interventions. The World Health Organization (WHO) relegates RT-PCR as the gold standard for laboratory diagnosis, being essential in the pandemic context.8

Discrimination of SARS-CoV-2 lineages is associated with important epidemiological and biological events and the detection of changes in amino acid sequences along virus encoding proteins are based on comparison to the first isolate from the Chinese Hubei province (the root sequence).9 Assignment of epidemiological lineages in an emerging pandemic is usually done by using the Pangolin tool.9,10 Camargo et al.11 report a predominance of B.1.1.33 (30.3%, n = 3028 sequences) and B.1.1.28 (26.6%) strains circulating in Brazil. In Sao Paulo state, Southeastern Brazil, B.1.1.28 was the first-ranked. It was the source for P.1 (the so-called Gama variant) and P.2, both commonly detected in the state, and also for P.3, not reported in the country during the study period. B.1.1.28 has been found in 46 countries, P.1 in 27, P.2 in 28, and P.3 has been reported only in the Philippines.12

The Adolfo Lutz Institute, Santo André Regional Center (IAL Santo André), Sao Paulo, Brazil, as the central public health laboratory, is responsible for laboratory tests of various diseases required for epidemiological surveillance. During the first year of the COVID-19 pandemic, the diagnosis of SARS-CoV-2 was included in a platform used in this study to describe the epidemiological characteristics of symptomatic patients from Sao Paulo state. Our purpose is to evaluate the association of positive RT-PCR results to demographic factors (age, sex, and race/ethnicity) in three cities with massive records (São Bernardo, Diadema, and Mauá), besides the identification of the main circulating virus lineages.

2 | MATERIALS AND METHODS

2.1 | Statistical analysis

From March 2020 to April 2021, 91,537 samples were collected and sent to the IAL Santo André for RT-PCR assessment of SARS-CoV-2. Patients are mainly from the seven municipalities of the Metropolitan Region of ABC (ordered by decreasing sample size: São Bernardo do Campo, Diadema, Mauá, Rio Grande da Serra, Ribeirão Pires, Santo André, and São Caetano do Sul), followed by 39 cities of the Paraíba Valley Region (being the top-three: Taubaté, São José dos Campos, and Pindamonhangaba), in Sao Paulo state, Southeastern Brazil. Age, sex, race/ethnicity, and purpose of such evaluation (clinical diagnosis of symptoms) were recorded for each person, as well as the date of collection, being the only available information. Sometimes sequential samples were taken from some individuals, which were excluded. For further analysis, the entire database was subset to acquire sampling homogeneity throughout the months of collection and among the different categories of subjects in three cities nearby the capital Sao Paulo that presented disproportionately massive records: São Bernardo do Campo, Diadema, and Mauá. We retained only data acquired since July 2020, spanning 10 months with no less than 3000 samples each. We included only the first, conclusive test for each individual, aged less than 90 years and affected with a mild respiratory syndrome. Scattered data from indigenous people were also discarded.

The three cities were compared according to temporal trends in the rate of positive test results, as assessed in the two waves of the pandemics which strike such regions during the time course of sampling. For practical reasons, these waves were detached by the lowering of the daily incidence rate around the October to November 2020 edge, as detected using external, unofficial data.13 The probability of a positive test result (response variable) was assessed by generalized linear models with a binomial error distribution, using wave, city, day of collection, age, sex, and race/ethnicity as risk factors (predictors). Models were built to investigate the additive contribution of such variables. Interactions were also assessed by partitioning the data set by wave, city, and combinations of both. R program was used in such analysis.

2.2 | Samples collection and SARS-CoV-2 detection

Nasopharyngeal secretion samples (combined nose and throat) were collected with synthetic fiber swabs (rayon) with a plastic rod between the third and seventh day after the onset of symptoms. In hospitalized patients, was collected bronchoalveolar lavage, tracheal or nasopharynx suction, adopting oropharynx as the preferred swab sample.

The RT-PCR reactions for SARS-CoV-2 detection were performed using different methodologies, described below into three protocols. Until July 2020, samples had RNA extracted with silica columns, using Biogene kit (Quibasa) and RT-PCR reactions were performed with SARS-CoV-2 molecular kit (Bio-Manguinhos) for individual gene E detection of SARS-CoV-2 and human RP gene as a positive control, as standardized in
Charité laboratory protocol. From August to November 2020, samples were subjected to heat/chemical extraction using QuickExtract DNA Extraction Solution kit (Lucigen), and RT-PCR reactions were performed with IDT primer kit and GoTaq Probe 1 enzyme Step RT-qPCR System (Promega) for individual detection of N1 and N2 genes, according to a protocol standardized by Centers for Disease Control and Prevention (CDC), in addition to detection of the RP gene as a positive control. From December 2020 to April 2021, samples were subjected to automated extraction of genetic material using magnetic microparticles from EXTRACTA kit—DNA and Viral RNA and EXTRACTA 32 equipment (Locuss). RT-PCR reactions were performed with Allplex 2019-nCov Assay kit (Seegene) with modifications for the triplex detection of the E and N genes from SARS-CoV-2, in addition to the human RP gene, whose protocol is also based on the primers designed by Charité laboratory. Part of the tests that were submitted to the IAL Santo André was carried out by other laboratory on the state diagnostic platform, which also adopted similar kits and the protocol of Charité laboratory. All extraction reactions of the genetic material and RT-PCR were performed according to the manufacturers’ instructions, as well as the analysis and interpretation of results.

2.3 Identification of SARS-CoV-2 lineages and genetic variation

The IAL Sao Paulo responds to emerging diseases of public health interest, and it was responsible for sequencing the complete genome of isolated SARS-CoV-2 to monitor the diversity of new coronaviruses. The sequences are deposited and available from the Global Initiative on Sharing All Influenza Data (GISAID). All SARS-CoV-2 lineages described in this study were extracted from GISAID on June 21, 2021. All available SARS-CoV-2 complete sequences from São Bernardo do Campo, Diadema, and Mauá were downloaded and the modified target encoding genes in reference lineages strains were evaluated individually by employing MegAlign of the Lasergene software.

3 RESULTS

After discarding scattered data, the resulting number of records for the remaining cities were São Bernardo do Campo (n = 39 034), Diadema (n = 15 604), and Mauá (n = 12 245). A decreasing number of reported cases around the October to November 2020 edge—indicated by the moving averages of external data—roughly separate the two epidemic waves in the three cities (Figure SA1). The median of the monthly sample size for RT-PCR tests obtained with our data in each city was 1711, ranging from 376 (Mauá, July 2020) to 6421 (São Bernardo do Campo, Mar 2021) (Figure SA2). The age distribution of sampled patients increased abruptly from 13 years, roughly peaking around 20–35 years, and then decreasing monotonically (Figure SA3).

The temporal trend in the fraction of positive RT-PCR results (Figure 1) roughly resembles that of the new cases reported from external data (Figure SA1): stabilized (around 25%) or slightly decreasing from July to October 2020 (the end of the first wave),

![Figure 1](image-url) Fraction of positive results in reverse transcription-polymerase chain reaction COVID-19 tests for samples from the cities of São Bernardo do Campo (SBC, total n = 39 034), Diadema (DIA, n = 15 604), and Mauá (MAU, n = 12 245), from July 2020 (Month 1) to April 2021. For each month, proportions are presented with 95% confidence intervals based on a normal approximation. The vertical dashed line roughly separates two epidemic waves by the lowest figures for new cases in the October to November 2020 edge.
increasing to the highest values (more than 50% in Mauá) in the first 2 months of the following year (the ascending slope of the second wave) and, finally, with a new decrease in the subsequent period. Despite the sequential employment of three different procedures of RNA extraction and RT-PCR along the period of sample scanning, such protocols were not treated as levels of another explaining factor in the analysis of the chance of positive tests. As the third protocol was almost temporally coincident with the second epidemic wave, it should be statistically impossible to isolate its effect (i.e., to detach the contribution of "WAVE" and "PROTOCOL"). So, we assumed that the trend described at the beginning of this paragraph (Figure 1), as detected also in the external data (Figure SA1), was entirely attributed to the waves and not to the methodological approach.

This same “waving” pattern was recurrent among 15-years age classes, especially those of 30, 45, and 60 years, as labeled by their upper limit (Figure 2). It also appears in distinct genders (Figure 3) and races, especially in categories "white," "mixed," and "unreported" (Figure 4). The age distribution of positive RT-PCR results indicates a monotonic increase until the classes of 60 and 75 years, followed by a plateau or a slight decrease in the last group (Figure SA4). This pattern was especially evident in the months of the second wave (Figure SA5). Mauá differed from the other cities by higher positive RT-PCR scores in January and February 2021 (Figure 1), mainly in age classes 30, 45, and 60 (Figures 2, SA4, SA5), in both gender (Figure 3) and in all race categories except “black” (Figure 4).

The first additive model using all data and all independent variables indicated a strong increase in the chance of being RT-PCR positive in the second wave (Table 1). As indicated by their significant coefficients, the next factors in importance were “city” (Diadema and Mauá with a lower and a higher chance, respectively, of a positive result than São Bernardo do Campo) and “sex” (females less prone to positive results than males). Each year in age promotes an increase of 1.6% in the odds of a positive outcome. Each day after July 1, 2020, represents an increase of less than 0.1% in the odds of a positive outcome but such effect did not resist an adjusted critical P. No race category differed from the reference level (“white”) except the “unreported,” which presented a significant, increased odds of positive results.

The models using data subset by wave (Table 2) mainly repeat such patterns, with some peculiar differences. Mauá has a slightly increased chance of a positive result in the first wave, maintaining the
strong, positive effect in the second wave. The effect in reducing the chance of a positive outcome was more intense in the first than in the second wave for Diadema. The day of the collection indicates a reducing chance of a positive outcome in the end of the first wave and a reversion of such trend in the start of the second wave, reinforcing the division between them, as indicated by the external data on the number of cases (Figure SA1). The significant positive effect of age was slightly more intense in the second wave. The significant effect of gender (males with an increased chance of a positive outcome) did not differ among waves. The significant positive effect of unreported race was restricted to the second wave.

Cutting the data by city (Table 3) highlights some new scenarios. The effect of the second wave was more intense in Mauá than in Diadema or São Bernardo do Campo. The effect of day of collection was absent (Diadema, Mauá) or negligible (São Bernardo do Campo): both waves were canceled out by opposite trends. The effect of gender in Mauá did not resist a Bonferroni correction. The effect of unreported race was restricted to São Bernardo do Campo. The effect of age was significantly positive and similar in all cities and waves.

When models were built by grouping data according to both wave and city (Table 4), some interactions were evinced. The effect of day of collection was negative in the first wave in São Bernardo do Campo but positive in the second wave. It was absent in both waves in Diadema. It was negative in the first wave but absent in the second wave in Mauá. The effect of sex was absent in both waves in Mauá but did not differ among Diadema and São Bernardo do Campo, where males were prone to a positive result. The effect of unreported race was restricted to the second wave in São Bernardo do Campo, but this effect did not resist a Bonferroni correction. The effects of age were significantly positive and similar in all cities and waves.

Thirty-eight isolates of the new coronavirus, detected in this study between October 2 and December 2, 2020, were sequenced by the IAL Sao Paulo and deposited on GISAID. Classification of SARS-CoV-2 isolates from São Bernardo do Campo, Diadema, and Mauá and amino acid variability in specific target viral open reading frames are described in Table SA1. Accession number of obtained sequences are described in Table A1, as isolate. Nine out of these 38 SARS-CoV-2 isolates were from São Bernardo do Campo, being five of them classified as lineage P.1, one as P.2, two as B.1.1.28, and one as B.1.1.332. Among 20 isolates from Diadema, eight were classified as P.1, six as P.2, five as B.1.1.28, and one as B.1.1.33. Nine isolates were obtained from Mauá and presented two different lineages: P.1 (six sequences) and three as P.2. In most of such isolates, specific modifications in amino acid sequences of the target ORFs were observed, being significant a substitution in the residue 33 of the ORF3a (a serine instead of an alanine) found in four of the six (67%) P.1 sequences from Mauá patients (Table SA1).

4 | DISCUSSION

COVID-19 is a novel disease that is revealing unprecedented challenges. Initially, in the Sao Paulo state of Brazil, only cases of SARS, health professionals with symptoms of flu syndrome (FS), or
investigation of outbreak in a closed or semiclosed community such as prisons, long-term care institutions, schools, daycare centers, companies, should be notified as suspected of COVID-19 and have their material collected and analyzed by RT-PCR. Afterward, in July 2020, the investigation of COVID-19 was extended to all symptomatic individuals, including therefore, patients with FS, that is, defined cases with mild respiratory symptoms, characterized by a feverish sensation, accompanied by cough or sore throat or runny nose or difficulty breathing.

The number of samples received at the IAL Santo André was not correlated only to the population size of the cities of origin but also to better testing efforts and policy of such municipalities. A gradual increase in the monthly average of exams performed, as well as in the fraction of positive cases mainly in the second wave (November 2020 to April 2021, in our data) was observed. Probably, it was due to the agglomerations caused by the Brazilian municipal elections in October 2020 and the Holiday season (Christmas and New Year’s festivities) and possibly combined with the presence of more infectious variants. The decline in the first wave (until the October to November 2020 edge) indicates that the epidemiological progression in Brazil was largely influenced by the introduction of lockdown measures in that period. However, the subsequent flexibilization of such policies may have directly promoted a resurgence, which takes the form of the second wave.

As pointed before, Mauá differed from the other two neighbor cities by higher positive RT-PCR scores in January and February 2021, mainly in age classes 30, 45, and 60, in both genders, in all race categories except “black,” and presented at least two different variants, P.1 and P.2. It is important to note that, so far, there is no evidence that the variant P.1 is more virulent or transmissible compared to other previously identified. P.1 strain has been found in 27 countries and in Brazil, and it had an increased incidence in the period of this study. Since the first identification of SARS-CoV-2 to date, more than 414,575 complete genomic sequences worldwide are available through public access databases. The ability to monitor viral evolution in near real-time has a direct impact on the public health response to the COVID-19 pandemic. Camargo et al. reported that in just over a year after the COVID-19 pandemic was decreed, Brazil deposited 4606 sequences on GISAID, being the IAL Sao Paulo
TABLE 1  A generalized linear model for all data. Results of COVID-19 RT-PCR tests as a function of additive risk factors in three cities of Sao Paulo state, southeastern Brazil from July 2020 to April 2021

| Model          | Factor | Level  | B     | SE   | OR    | CI+   | CI−   | p      |
|----------------|--------|--------|-------|------|-------|-------|-------|--------|
| All data       | City   | Diadema| −0.197| 0.021| 0.821 | 0.856 | 0.788 | 0.0001**|
|                | Mauá   | 0.192  | 0.022 | 1.212| 1.265 | 1.161 | 0.0001**|
| Wave           |        |        | 0.713 | 0.034| 2.040 | 2.179 | 1.910 | 0.0001**|
| Collection     | 0.001  | 0.001  | 1.000 | 1.001| 1.000 | 0.0407*|
| Age            | 0.016  | 0.001  | 1.016 | 1.017| 1.015 | 0.0001**|
| Sex            | Female | −0.144 | 0.017 | 0.866| 0.895 | 0.838 | 0.0001**|
| Race           | Asian  | 0.010  | 0.028 | 1.010| 1.068 | 0.955 | 0.7195|
|                | Black  | −0.030 | 0.042 | 0.970| 1.055 | 0.893 | 0.4774|
|                | Mixed  | 0.014  | 0.024 | 1.014| 1.064 | 0.967 | 0.5679|
|                |        |        | 0.066 | 0.021| 1.068 | 1.112 | 1.025 | 0.0016**|

Note: The response variable is the probability of a positive outcome. “São Bernardo do Campo” is the reference level for cities, as well as “male” for sex and “white” for the race. Two waves defined by lower values of reported cases in external data were separated by the edge October to November 2020. Days of collection starting at July 11, 2020; age measured in years. Logistic regression coefficients (B) are shown with respective standard errors (SE), odds ratios (OR), and 95% confidence limits (CI). *p < 0.05; **p < 0.0167 (= 0.05/3) for City, p < 0.0125 (=0.05/4) for Wave and p < 0.0042 (=0.05/12) for the remainder; after a Bonferroni correction for the number of times a given factor was included in an analysis (a given model) in the whole paper.

TABLE 2  Two generalized linear models for epidemic waves. Results of COVID-19 RT-PCR tests as functions of additive risk factors during two epidemic waves in three cities of Sao Paulo state, southeastern Brazil

| Model  | Factor | Level  | B     | SE   | OR    | CI+   | CI−   | p      |
|--------|--------|--------|-------|------|-------|-------|-------|--------|
| Wave 1 | City   | Diadema| −0.290| 0.049| 0.749 | 0.824 | 0.680 | 0.0001 |
|        | Mauá   | 0.096  | 0.048 | 1.101| 1.211 | 1.002 | 0.0462*|
|        | Age    | −0.003 | 0.001 | 0.977| 0.998 | 0.996 | 0.0001**|
|        | Sex    | Female | −0.139| 0.033| 0.870 | 0.928 | 0.816 | 0.0001**|
|        | Race   | Asian  | −0.067| 0.057| 0.925 | 1.046 | 0.836 | 0.2407 |
|        | Black  | −0.145 | 0.088 | 0.865| 1.028 | 0.728 | 0.1002|
|        | Mixed  | 0.022  | 0.049 | 1.023| 1.126 | 0.928 | 0.6511|
|        | Unreported | 0.036 | 0.040 | 1.036| 1.120 | 0.959 | 0.3698|
| Wave 2 | City   | Diadema| −0.139| 0.024| 0.870 | 0.912 | 0.830 | 0.0001**|
|        | Mauá   | 0.239  | 0.025 | 1.270| 1.334 | 1.210 | 0.0001**|
|        | Collection | 0.001 | 0.001 | 1.001| 1.001 | 1.001 | 0.0001**|
|        | Age    | 0.017  | 0.001 | 1.017| 1.018 | 1.016 | 0.0001**|
|        | Sex    | Female | −0.144| 0.020 | 0.866| 0.900 | 0.833 | 0.0001**|
|        | Race   | Asian  | 0.032  | 0.033 | 1.033 | 1.101 | 0.968 | 0.3298|
|        | Black  | 0.004  | 0.049 | 1.004| 1.104 | 0.912 | 0.9423|
|        | Mixed  | 0.011  | 0.028 | 1.011| 1.069 | 0.957 | 0.6933|
|        | Unreported | 0.072 | 0.024 | 1.075| 1.128 | 1.025 | 0.0031**|

Note: The response variable is the probability of a positive outcome. The first wave lasts from July to October 2020 and the second, from November 2020 to April 2021. They were defined by lower values of reported cases in external data. “São Bernardo do Campo” is the reference level for cities, as well as “male” for sex and “white” for the race. Days of collection starting at July 1, 2020; age measured in years. Logistic regression coefficients (B) are shown with respective standard errors (SE), odds ratios (OR), and 95% confidence limits (CI). *p < 0.05; **p < 0.0167 (= 0.05/3) for City, p < 0.0125 (=0.05/4) for Wave and p < 0.0042 (=0.05/12) for the remainder; after a Bonferroni correction for the number of times a given factor was included in an analysis (a given model) in the whole paper.
responsible for 18.3% of these records. The predominant lineages in Brazil during the period of the study are B.1.1.33, B.1.1.28, P.1, and P.2. We observed that the variants isolated in the tree cities of the present study were those plus B.1.1.332 (Table SA1). In Mauá, four out of six P.1 isolates presented a significant specific amino acid substitution in the ORF3a, a serine (polar) at the place of an alanine (nonpolar) (Table SA1). Structural study of SARS-CoV-2 ORF3a revealed it is a nonselective Ca²⁺ permeable cation channel.²³ Using an infection model of lung Type II pneumocytes, where Ca²⁺ plays an important role in maintaining airway homeostasis, SARS-CoV-2 ORF3a could disrupt this homeostasis, affecting COVID-19 pathogenesis. There is evidence that SARS-CoV-2 ORF3a is implicated in activation of the inflammasome and null mutation mouse model for this encoding gene reduces viral titer and morbidity.²⁴ SARS-CoV-2 ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of apoptosis-associated speck-like protein containing a caspase recruit domain (ASC).²⁴,²⁵ Thus, the A33S specific mutation found in ORF3a of P.1 isolates from Mauá patients could interfere in the protein activity, increasing pathogenicity and consequently, the number of symptomatic individuals, which could explain the higher number of positive tests in this city during January and February 2021. Genomic sequencing of a higher number of isolates from Mauá and association to clinical and laboratory findings of infected individuals should be performed to investigate this hypothesis.

| Model                  | Factor   | Level | B    | SE     | OR    | CI+   | CI−   | p      |
|------------------------|----------|-------|------|--------|-------|-------|-------|--------|
| São Bernardo do Campo  | Wave     |       | 0.614| 0.047  | 1.848 | 2.025 | 1.687 | 0.0001**|
|                        | Collection|       | 0.001| 0.001  | 1.001 | 1.001 | 1.000 | 0.0105* |
|                        | Age      |       | 0.016| 0.001  | 1.016 | 1.018 | 1.015 | 0.0001**|
|                        | Sex      | Female| −0.155| 0.022 | 0.856 | 0.894 | 0.820 | 0.0001**|
|                        | Race     | Asian | 0.006| 0.037  | 1.006 | 1.081 | 0.936 | 0.8657  |
|                        |          | Black | −0.053| 0.058 | 0.949 | 1.062 | 0.847 | 0.3603  |
|                        |          | Mixed | 0.005| 0.035  | 1.005 | 1.076 | 0.939 | 0.8764  |
|                        |          | Unreported | 0.069| 0.027 | 1.072 | 1.129 | 1.017 | 0.0098* |
| Diadema                | Wave     |       | 0.805| 0.065  | 2.236 | 2.541 | 1.968 | 0.0001**|
|                        | Collection|       | 0.001| 0.001  | 1.001 | 1.001 | 1.000 | 0.0813  |
|                        | Age      |       | 0.015| 0.001  | 1.015 | 1.017 | 1.013 | 0.0001**|
|                        | Sex      | Female| −0.168| 0.036 | 0.845 | 0.906 | 0.788 | 0.0001**|
|                        | Race     | Asian | 0.036| 0.059  | 1.037 | 1.163 | 0.925 | 0.5342  |
|                        |          | Black | −0.001| 0.077 | 0.999 | 1.161 | 0.860 | 0.9900  |
|                        |          | Mixed | 0.019| 0.046  | 1.019 | 1.115 | 0.932 | 0.6784  |
|                        |          | Unreported | 0.088| 0.049 | 1.092 | 1.202 | 0.993 | 0.0706  |
| Mauá                   | Wave     |       | 0.885| 0.076  | 2.424 | 2.812 | 2.090 | 0.0001  |
|                        | Collection|       | 0.001| 0.001  | 1.000 | 1.001 | 0.999 | 0.8498  |
|                        | Age      |       | 0.016| 0.001  | 1.016 | 1.018 | 1.014 | 0.0001  |
|                        | Sex      | Female| −0.077| 0.038 | 0.926 | 0.998 | 0.859 | 0.0431  |
|                        | Race     | Asian | −0.015| 0.071 | 0.985 | 1.133 | 0.857 | 0.8363  |
|                        |          | Black | −0.001| 0.110 | 0.999 | 1.240 | 0.805 | 0.9937  |
|                        |          | Mixed | 0.023| 0.054  | 1.024 | 1.137 | 0.922 | 0.6631  |
|                        |          | Unreported | 0.034| 0.045 | 1.034 | 1.130 | 0.947 | 0.4538  |

Note: The response variable is the probability of a positive outcome. The first wave lasts from July to October 2020 and the second, from November 2020 to April 2021. They were defined by lower values of reported cases in external data.¹³ “Male” is the reference level for sex and “white” for the race. Days of collection starting at July 1, 2020; age measured in years. Logistic regression coefficients (B) are shown with respective standard errors (SE), odds ratios (OR), and 95% confidence limits (CI). *p < 0.05; **p < 0.0167 (= 0.05/3) for City, *p < 0.0125 (= 0.05/4) for Wave and p < 0.0042 (=0.05/12) for the remainder; after a Bonferroni correction for the number of times a given factor was included in an analysis (a given model) in the whole paper.
| Model          | Factor       | Level| B    | SE  | OR  | CI+ | CI− | p       |
|----------------|--------------|------|------|-----|-----|-----|-----|---------|
| Wave 1         | São Bernardo do Campo | Collection | -0.002 | 0.001 | 0.998 | 0.999 | 0.997 | 0.0002** |
|                | Age          | 0.013 | 0.001 | 1.013 | 1.016 | 1.011 | 0.0001** |
|                | Sex Female   | -0.114 | 0.039 | 0.892 | 0.962 | 0.827 | 0.0031** |
|                | Race Asian   | -0.124 | 0.067 | 0.884 | 1.007 | 0.776 | 0.0634 |
|                | Race Black   | -0.184 | 0.107 | 0.832 | 1.026 | 0.675 | 0.0859 |
|                | Race Mixed   | -0.014 | 0.061 | 0.986 | 1.112 | 0.874 | 0.8177 |
|                | Race Unreported | 0.052 | 0.046 | 1.054 | 1.153 | 0.963 | 0.2519 |
|                | Diadema      | Collection | -0.001 | 0.001 | 0.999 | 1.002 | 0.997 | 0.6444 |
|                | Age          | 0.009 | 0.003 | 1.009 | 1.014 | 1.003 | 0.0018** |
|                | Sex Female   | -0.267 | 0.091 | 0.766 | 0.915 | 0.641 | 0.0033** |
|                | Race Asian   | 0.185 | 0.143 | 1.204 | 1.593 | 0.909 | 0.1950 |
|                | Race Black   | 0.055 | 0.190 | 1.056 | 1.534 | 0.728 | 0.7733 |
|                | Race Mixed   | 0.110 | 0.116 | 1.117 | 1.402 | 0.889 | 0.3422 |
|                | Race Unreported | 0.024 | 0.123 | 1.024 | 1.304 | 0.805 | 0.8457 |
|                | Mauá         | Collection | -0.007 | 0.001 | 0.993 | 0.996 | 0.991 | 0.0001** |
|                | Age          | 0.012 | 0.003 | 1.012 | 1.017 | 1.007 | 0.0001** |
|                | Sex Female   | -0.135 | 0.090 | 0.874 | 1.042 | 0.733 | 0.1338 |
|                | Race Asian   | 0.007 | 0.177 | 1.007 | 1.425 | 0.712 | 0.9691 |
|                | Race Black   | -0.198 | 0.278 | 0.820 | 1.414 | 0.476 | 0.4757 |
|                | Race Mixed   | 0.089 | 0.123 | 1.094 | 1.392 | 0.859 | 0.4671 |
|                | Race Unreported | -0.057 | 0.106 | 0.944 | 1.163 | 0.767 | 0.5906 |
| Wave 2         | São Bernardo do Campo | Collection | 0.001 | 0.001 | 1.001 | 1.002 | 1.001 | 0.0001** |
|                | Age          | 0.017 | 0.001 | 1.018 | 1.019 | 1.016 | 0.0001** |
|                | Sex Female   | -0.175 | 0.027 | 0.840 | 0.885 | 0.796 | 0.0001** |
|                | Race Asian   | 0.058 | 0.044 | 1.059 | 1.155 | 0.971 | 0.1927 |
|                | Race Black   | 0.001 | 0.069 | 1.001 | 1.147 | 0.874 | 0.9876 |
|                | Race Mixed   | 0.018 | 0.042 | 1.018 | 1.106 | 0.937 | 0.6787 |
|                | Race Unreported | 0.074 | 0.033 | 1.076 | 1.148 | 1.009 | 0.0263*  |
|                | Diadema      | Collection | 0.001 | 0.001 | 1.001 | 1.002 | 1.000 | 0.0676 |
|                | Age          | 0.016 | 0.001 | 1.016 | 1.018 | 1.014 | 0.0001** |
|                | Sex Female   | -0.152 | 0.039 | 0.859 | 0.927 | 0.796 | 0.0001** |
|                | Race Asian   | 0.008 | 0.064 | 1.008 | 1.143 | 0.889 | 0.9005 |
|                | Race Black   | -0.011 | 0.084 | 0.989 | 1.166 | 0.839 | 0.8942 |
|                | Race Mixed   | 0.002 | 0.050 | 1.002 | 1.104 | 0.909 | 0.9673 |
|                | Race Unreported | 0.101 | 0.053 | 1.106 | 1.228 | 0.996 | 0.0588 |
|                | Mauá         | Collection | 0.001 | 0.001 | 1.001 | 1.002 | 1.000 | 0.1327 |
|                | Age          | 0.017 | 0.001 | 1.017 | 1.019 | 1.014 | 0.0001** |
|                | Sex Female   | -0.059 | 0.042 | 0.942 | 1.024 | 0.868 | 0.1610 |
|                | Race Asian   | -0.015 | 0.078 | 0.985 | 1.147 | 0.846 | 0.8472 |

(Continues)
TABLE 4 (Continued)

| Model       | Factor | Level | B   | SE  | OR     | CI− | CI+   | p       |
|-------------|--------|-------|-----|-----|--------|-----|-------|---------|
| Black       |        |       | 0.042 | 0.121 | 1.043 | 1.323 | 0.823 | 0.7265 |
| Mixed       |        |       | 0.008 | 0.059 | 1.008 | 1.132 | 0.897 | 0.8985 |
| Unreported  |        |       | 0.045 | 0.050 | 1.046 | 1.153 | 0.948 | 0.3712 |

Note: The response variable is the probability of a positive outcome. The first wave lasts from July to October 2020 and the second, from November 2020 to April 2021. They were defined by lower values of reported cases in external data.13 “Male” is the reference level for sex and “white” for the race. Days of collection starting at July 1, 2020; age measured in years. Logistic regression coefficients (B) are shown with respective standard errors (SE), odds ratios (OR), and 95% confidence limits (CI).

The high percentage of undetected samples for SARS-CoV-2 can have several explanations, such as inappropriate collection, problems of conservation or transport of the sample, collection at an inopportune date, which can lead to false-negative results. Symptomatic negative results could be attributed to other pathogens such as common cold coronaviruses, for example. However, this statement cannot be confirmed, as other pathogens are not tested at IAL Santo André, except Influenza A and B viruses, for all samples of patients with SARS and which were negative for SARS-CoV-2. It was observed that practically all SARS-CoV-2 negative samples were also negative for Influenza viruses (data not shown). The great number of negative results for SARS-CoV-2 can also be attributed to investigations of FS outbreaks in a closed or semi-closed community. In these cases, up to 25 people are evaluated from the surroundings where two suspected or confirmed cases were detected, according to IAL guidelines, which are periodically updated.26 It includes professionals who work in that environment, even if they are asymptomatic so that anyone with the virus can be segregated and isolated. In these outbreak investigations, most of the individuals analyzed showed undetectable results, contributing to the increase in negative monthly samples of the IAL Santo André.

ETHICS STATEMENT
Not applicable. No human sample was collected for the purpose of this study; they were collected for the diagnostic routine of the laboratory. Private information from patients was omitted and data was assessed in a retrospective, aggregated mode, disabling personal identification.

AUTHOR CONTRIBUTIONS
Data collection on the database platform: André Eterovic, Ivana Barros de Campos, and Maria Cecilia Cergole-Novella. Writing, review editing, and original draft preparation: André Eterovic, Andriá Moreira dos Santos Carmo, Elaine Cristina de Mattos, Ivana Barros de Campos, Márcia Aparecida Sperança, and Maria Cecilia Cergole-Novella. Laboratory diagnosis analysis: Daniela Rodrigues Colpas, Ivana Barros de Campos, and Itatiana Ferreira Rodart. Conception and design of the study: André Eterovic, Andriá Moreira dos Santos Carmo, Elaine Cristina de Mattos, Ivana Barros de Campos, Márcia Aparecida Sperança, Maria Cecilia Cergole-Novella, and Vilma dos Santos Menezes Gaiotto Daros. All authors have read and approved the final manuscript.

DATA AVAILABILITY STATEMENT
The data sets used and analyzed during the current study are available from the corresponding author on reasonable request.

ORCID
André Eterovic https://orcid.org/0000-0002-3875-4946
Márcia Aparecida Sperança https://orcid.org/0000-0002-1178-3747
Ivana Barros de Campos https://orcid.org/0000-0003-0334-0572
Andriá Moreira dos Santos Carmo https://orcid.org/0000-0002-0602-4623
Elaine Cristina de Mattos https://orcid.org/0000-0002-1052-8883
Maria Cecilia Cergole-Novella https://orcid.org/0000-0001-9671-825X

REFERENCES
1. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727-733.
2. WHO. Novel Coronavirus (2019-nCoV) Situation Report-1:2020a:5. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf

3. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome‐related coronavirus: classifying 2019‐nCoV and naming it SARS‐CoV‐2. Nat Microbiol. 2020. doi:10.1038/s41564-020-0695-z

4. Chu DKW, Pan Y, Cheng SMS, et al. Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia. Clin Chem. 2020;66:549-555.

5. Garg S, Kim L, Whitaker M, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019—COVID-NET, 14 United States, March 1-30, 2020. Morb Mortal Wkly Rep. 2020;69:458-464.

6. Hooper MW, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA. 2020;323:2466-2467.

7. Ministério da Saúde (MS). Guia de Vigilância Epidemiológica. Emergência de Saúde Pública de Importância Nacional pelo Coronavírus 2020. Vigilância de Síndromes Respiratórias Agudas de Saúde Pública de Importância Nacional pela Doença pelo Coronavírus SARS-COV-2 lineagens a assistir à epidemiologia. https://cidades.ibge.gov.br/brasil/sp/cidades/operacao+-+2020a:5.

8. World Health Organization (WHO). 2020b. https://www.who.int/publications/i/item/diagnostic-testing-for-sars-cov-2

9. O’toule Á, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7:veab064.

10. PANGO Lineages. Latest epidemiological lineages of SARS-CoV-2. https://cov-lineages.org/lineage_list.html

11. Camargo CH, Gonçalves CR, Pagnoca EVRG, et al. A year of the PANGO Lineages. Latest epidemiological lineages of SARS-CoV-2. Nat Microbiol. 2020. doi:10.1038/s41564-020-0695-z

12. Rambaut A, Holmes EC, O’toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403-1407.

13. Brasil IO. Boletins epidemiológicos da COVID-19 por município por dia. https://brasil.io/dataset/covid19

14. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. 2020;25(3):200045.

15. Centers for Disease Control and Prevention (CDC). Division of Viral Diseases. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel. Centers for Disease Control and Prevention; 2020a:1-80. https://www.fda.gov/media/134922/download

16. Centers for Disease Control and Prevention (CDC). Division of Viral Diseases. 2019-Novel Coronavirus (2019-nCoV) Real-time RT-PCR Panel Primers and Probes. Centers for Disease Control and Prevention; 2020b:1-2. https://www.cdc.gov/coronavirus/2019-ncov/downloads/rt-PCR-panel-primer-probes.pdf

17. Global Initiative on Sharing All Influenza Data (GISAID). (https://www.gisaid.org).

18. Burland TG. DNASTAR’s Lasergene sequence analysis software. In: Misener S, Krawetz SA, eds. Bioinformatics Methods andProtocols. Methods in Molecular Biology™. Vol 132. Humana Press; 2000:71-91. doi:10.1385/1-59259-192-2:71

19. Instituto Brasileiro de Geografia e Estatística (IBGE). Portal do Instituto. Disponível em: https://cidades.ibge.gov.br/brasil/sp/panorama

20. Silva SJRD, Pena L. Collapse of the public health system and the emergence of new variants during the second wave of the COVID-19 pandemic in Brazil. One Health. 2021;13:100287. doi:10.1016/j. onehilt.2021.100287

21. News. Welcomed operation: Adolfo Lutz Institute confirms cases of the new variant P.1 of SARS-CoV-2 in soldiers of the Welcomed Operation in Roraima. BEPA. 2021;18(207):54. https://www.saude.sp.gov.br/resources/ccd homepage/bepa/edicoes-2021/edicao-207_do_bepa_-_marco.pdf

22. OPAS/PAHO/WHO. Occurrence of Variants of SARS-CoV-2 in the Americas. PAHO/WHO. 2021. http://bit.ly/2KCKUy

23. Kern DM, Sorum B, Mali SS, et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. Nat Struct Mol Biol. 2021;28:1-10. doi:10.1038/S41594-021-00619-0

24. Siu KL, Yuen KS, Castaño-Rodriguez C, et al. Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. FASEB J. 2019;33(8):8865-8877.

25. Castaño-Rodriguez C, Honrubia JM, Gutiérrez-Álvarez J, et al. Role of severe acute respiratory syndrome coronavirus virions E, 3a, and 8a in replication and pathogenesis. mBio. 2018;9(3):e02325-17.

26. Instituto Adolfo Lutz (IAL). Protocolo laboratorial para coleta, acondicionamento e transporte de amostras biológicas para investigação de SG por SARS-CoV-2. IAL; 2021:1-6. http://www.ial.sp.gov.br/resources/inaludof-lutz/publicacoes/protocolo_laboratorial_para_coleta_sg_covid_23092020.pdf

SUPPORTING INFORMATION
Additional supporting information may be found in the online version of the article at the publisher’s website.