Title: Extensive transmission of SARS-CoV-2 BQ.1* variant in a population with high levels of hybrid immunity: A prevalence survey

Juan P. Aguilar Ticona[1,2,3]*, Meng Xiao[3,4]*, Dan Li[3,5]*, Nivison Nery Jr.[1,2,3], Matt Hitchings[6], Emilia M. M. Andrade Belitardo[2], Mariam O. Fofana[3], Renato Victoriano[2], Jaqueline S. Cruz[2], Laise Eduarda Paixão de Moraes[2], Icaro Morais Strobel[2], Jessica Jesus Silva[2], Ananias Sena do Aragão Filho[2], Guilherme S. Ribeiro[2], Mitermayer G. Reis[2,3,7], Federico Costa[1,2,3], Ricardo Khouri[2], Albert I. Ko[2,3]*, Derek A.T. Cummings[8,9]*

1. Instituto de Saúde Coletiva, Universidade Federal da Bahia, Salvador, BA, Brazil
2. Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Ministério da Saúde, Salvador, BA, Brazil
3. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut, United States
4. Department of Laboratory Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
5. Public Health Emergency Center, Chinese Center for Disease Control and Prevention, Beijing, China
6. Department of Biostatistics, University of Florida, Gainesville, FL, United States
7. Faculdade de Medicina da Bahia, Universidade Federal da Bahia, Salvador, BA, Brazil
8. Department of Biology, University of Florida, Gainesville, FL, United States
9. Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States

* These authors contributed equally to this article.

***Corresponding author:** Juan P. Aguilar Ticona, Instituto de Saúde Coletiva, Universidade Federal da Bahia R. Basílio da Gama, s/n - Canela, Salvador - BA, 40110-040; email: pkjpablo@gmail.com (JPAT)

Abstract

Background: The SARS-CoV-2 BQ.1* variant rapidly spread globally in late 2022, posing a challenge due to its increased immune evasion.

Methods: We conducted a prevalence survey in Brazil from November 16th to December 22nd, 2022, as part of a cohort study. We conducted interviews and collected nasal samples for RT-PCR testing and whole-genome sequencing. Cumulative incidence was estimated using RT-PCR positivity, cycle threshold values, and external data on the dynamics of RT-PCR positivity following infection.

Results: Among 535 participants, 54% had documented SARS-CoV-2 exposure before this outbreak and 74% had received COVID-19 vaccination. In this study, 14.8% tested positive for SARS-CoV-2, with BQ.1* identified in 90.7% of cases. Using case data and cycle threshold values, cumulative incidence was estimated at 56% (95%CI, 36-88%). Of the 79 positive participants, 48.1% had a symptomatic illness, with a lower proportion fulfilling the WHO COVID-19 case definition compared to prior Omicron waves. No participants required medical attention.

Conclusion: Despite high population-level hybrid immunity, the BQ.1* variant attacked 56% of our population. Lower disease severity was associated with BQ.1* compared to prior Omicron variants. Hybrid immunity may provide protection against future SARS-CoV-2 variants but in this case was not able to prevent widespread transmission.

Keywords: SARS-CoV-2; BQ.1 variant; high incidence; hybrid immunity

Introduction

The Omicron variant of SARS-CoV-2 has been characterized by high levels of immune evasion [1]. The most recently emerged subvariants, BQ.1.1 and XBB, have been shown to effectively evade immunity generated by vaccines, including bivalent formulations designed specifically to target Omicron BA.5 [1-3]. In addition to diminishing vaccine effectiveness, the continued evolution of Omicron variants may limit the utility of available treatment options such as Nirmatrelvir/ritonavir or molnupiravir [4, 5]. Moreover, changes in the clinical spectrum of disease may result in biased estimates of transmission from symptom-based surveillance [6, 7].

Laboratory studies have identified mutations that confer twice as much immune evasion in BQ.1 and BQ.1.1 subvariants (here referred to collectively as BQ.1*) compared to the BA.4 and BA.5 subvariants [8]. However, it remains unknown how much BQ.1* associated immune evasion affects transmission among populations with pre-existing immunity, especially those with hybrid immunity (immunity due to exposure to both infection and vaccination). Prior studies of transmission during the circulation of the omicron BA.1 subvariant demonstrated a high incidence of reinfection and breakthrough infections among vaccinated individuals, and the degree of protection conferred by prior infection and vaccination is known to decline over time [9-11].

This study aims to estimate the incidence of PCR-confirmed infection with the SARS-CoV-2 Omicron BQ.1* subvariant in a population in Salvador, Brazil with high prevalence of hybrid immunity. We performed a population-based prevalence survey of SARS-CoV-2 infection using molecular diagnostics and whole-genome sequencing and applied novel computational approaches to infer the incidence of infection using the distribution of PCR cycle threshold (Ct) values [12]. We also used the Ct values of

samples to gain broader insights on BQ.1* transmission. We compared the severity of illness associated with BQ.1* infection to other Omicron variants, estimated household secondary attack rate and examined risk factors associated with acquisition of infection.

Methods

Setting and Study Design

This study was conducted in Salvador, the capital of the state of Bahia, Brazil, which has experienced five major COVID-19 waves since early 2020 (Figure 1A), with the three most recent waves in 2022 driven by Omicron subvariants (Figure 1B). The first Omicron wave occurred from January to March 2022, mainly attributed to BA.1* and BA.2* subvariants, while the second wave, from June to September 2022, was attributed to BA.4* and BA.5*. The third wave in November 2022 was predominantly due to the BQ.1* subvariant (Figure 1B). By November 2022, 86% of Salvador's residents had received at least one dose of a COVID-19 vaccine, and 75% had received at least two doses (Figure 1C).

In this context, we conducted a population-based prevalence survey in Pau da Lima, a slum community in Salvador. This community is in an area of 0.17 km2. Approximately 85% of inhabitants were squatters without legal title to their homes, and 50% had aper capita household income of less than $1.25 per day. In 2003, an open cohort study was initiated in the area to investigate infectious diseases including leptospirosis and arbovirus infections, with bi-annual or annual follow-up. After 2020, the study's scope was expanded to include COVID-19 studies [13]. Three serosurveys conducted from November 2020 to August 2022 (Surveys 1-3) showed an increase in seropositivity (tested by SARS-CoV-2 anti-S IgG) among participants (Figure 1D). An

active case-finding study between November 2021 to October 2022 identified symptomatic SARS-CoV-2 cases and their contacts in the same area (Figure 1E). During the active case finding study, our field teams visited study households every two weeks to screen residents for symptoms and collect nasal swabs for SARS-CoV-2 molecular diagnostics.

The results of this work are part of the fourth COVID-19 survey conducted in the cohort between November 16 and December 22, 2022. During this period, Salvador, as well as the Pau da Lima community, experienced a high increase in the number of cases associated with the transmission of the SARS-CoV-2 variant BQ.1.

Participants and Study Procedures

We included individuals aged 2 years or older who slept at least 3 nights per week within the study area and provided consent to participate. Field technicians performed data collection, including interviews and collecting biological samples. After obtaining informed consent, a standardized questionnaire was administered to collect sociodemographic information (age, sex, schooling, self-reported ethnicity, and income), COVID-19 symptoms, and vaccination history. Symptomatic individuals were defined as those who reported any of the following symptoms in the week preceding or during the visit: fever, cough, fatigue, headache, myalgia, sore throat, congestion or runny nose, dyspnea, nausea, diarrhea, anorexia, loss of taste, loss of smell or mental state altered [14]. Each participant provided an anterior nasal swab for SARS-CoV-2 molecular testing, and symptomatic cases and their household contacts were administered a rapid antigen test during the initial visit. Positive cases were immediately informed, and healthcare assistance recommendations were given.

Laboratory Examination of SARS-Cov-2 Infection

Real-time reverse transcription-polymerase chain reaction (RT-PCR) was conducted to confirm SARS-CoV-2 infection, and the PCR Ct values for the ORF1ab gene were recorded for positive samples. Next-generation sequencing (NGS) using the Illumina method was performed on positive samples to identify variants of concern (VOCs) and/or variants of interest (VOIs). Both molecular diagnostic tests were conducted by the COVID Platform of Fiocruz-Bahia, Brazil. For phylogenetic analysis, Omicron lineage sequences collected during both the active case-finding and prevalence survey periods from Pau da Lima were compared with SARS-CoV-2 Omicron variant data from Salvador, Brazil, obtained from the GISAID database between January 1, 2022, and December 31, 2022 (see details in Supplementary Material 1).

Data Analysis: Descriptive analyses

To describe the characteristics of the study participants, we used absolute frequencies and percentages for categorical variables and median and interquartile range (IQR) for numeric variables. We compared continuous variables with Mann-Whitney U and categorical variables with Fisher's exact test or chi-square test as appropriate, and linear by linear chi-square tests for ordinal categorical variables. Statistical analysis was performed using R Statistical Software version 3.1.6.

Data Analysis: Estimation of Cumulative Incidence

We used a method for estimating the epidemic growth rate of SARS-CoV-2 using RT-PCR Ct values, as previously described by Hay et al. [12]. Briefly, the daily prevalence of RT-PCR positivity together with the Ct values among RT-PCR positive samples was used to estimate the daily probability of infection. To ensure we used only tests that

represented a random sample of individuals with respect to infection risk, we excluded tests collected at the Day 7 follow-up visit. To inform the distribution of Ct following infection we used published data on Omicron infections (see details in Supplementary Material 1). The assumed Ct distribution over time since infection was consistent with the observed Ct over time from symptom onset observed in symptomatic individuals in our population (Supplementary Figure 3). We estimated the overall cumulative incidence of infection from October 19, 2022 to December 22, 2022. Based on available literature, the probability of testing RT-PCR positive 28 days after Omicron infection is small [15, 16], meaning that Ct values we measured provided no information about incidence before October 19. Using the estimated incidence over time, we estimated the day of peak incidence, as well as RT-PCR positivity prevalence by week to assess goodness of fit.

We conducted sensitivity analyses to compare the recruited and nonrecruited participants, to determine the robustness of our sample to identify PCR-positive participants in our cohort that was used to estimate the incidence. Also, we performed sensitivity analyses to check the robustness of the cumulative incidence estimate to changes in the CT distribution and PCR positivity probability over time (Supplementary Material 1).

Data Analysis: Symptom Evaluation

We compared the frequency of reported symptoms and medical attention between participants identified through active case finding during a period dominated by BA.1 and BA.5 variants, and those recruited in the current survey. Symptomatic cases were identified based on any symptom associated with COVID-19, as mentioned previously in the text. Then, we proceeded to evaluate the proportion of infections meeting WHO's

definition of a symptomatic case, which includes an acute onset of fever and cough, or three or more of the following symptoms: fever, cough, weakness/fatigue, headache, myalgia, sore throat, coryza, dyspnea, nausea, diarrhea, and anorexia [14].

Data Analysis: Secondary Attack Rate

To estimate the secondary attack rate (SAR), we defined the index case as the individual with the earliest positive COVID-19 test or symptom onset. Co-index cases were two or more household members who tested positive or had symptom onset on the same date. One co-index case was selected randomly as the index case to calculate the SAR. Household contacts were individuals who lived in the same household as the index case within 7 days after the positive PCR test result or onset of symptoms. A secondary case was a household contact who tested positive for SARS-CoV-2. The SAR was calculated by dividing the number of secondary cases by the total number of non-index household residents. Additionally, using the imputed datasets generated to evaluate the sample selection, we estimated the SAR for the entire cohort and compared it with the observed result.

Secondary Data Resource

To describe the context of the SARS-CoV-2 transmission in Salvador, we used data on daily infections and deaths in Salvador and the Pau da Lima sanitary district since the beginning of the pandemic from the Brazil Ministry of Health (https://covid.saude.gov.br) and the Center for Strategic Information for Health Surveillance (CIEVS) (http://www.cievs.saude.salvador.ba.gov.br/), respectively. The prevalence of SARS-CoV-2 variants in Salvador over time was obtained from the Fiocruz COVID-19 Genomic Surveillance Network (https://pvm-igm.github.io), while

data on vaccination were obtained from the Brazil Ministry of Health (https://opendatasus.saude.gov.br/).

Results

Participants Characteristics

We surveyed 293 households, totaling 929 residents, with 535 meeting the inclusion criteria and participating in the study by completing questionnaires and providing biological samples. The remaining 378 residents were excluded due to reasons such as moving out, absence during visits, or declining to participate. Additionally, 16 residents were excluded due to invalid PCR results (Supplementary Figure 1).

Sociodemographic characteristics of the participants are presented in Table 1 according to the SARS-CoV-2 immunological status. Briefly, 57.9% (310/535) were female, and the median age was 32 years (IQR 16-47 years). 49.0% (262/535) self-identified as black, and 46.7% (250/535) reported an income below the international poverty line (US$2.15 per person per day). Overall, 95.8% (518/535) of participants have received at least one dose of a COVID-19 vaccine or have a history of SARS-CoV-2 infection.

Crude and Variant Specific Prevalence

A total of 79 cases of SARS-CoV-2 were identified, with an overall crude prevalence of 14.8% (95% CI 11.8% – 17.8%) (Supplementary Figure 1). Among 58 positive RT-PCR samples analyzed using NGS, 15 cases (25.9%) could not be classified at the subvariant level due to low genome coverage (<70%). Among the remaining 43 cases, BQ.1* was detected in 39 cases (90.7%). Of these, 30 cases (69.8%) were BQ.1.1, 8 cases (18.6%) were BQ.1, and 1 case (2.3%) was BQ.1.22. The BA.5.1 and BE.9

Omicron subvariants each accounted for 1 case (2.3%), while 2 cases (4.7%) of XBB.1 were identified during the second week of sample collection (Figure 2A).

Phylogenetic analysis included 1263 samples from Salvador, 88 from the previous active case finding period, and 43 from the present survey. Viruses from Pau da Lima and Salvador were closely related, and no genetic clustering within these two geographic areas was identified. Like Pau da Lima, the circulation of XBB in Salvador was lower than that of BQ.1 during the study period (Figure 3).

Prevalence Over Time and Cumulative Incidence

Figure 2B shows the distribution of Ct values in SARS-CoV-2 RT-PCR-positive samples by epidemiological week. Lower Ct values were observed in the first two weeks, with a subsequent increase in the following three weeks. These changes matched the observed weekly SARS-CoV-2 prevalence trends, which peaked at 32.1% in the second week (Nov 23 – 29th, 2022) before gradually decreasing (Figure 2C). The prevalence trends in the study population were consistent with data from the Pau da Lima sanitary district, with a two-week lag for the peak of the BQ.1 wave compared to Salvador's overall peak (Figure 2D and 2E).

The estimated cumulative incidence of infection from October 19 to December 22, 2022, was 56% (95% CrI = 36 to 88%), with the peak incidence on November 17th (95% CrI = 9th to 21st) during the first sampling week (Figure 2F). Due to the lag between incident infections and viral clearance, the peak of the estimated incidence curve appeared earlier than the peak of observed prevalence (Figures 2E and 2F). The overall RT-PCR positivity was well-fitted by the model (11.7% vs. observed 12.4%

among individuals swabbed at the initial household visit), but the peak prevalence in week 47 was underestimated (22.0% vs. observed 32.1%) (Figure 2C).

In additional analyses, using multiple imputations to account for missing data and estimate the proportion of participants PCR-positive (Supplementary Table 4 – 6), our findings remained unchanged. In sensitivity analysis varying key features of the assumed Ct distribution over time following infection, the estimated cumulative incidence ranged from 49% to 62% (Supplementary Table 7).

Clinical Symptoms and Medical Attention After Infection

Clinical symptoms were assessed in 38 (48.1%, 95% CI = 37.1 – 59.1) SARS-CoV-2-positive symptomatic individuals during the BQ.1* wave and compared to 103 positive cases from prior Omicron waves. Rhinorrhea was the most frequently reported symptom during the BQ.1* wave (78.9%), followed by cough, headache, and sore throat, each reported by more than 50% of participants (Table 2). The number of symptoms reported was similar between the BQ.1* wave and previous waves. However, individuals in the BQ.1* wave were more likely to report shortness of breath (47.4% vs. 14.6%, $p < 0.001$) and less likely to report diarrhea (2.6% vs. 16.5%, $p = 0.043$) (Table 2). Additionally, the proportion of symptomatic cases meeting the WHO definition criteria was significantly lower during the BQ.1* wave compared to previous waves (47.4% vs. 69.9%, $p = 0.023$) (Table 2). None of the SARS-CoV-2-positive individuals during the BQ.1* wave required medical attention, in contrast to 3.8% (95%CI = 1.2 – 9.1%) during previous waves (Table 2).

Household Secondary Attack Rate

Among 54 households with at least one confirmed case of SARS-CoV-2 we selected 115 residents from 35 households with more than one resident to estimate the secondary attack rate (SAR). Among these participants, 35 were classified as index cases, 25 were secondary cases, and 55 were negative contacts (Supplementary Figure 2). The crude SAR was 31.3% (95% CI = 22.2 – 42.1), and other SARs stratified by non-index characteristics are presented in Supplementary Table 1. Individuals under 18 were more likely to be secondary cases compared to those 18 and older (RR = 2.03, 95% CI = 1.04 – 3.95). Using multiple imputations to estimate the SAR in the entire cohort while considering the household number of residents distribution, sex, age, vaccination, and previous participation in the previous survey, we found a low SAR in households with a high number of residents (Supplementary Figure 4). However, the 95% CIs from the observed and estimated data overlap in both the pooled SAR and the stratified analysis by household number of residents (Supplementary Figure 4 and Supplementary Table 8).

Documented prior exposure

A detailed description of the evolution of seroprevalence and vaccination in the cohort before the outbreak described here, aimed at understanding hybrid immunity in this community, is provided in Supplementary Table 2. However, due to uncertainties regarding seropositivity associated with vaccination or infection, and the loss of follow-up, it was not possible to clearly define prior exposure associated with either or both once vaccination became available (after survey 1). The evaluation of risk factors associated with BQ.1 PCR positivity is outlined in Supplementary Table 3. We identified a signal of protection (OR= 0.50; 95%CI = 0.25 – 0.97) suggesting that previous infection during Survey 1, conducted from November 2020 to February 2021,

may serve as a proxy for a potentially lower risk of reinfection during the subsequent months until the BQ.1* outbreak in this community.

Discussion

We describe a rapid and large outbreak predominantly caused by BQ.1* that we estimated affected 56% (95% CrI = 36 to 88%) of individuals in our population over five weeks. Our population was previously highly exposed with 97% having detectable immunity to SARS-CoV-2 from prior infection and/or high rates of vaccination before the outbreak we describe here. Our findings highlight that even populations in which a high proportion of individuals have been previously infected and/or vaccinated can experience substantial outbreaks of BQ.1* [17, 18]. During the study period, BQ.1* was the most prevalent variant (90.7%) compared to XBB. This differs from other regions such as Singapore and India [19-22] where XBB emerged as the most common variant at the end of 2022. While BQ.1 remained the predominant variant in the US and Europe until the last weeks of 2022, increasing trends of XBB have been observed in these regions. In the first and sixth week of 2023, XBB became the most prevalent variant in the US and Europe, respectively [22, 23], while incidence of XBB in Brazil remained low. The mechanisms driving emergence of one strain over the other are not understood [23].

Although this population had a high incidence of infection, medically attended illness rates were extremely low. Compared to a previous period of BA.1 predominance, fewer individuals met WHO clinical diagnosis criteria during the BQ.1* wave. This change in symptom presentation may lead to an underestimation of BQ.1* incidence from surveillance based on clinical criteria. Similar shifts in symptom patterns were observed during the previous BA.1 and BA.2 transmission periods compared to the Delta variant

[7]. Additionally, during the Omicron BA.1 period, there was a decrease in the severity of symptoms, hospitalizations, and deaths compared to pre-Alpha variants and the displaced Delta variant [24, 25]. This difference could be due to the high prior exposure [13], changes in health seeking behavior or intrinsic differences between viral lineages. While PCR tests were useful in identifying cases during epidemic SARS-CoV-2 waves, they may not be affordable for community-based surveys, particularly in resource-limited settings. Therefore, it is important to update diagnostic algorithms that consider the presence and combination of symptoms associated with the emergence of new variants.

We found some evidence that immune status was linked to the risk of RT-PCR detected infection in this population. Individuals who were first infected before the first round of surveys (before November 2020) had a reduced risk of infection. As these people had the greatest opportunity to acquire multiple infections, our results suggest that people who were frequently exposed to SARS-CoV-2 may accumulate protective immunity from multiple prior exposures [19, 26, 27]. Low rates of reporting to national surveillance systems over time mean that cohort studies will become increasingly relied upon to understand immunity to SARS-CoV-2. Such studies should measure immune status, exposure history, and detect incident infections. Assessing COVID-19 transmission through serosurveys can be challenging for open cohorts that may face issues such as loss to follow-up and incomplete registration. Additionally, the presence of vaccines can complicate the interpretation of serological results as they may reflect either infection or vaccination. Here, we use novel methods to integrate PCR confirmed infections with Ct values to reconstruct the dynamics of infection in this cohort. Due to the challenge of identifying cases through passive surveillance, future studies, including ours, will need to integrate multiple sources of information to characterize the dynamics

of infections in populations. We identified a high secondary transmission rate of 31.3% (95% CI 22.2 – 42.1). While there are no epidemiological studies that confirm the increased infectiousness of the Omicron BQ.1 variant, we used insights from previous variants, such as BA.1 and BA.2, to contextualize our findings [28, 29]. It has been reported that the Omicron variant is associated with a ~50% secondary household transmission [29, 30]. The high attack rate observed in our study underscores the urgent need to implement prevention measures in addition to vaccine campaigns to limit transmission.

We acknowledge limitations in our study. Firstly, the study was conducted during the peak of the outbreak, which may limit our ability to fully characterize the outbreak. Although we estimated cumulative incidence, the uncertainty during the pre-study recruitment period is reflected in the wide 95% credible interval during this period, and our estimate relied on a small number of studies measuring RT-PCR positivity over time following an Omicron infection. Secondly, as described above there was likely misclassification in our identification of prior SARS-CoV-2 exposure using previous serosurveys. Moreover, the use of RT-PCR positivity as the outcome of interest in our regression analysis likely induced misclassification of the outcome of interest (i.e. infection during the outbreak). Thirdly, self-reported data were used to evaluate symptoms, which may have introduced recall bias. Finally, our assumption that all secondary cases within a household were infected by the primary case in the SAR analysis was a simplification and did not account for infections acquired outside of the household. Additionally, as a prevalence survey study, our estimates of incidence outside the study period were moderately sensitive to model assumptions in a sensitivity analysis. Finally, we assumed that symptomatic and asymptomatic individuals had the same Ct distribution following infection, which may have biased our estimate of

cumulative incidence. The direction of bias depends on which features of the Ct curve differ by symptom status.

Our findings emphasize the importance of monitoring new variants and their clinical outcomes during the ongoing COVID-19 pandemic. Utilization of new tools, such as mathematical modeling and phylogenetic analysis can improve outbreak characterization and allow for continued monitoring of incidence as the COVID-19 outbreak continues.

Acknowledgments

The authors would like to thank the members of the Pau da Lima community in Salvador, Brazil for their support. We would also like to thank team members from the Collective Health Institute Oswaldo Cruz Foundation, Yale University and University of Florida.

Conflict of Interest

A.I.K serves as an expert panel member for Reckitt Global Hygiene Institute, scientific advisory board member for Revelar Biotherapeutics and a consultant for Tata Medical and Diagnostics and Regeneron Pharmaceuticals, and has received grants from Merck, Regeneron Pharmaceuticals and Tata Medical and Diagnostics for research related to COVID-19, all of which are outside the scope of the submitted work. D.A.T.C. has received a grant from Merck for research unrelated to COVID-19, outside of the scope of this work. Other authors declare no competing interests.

Funding Source

This work was supported by grants from the National Institutes of Health (https://www.nih. gov; R01 AI052473, U01AI088752, R01 TW009504 and R25

TW009338 to A.I.K), the UK Medical Research Council (https://mrc.ukri.org; MR/T029781/1 to F.C.), the Wellcome Trust (https://wellcome.org; 102330/Z/13/Z; 218987/Z/19/Z to F.C.), the Bill and Melinda Gates Foundation (https://www.gatesfoundation.org; OPP1211988 to M.G.R. and F.C.), the Conselho Nacional de Desenvolvimento Científico e Tecnológico [Brazilian National Council for Scientific and Technological Development] (https://www.gov.br/cnpq/pt-br; CNPq 311365/2021-3 to G.S.R.), the Fundação de Amparo à Pesquisa do Estado da Bahia [Bahia Foundation for Research Support] (http://www.fapesb.ba.gov.br; FAPESB SUS0019/2021 and PET0022/2016 to G.S.R.), the Burroughs-Wellcome Fund (https://www.bwfund.org; ASTMH Postdoctoral Fellowship to M.O.F.), the William H. Prusoff Foundation (Postdoctoral Fellowship to M.O.F.), a US NSF RAPID award (to M.D.T.H. and D.A.T.C.), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil [Coordination for the Improvement of Higher Education Personnel] (Doctoral scholarship to J.P.A.T., Finance Code 001) and the Sendas Family Fund at the Yale School of Public Health (to A.I.K.). This funding source had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, decision to publish, or preparation of the manuscript.

Ethical Approval statement

The study was approved by the Ethics Committee of the Institute of Collective Health (35405320.0.1001.5030), the Institutional Review Boards of the Instituto Gonçalo Moniz, Oswaldo Cruz Foundation (Fiocruz) and the Brazilian National Commission for Ethics in Research (CAAE 45217415.4.0000.0040, 35405320.0.1001.5030 and 59889922.6.0000.0040), and the Yale University Human Research Protection Program (2000031554).

406

Reference

[1] Davis-Gardner ME LL, Wali B, et al. . Neutralization against BA.2.75.2, BQ.1.1, and XBB from mRNA Bivalent Booster. N Engl J Med 2023;388(2):183-5.

[2] Kurhade C, Zou J, Xia H, Liu M, Chang HC, Ren P, et al. Low neutralization of SARS-CoV-2 Omicron BA.2.75.2, BQ.1.1 and XBB.1 by parental mRNA vaccine or a BA.5 bivalent booster. Nat Med. 2022. Epub 2022/12/07. doi: 10.1038/s41591-022-02162-x.

[3] Hoffmann M, Behrens GMN, Arora P, Kempf A, Nehlmeier I, Cossmann A, et al. Effect of hybrid immunity and bivalent booster vaccination on omicron sublineage neutralisation. Lancet Infect Dis. 2023;23(1):25-8. Epub 2022/12/09. doi: 10.1016/S1473-3099(22)00792-7.

[4] Arora P, Kempf A, Nehlmeier I, Schulz SR, Jäck H-M, Pöhlmann S, et al. Omicron sublineage BQ.1.1 resistance to monoclonal antibodies. The Lancet Infectious Diseases. 2023;23(1):22-3. doi: 10.1016/s1473-3099(22)00733-2.

[5] .Imai M IM, Kiso M, et al. Efficacy of Antiviral Agents against Omicron Subvariants BQ.1.1 and XBB. N Engl J Med 2023;2023;388(1)::89-91.

[6] Sigal A, Milo R, Jassat W. Estimating disease severity of Omicron and Delta SARS-CoV-2 infections. Nature Reviews Immunology. 2022;22(5):267-9. doi: 10.1038/s41577-022-00720-5.

[7] Vihta K-D, Pouwels KB, Peto TE, Pritchard E, House T, Studley R, et al. Omicron-associated changes in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) symptoms in the United Kingdom. Clinical Infectious Diseases. 2023;76(3):e133-e41.

[8] Qu P, Evans JP, Faraone JN, Zheng YM, Carlin C, Anghelina M, et al. Enhanced neutralization resistance of SARS-CoV-2 Omicron subvariants BQ.1, BQ.1.1, BA.4.6, BF.7, and BA.2.75.2. Cell Host Microbe. 2023;31(1):9-17 e3. Epub 2022/12/09. doi: 10.1016/j.chom.2022.11.012.

[9] Bobrovitz N, Ware H, Ma X, Li Z, Hosseini R, Cao C, et al. Protective effectiveness of previous SARS-CoV-2 infection and hybrid immunity against the omicron variant and severe disease: a systematic review and meta-regression. Lancet Infect Dis. 2023. Epub 2023/01/22. doi: 10.1016/S1473-3099(22)00801-5.

[10] Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman LS, Ash N, et al. Protection and Waning of Natural and Hybrid Immunity to SARS-CoV-2. N Engl J Med. 2022;386(23):2201-12. Epub 2022/05/26. doi: 10.1056/NEJMoa2118946.

[11] Chemaitelly H AH, Tang P, et al. . Immune Imprinting and Protection against Repeat Reinfection with SARS-CoV-2. . N Engl J Med. 2022;387(2022;387(18)):1716-8. Epub 2022/11/03. doi: 10.1056/NEJMc2210175.

[12] Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. Science. 2021;373(6552). Epub 2021/06/05. doi: 10.1126/science.abh0635.

[13] Fofana MO, Nery N, Jr., Aguilar Ticona JP, de Andrade Belitardo EMM, Victoriano R, Anjos RO, et al. Structural factors associated with SARS-CoV-2 infection risk in an urban slum setting in Salvador, Brazil: A cross-sectional survey. PLoS Med. 2022;19(9):e1004093. Epub 2022/09/09. doi:

10.1371/journal.pmed.1004093.
[14] World Health Organization. WHO COVID-19 case definition. World Health Organization, 2020.
[15] Hay JA, Kissler SM, Fauver JR, Mack C, Tai CG, Samant RM, et al. Quantifying the impact of immune history and variant on SARS-CoV-2 viral kinetics and infection rebound: A retrospective cohort study. eLife. 2022;11:e81849. doi: 10.7554/eLife.81849.
[16] Boucau J, Marino C, Regan J, Uddin R, Choudhary MC, Flynn JP, et al. Duration of Shedding of Culturable Virus in SARS-CoV-2 Omicron (BA.1) Infection. New England Journal of Medicine. 2022;387(3):275-7. doi: 10.1056/NEJMc2202092.
[17] Cui Z, Liu P, Wang N, Wang L, Fan K, Zhu Q, et al. Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. Cell. 2022;185(5):860-71.e13. Epub 20220125. doi: 10.1016/j.cell.2022.01.019.
[18] Shrestha LB, Foster C, Rawlinson W, Tedla N, Bull RA. Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. Rev Med Virol. 2022;32(5):e2381. Epub 20220720. doi: 10.1002/rmv.2381.
[19] Tan CY, Chiew CJ, Pang D, Lee VJ, Ong B, Lye DC, et al. Protective Effectiveness of Natural SARS-CoV-2 Infection and Vaccines against Omicron BA. 4/BA. 5 and XBB Reinfection in Singapore: A National Cohort Study. Available at SSRN 4308740. 2022.
[20] Scarpa F, Sanna D, Azzena I, Casu M, Cossu P, Fiori PL, et al. Genome-based comparison between the recombinant SARS-CoV-2 XBB and its parental lineages. bioRxiv. 2022:2022.12.20.521197. doi: 10.1101/2022.12.20.521197.
[21] Karyakarte R, Das R, Dudhate S, Agarasen J, Pillai P, Chandankhede P, et al. Clinical Characteristics and Outcomes of Laboratory-Confirmed SARS-CoV-2 Cases Infected with Omicron subvariants and XBB recombinant variant. medRxiv. 2023:2023.01.05.23284211. doi: 10.1101/2023.01.05.23284211.
[22] Wang Q, Iketani S, Li Z, Liu L, Guo Y, Huang Y, et al. Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants. Cell. 2023;186(2):279-86 e8. Epub 2022/12/30. doi: 10.1016/j.cell.2022.12.018.
[23] Centers for Disease Control and Prevention. COVID Data Tracker. Atlanta, GA: US Department of Health and Human Services, CDC; 2023, February 12. https://covid.cdc.gov/covid-data-tracker.
[24] Mattiuzzi C, Lippi G. Timeline analysis of clinical severity of COVID-19 in the general population. European Journal of Internal Medicine. doi: 10.1016/j.ejim.2022.12.007.
[25] Madhi SA, Kwatra G, Myers JE, Jassat W, Dhar N, Mukendi CK, et al. Population Immunity and Covid-19 Severity with Omicron Variant in South Africa. N Engl J Med. 2022;386(14):1314-26. Epub 20220223. doi: 10.1056/NEJMoa2119658.
[26] Chemaitelly H, Ayoub HH, Coyle P, Tang P, Yassine HM, Al-Khatib HA, et al. Protection of Omicron sub-lineage infection against reinfection with another Omicron sub-lineage. Nature Communications. 2022;13(1):4675. doi: 10.1038/s41467-022-32363-4.
[27] Chemaitelly H, Tang P, Coyle P, Yassine HM, Al-Khatib HA, Smatti MK, et al. Protection against Reinfection with the Omicron BA.2.75 Subvariant. New England Journal of Medicine. 2023;388(7):665-7. doi:

10.1056/NEJMc2214114.
[28] Jørgensen SB, Nygård K, Kacelnik O, Telle K. Secondary Attack Rates for Omicron and Delta Variants of SARS-CoV-2 in Norwegian Households. JAMA. 2022;327(16):1610-1. doi: 10.1001/jama.2022.3780.
[29] Kim EY, Choe YJ, Park H, Jeong H, Chung JH, Yu J, et al. Community Transmission of SARS-CoV-2 Omicron Variant, South Korea, 2021. Emerg Infect Dis. 2022;28(4):898-900. Epub 20220216. doi: 10.3201/eid2804.220006.
[30] Baker JM, Nakayama JY, O'Hegarty M, McGowan A, Teran RA, Bart SM, et al. SARS-CoV-2 B.1.1.529 (Omicron) Variant Transmission Within Households - Four U.S. Jurisdictions, November 2021-February 2022. MMWR Morb Mortal Wkly Rep. 2022;71(9):341-6. Epub 20220304. doi: 10.15585/mmwr.mm7109e1.

Table 1. Demographic and SARS-CoV-2 immunological characteristics of participants, Salvador, Brazil.

Characteristics	No. (%)		p value
	SARS-CoV-2 positive n = 79	SARS-CoV-2 negative n = 456	
Sex			0.426
Female	49 (62.0)	261 (57.2)	
Male	30 (38.0)	195 (42.8)	
Age group, y			0.941
<18	24 (30.4)	126 (27.6)	
18-35	21 (26.6)	131 (28.7)	
36-59	26 (32.9)	147 (32.2)	
≥60	8 (10.1)	52 (11.4)	
Ethnicity [a]			0.384
Black	35 (45.5)	227 (50.0)	
Brown	35 (45.5)	203 (44.7)	
Other	7 (9.1)	24 (5.3)	
Education [a]			0.104
Never studied	7 (8.9)	19 (4.2)	
Primary and middle school	51 (64.6)	276 (60.9)	
High school and higher	21 (26.6)	158 (34.9)	
Income category			0.105
<US$2.15/day	42 (53.2)	208 (45.6)	
US$2.15-3.63 /day	16 (20.3)	84 (18.4)	
>US$3.63 /day	21 (26.6)	164 (36.0)	
Prior vaccination			0.516
≥3 doses	42 (53.2)	220 (48.2)	
2 doses	17 (21.5)	124 (27.2)	
1 dose	11 (13.9)	36 (7.9)	
0 dose	9 (11.4)	76 (16.7)	
Prior documented SARS-CoV-2 exposure			0.004
Yes [B]	31 (39.2)	258 (56.6)	
No	48 (60.8)	198 (43.4)	
Prior documented SARS-CoV-2 exposure and vaccination [c]			0.032
Yes	24 (30.4)	197 (43.2)	
No	55 (69.6)	259 (56.8)	
Prior documented SARS-CoV-2 exposure or vaccination [d]			>0.999
Yes	77 (97.5)	441 (96.7)	
No	2 (2.5)	15 (3.3)	

[a] There were two and three individuals having missing values of their ethnicity and education, respectively, in the SARS-CoV-2 negative group.
[b] A SARS-CoV-2 seroconversion observed before the first dose of vaccination, or previous molecular confirmed infection during active case finding.
[c] "Yes" indicates individuals with ≥1 dose of vaccination and evidence of prior exposure at the

same time; "No" indicates individuals without prior vaccination or without evidence of prior exposure.

[d] "Yes" indicates individuals with ≥1 dose of vaccination or evidence of prior exposure; "No" indicates individuals without vaccination and without having evidence of prior exposure.

Table 2. Symptoms and severity outcomes of symptomatic SARS-CoV-2 positive participants during the BQ.1 wave versus in previous omicron waves, Salvor, Brazil.

Characteristics	No. (%) or median (IQR)		p-value
	SARS-CoV-2 positive in the BQ.1 survey [a] n = 38	SARS-CoV-2 positive in the active case finding [a] n = 103	
No. of symptoms	4.0 (2.0-6.0)	4.0 (2.5-6.5)	0.590
Frequency of symptoms			
Rhinorrhea	30 (78.9)	70 (68.0)	0.287
Cough	24 (63.2)	78 (75.7)	0.205
Headache	19 (50.0)	64 (62.1)	0.268
Sore throat	19 (50.0)	58 (56.3)	0.633
Short of breath	18 (47.4)	15 (14.6)	<0.001
Fever	14 (36.8)	52 (50.5)	0.211
Fatigue	10 (26.3)	26 (25.2)	1
Shiver	7 (18.4)	20 (19.4)	1
Myalgia	6 (15.8)	29 (28.2)	0.198
Anorexia	4 (10.5)	15 (14.6)	0.781
Loss of taste	4 (10.5)	10 (9.7)	1
Loss of smell	3 (7.9)	8 (7.8)	1
Diarrhea	1 (2.6)	17 (16.5)	0.043
Nausea	1 (2.6)	12 (11.7)	0.186
Mental state altered	1 (2.6)	2 (1.9)	1
Other symptoms [b]	4 (10.5)	7 (6.8)	0.705
Meet the WHO COVID-19 case definition [c]	18 (47.4)	72 (69.9)	0.023
Healthcare need			
Medical attention, n (%)	0 (0)	4 (3.8)	0.567
Urgent care visit, n (%)	0 (0)	3 (2.7)	0.773
Hospitalization, n (%)	0 (0)	0 (0)	NA

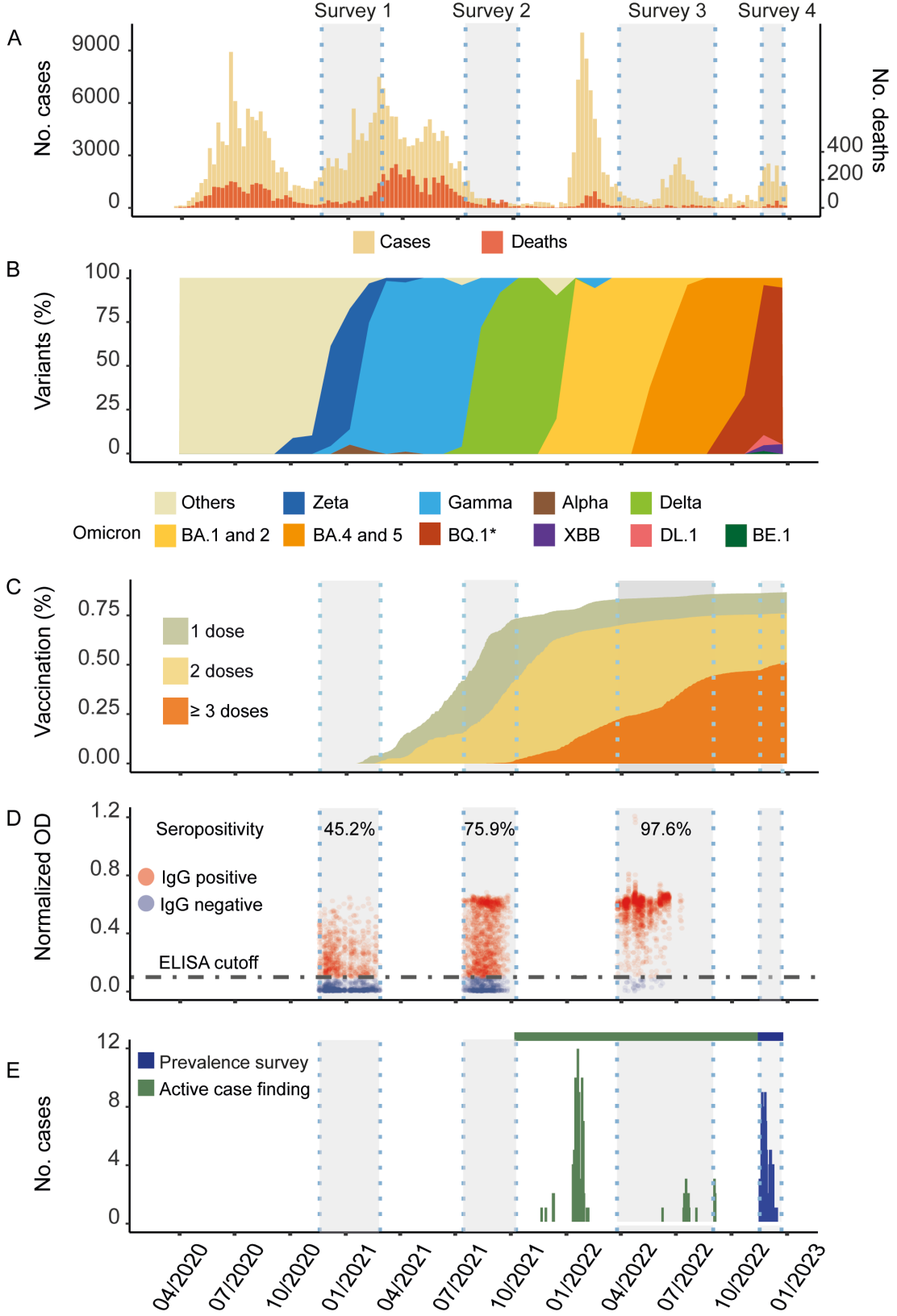
[a] BQ.1 survey was conducted between November 16 and December 22, 2022 and the active case finding was conducted between November 20, 2021, to October 26, 202.
[b] Other symptoms, besides at least one mentioned in the list, included eye discomfort, knuckle, abdominal, chest or lower back pain, itching, and bitterness in the mouth.
[c] WHO definition: acute onset of fever and cough, or acute onset of any three or more of the following signs or symptoms: fever, cough, general weakness/fatigue, headache, myalgia, sore throat, coryza, dyspnea, nausea, diarrhea and anorexia.

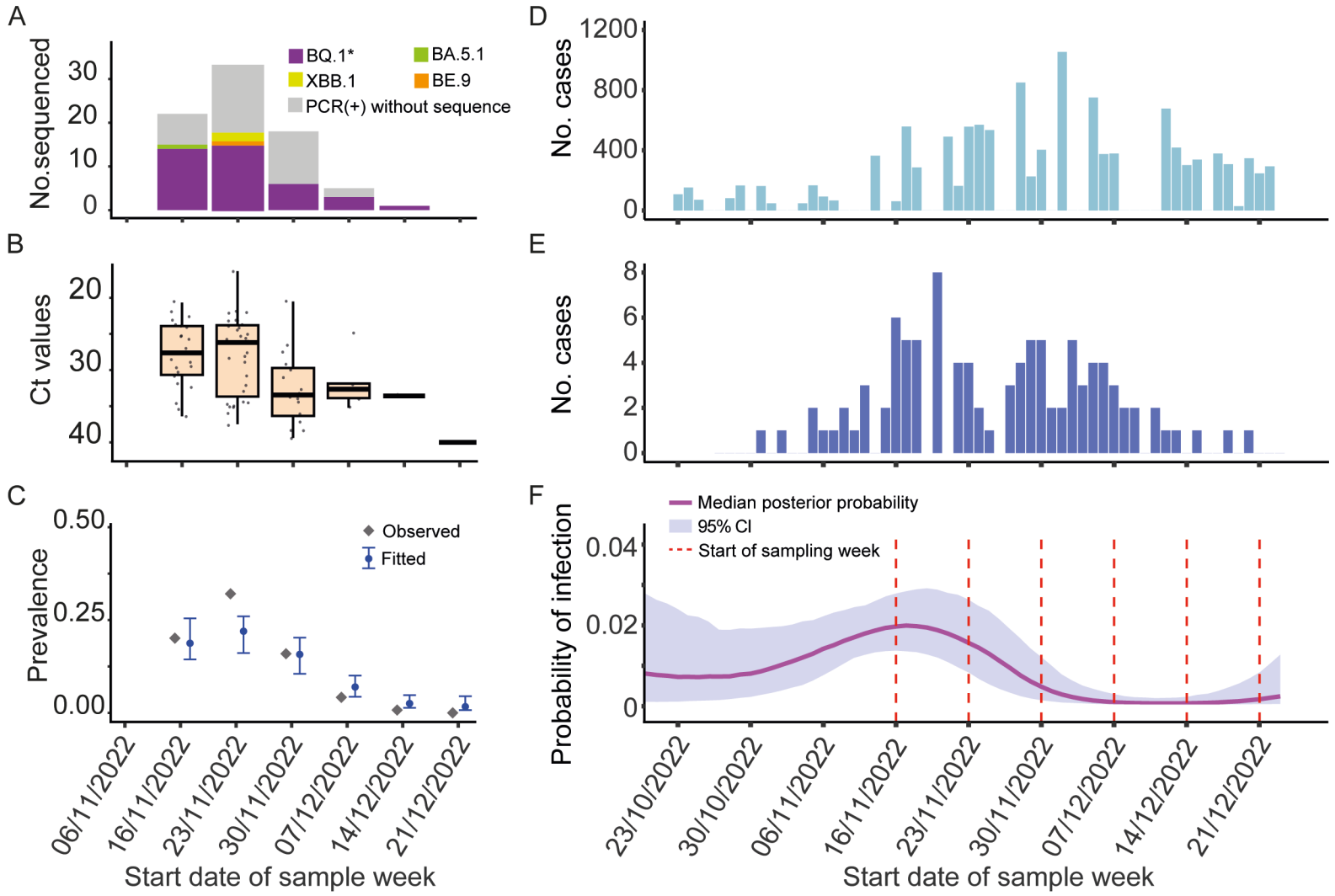
Figure Legend

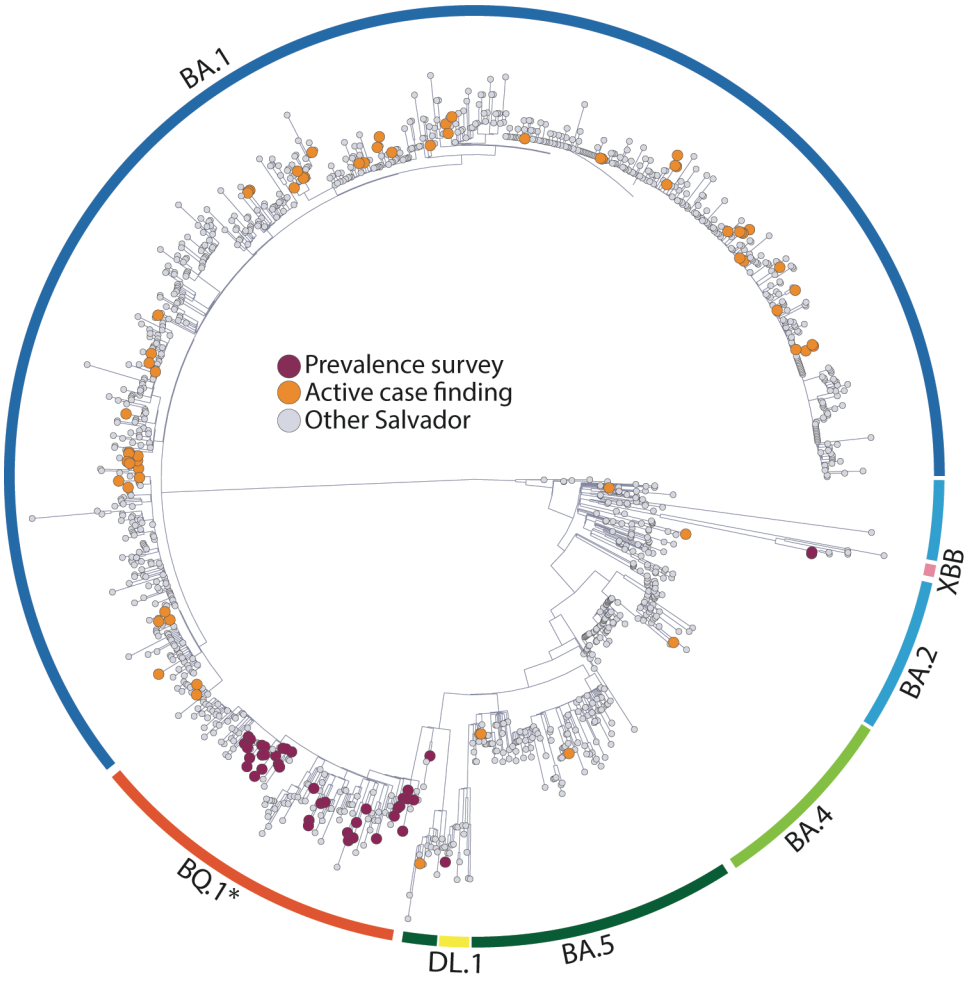
Figure 1. COVID-19 pandemic in Salvador and in the study site. (A) Weekly number of SARS-CoV-2 cases and deaths in Salvador, Brazil. (B) Distribution of SARS-CoV-2 subvariants in Salvador (C) Cumulative proportion of COVID-19 vaccination dose administered amongst Salvador residents. (D) SARS-CoV-2 IgG testing results in previous seroprevalence surveys in recruited individuals. Red dots: SARS-CoV-2 IgG; blue dots: SARS-CoV-2 IgG negative; horizontal grey dot-dash line: OD cut-off value of 0.5). (E) Number of SARS-CoV-2 cases identified in Pau da Lima.

Figure 2. Characterizations of the BQ.1* wave. (A) Number of different subvariant amongst molecular testing positive individuals. (B) Ct value of SARS-CoV-2 cases grouped by week. (C) Weekly observed prevalence (grey diamonds) and fitted median prevalence with 95% CI (blue points and error bars). (D) and SARS-CoV-2 daily cases reported in Salvador. (E) SARS-CoV-2 daily cases reported in Pau da Lima sanitary district (F) Median posterior trajectory for the incidence curve.

Figure 3. Genome-based phylogenetic tree of SARS-CoV-2 Omicron subvariants identified in this study and in the city of Salvador, Brazil.


A
Survey 1
Survey 2
Survey 3
Survey 4
No. cases
9000
6000
3000
0
No. deaths
400
200
0
Cases
Deaths
B
Variants (%)
100
75
50
25
0
Others
Zeta
Gamma
Alpha
Delta
Omicron
BA.1 and 2
BA.4 and 5
BQ.1*
XBB
DL.1
BE.1
C
Vaccination (%)
0.75
0.50
0.25
0.00
1 dose
2 doses
≥ 3 doses
D
Normalized OD
1.2
0.8
0.4
0.0
Seropositivity
45.2%
75.9%
97.6%
IgG positive
IgG negative
ELISA cutoff
E
No. cases
12
8
4
0
Prevalence survey
Active case finding
04/2020
07/2020
10/2020
01/2021
04/2021
07/2021
10/2021
01/2022
04/2022
07/2022
10/2022
01/2023


A
No.sequenced
30
20
10
0
BQ.1*
BA.5.1
XBB.1
BE.9
PCR(+) without sequence
B
Ct values
20
30
40
C
Prevalence
0.50
0.25
0.00
Observed
Fitted
06/11/2022
16/11/2022
23/11/2022
30/11/2022
07/12/2022
14/12/2022
21/12/2022
Start date of sample week
D
No. cases
1200
800
400
0
E
No. cases
8
6
4
2
0
F
Probability of infection
0.04
0.02
0
Median posterior probability
95% CI
Start of sampling week
23/10/2022
30/10/2022
06/11/2022
16/11/2022
23/11/2022
30/11/2022
07/12/2022
14/12/2022
21/12/2022
Start date of sample week


BA.1
Prevalence survey
Active case finding
Other Salvador
XBB
BA.2
BA.4
BA.5
DL.1
BQ.1*

Extensive transmission of SARS-CoV-2 BQ.1* variant in a population with high levels of hybrid immunity: A prevalence survey

Supplementary Material 1

Prior SARS-CoV-2 Exposure Documented in Study Population

As mentioned in the main text, our study population had a high prior exposure to SARS-CoV-2, which was identified through serosurveys and active case-finding before the present study. Three serosurveys were conducted, showing an increase in seropositivity (tested by SARS-CoV-2 anti-S IgG) among participants (Figure 1D). The interval periods for the serosurveys were November 2020 to February 2021 (Survey 1), from June to October 2021 (Survey 2), and from March to August 2022 (Survey 3). Additionally, our team conducted active case-finding between November 2021 to October 2022 to identify symptomatic SARS-CoV-2 cases and their contacts in the same area (Figure 1E). Field teams visited study households every two weeks to screen residents for symptoms and collect nasal swabs for SARS-CoV-2 molecular diagnostics.

To determine prior exposure for each recruited resident, we considered seroconversion occurring before the first dose of vaccination as evidence of prior exposure. We also used PCR-confirmed infection during active case-finding to identify additional prior exposures. Based on a summary of all available evidence, we assigned the final prior SARS-CoV-2 exposure status for each individual into four classes:

1. "Yes": Individuals with prior exposure identified in Survey 1, 2, or 3, or during active case-finding, regardless of follow-up status in other study periods.

2. "No": Individuals with complete follow-up, where prior exposure was not found in Surveys 1, 2, and 3, and during active case-finding.

3. "Unknown": Individuals with complete follow-up, where seroconversion was observed during Surveys 1 to 3 but after the first dose of vaccination (thus, the seroconversion cannot be attributed to prior exposure or vaccination). Additionally, prior exposure was not found during active case-finding.

4. "Missing": Individuals with incomplete follow-up in Surveys 1 to 3 or during active case-finding, with no evidence of prior exposure identified.

Furthermore, we used SARS-CoV-2 IgG levels as a proxy for the immune response. Subsequently, we employed logistic regression models to investigate the association between documented prior exposure in surveys, active case-finding before the BQ.1 outbreak, and the subsequent risk of SARS-CoV-2 infection, while adjusting for the number of COVID-19 vaccine doses, week of sample collection, and age due to their potential confounding role.

Molecular diagnosis

- **RNA extraction and RT-qPCR**

Samples were extracted from 200 µL with Quick-DNA/RNA Viral MagBead Kit (Zymo Research, Cat. no. R2141) using KingFisher Flex System (Thermo Fisher Scientific, Cat. no. 5400630).

The detection of SARS-CoV-2 RNA was performed by RT-qPCR using BIOMOL-OneStep/COVID-19 Kit (Instituto de Biologia Molecular do Paraná, ANVISA no. 80780040004), Molecular SARS-CoV-2 Kit EDx (Bio-Manguinhos, ANVISA no. 80142170045) or CDC 2019-nCoV Reverse Transcriptase PCR Assay [1] on a 7500 Real Time PCR System (Applied Biosystems, Cat. no. 4351105) or QuantStudio 5 Real-Time PCR System (Applied Biosystems, Cat. no. A28574). All protocols followed manufacturer's instructions.

- **NGS library preparation, sequencing and genome assembly**

The libraries were prepared using the COVIDSeq Test (Illumina, catalog numbers 20043675 and 20043137) with the ARTIC V4 or V4.1 primer set as they became available. Equimolar amounts of all libraries were then pooled together. The fragment length distribution was evaluated using the Agilent Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, catalog number 5067-4626) on the Agilent 2100 Bioanalyzer (Agilent Technologies, catalog number G2939BA). The concentration was determined using the Qubit 1X dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, catalog numbers Q33230 or Q33231) on the Qubit 3 Fluorometer (Thermo Fisher Scientific, catalog number Q33216). The library pool was denatured and diluted to a final loading concentration of 8 pM, and then loaded into either the 300-cycle MiSeq Reagent Kit v2 (Illumina, catalog number MS-102-2002) or the 600-cycle MiSeq Reagent Kit v3 (Illumina, catalog number MS-102-3003). The paired-end sequencing was performed using the Illumina MiSeq (Illumina, catalog number SY-410-1003) with a read length of 150 bp. All protocols were carried out following the manufacturer's instructions.

The fastq files generated were submitted to the pipeline defined by Dezordi and colleagues [2] with minor modifications. In brief, the reads were trimmed to remove low-quality base pairs and primers using fastp v0.23.2, [3] the assembly was performed by Burrows-Wheeler Aligner (BWA) v0.7.17 [4] using NCBI GenBank accession no. MN908947.3 as genome reference. The consensus sequences were then masked with "N" at regions with coverage depth <10, and the variant candidates were incorporated into the consensus genome by using iVAR v1.3.1. [5] The assembly statistics were calculated with SAMtools v.1.16.1 (using HTSlib 1.16.1) [6] and Seqtk v1.3-r106 (https://github.com/lh3/seqtk). The sequences generated in this study is available via the GISAID Epi Set identifier EPI_SET_230212yo (doi: 10.55876/gis8.230212yo).

- **Phylogenetic Analysis**

We obtained data for the SARS-CoV-2 Omicron variant from Salvador, Northeast Brazil, Bahia, from the GISAID database [7] for the period between January 1, 2022, and December 31, 2022, using R version 4.2.2 [8] and the GISAIDR package [9]. To ensure the quality of the analyzed data, only genomes greater than 29,000 base pairs with a variant assignment provided by the Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) [10] were considered (n = 1,240). The complete set of sequences used in the analysis can be accessed via the GISAID Epi Set identifier EPI_SET_230211te.

A multiple sequence alignment was performed using MAFFT version 7.511 with the options --6merpair and –add fragments [11, 12]. Problematic sites were masked with "N" in the alignment [13] and manually reviewed using AliView version 1.28 [14].

The maximum likelihood (ML) phylogenetic analysis was performed using IQ-TREE version 2.2.0.3 [15] under the generalized time-reversible (GTR) model of nucleotide substitution, incorporating empirical base frequencies (+F), a proportion of invariant sites (+I), and gamma rate heterogeneity across sites with 4 categories (+G4). The model was chosen based on the Bayesian Information Criterion by ModelFinder [16] and included 1,000 replicates of ultrafast bootstrapping (--B 1000) and the SH-aLRT branch test (--alrt 1000) [17]. The ML consensus tree was visualized using R version 4.2.2 with the ggtree package [18-20] and Adobe Illustrator CC 2023 (http://www.adobe.com).

Assessment of the robustness of the sample

We conducted a sensitivity analysis to assess whether the selected sample might not be representative of the community. Firstly, we used data obtained from the census conducted in the previous survey (Mar 2022 to Aug 2022), during active finding, and in the current study to evaluate the main characteristics associated with participants recruited and not recruited in this study. We observed that the number of residents was associated with participation (<0.001), as identified by the reviewer, as well as participation in the previous survey (<0.001). However, gender, age, previous vaccination, and seropositivity in the previous survey were similar in both populations (Supplementary Table 4).

Then, we evaluated whether the number of residents predicted a positive result in the PCR test for SARS-CoV-2. To do this, we used a logistic regression model and found no association between the number of residents and testing positive for SARS-CoV-2 in the PCR test (Supplementary Table 5). Subsequently, we performed multiple imputation to address missing values in the PCR results using the binary logistic regression method, considering gender, age, participation in the previous survey, and the number of residents in the houses, creating five complete datasets. Then, we used bootstrap sampling to select 1000 samples from each of the 5 datasets, each with 535 participants, as in our study to evaluate the robustness of the sample. With that we calculated the proportions of positive PCR cases in these random samples and the final results show that the obtained proportion of 14.8% positive PCR results used to estimate the incidence in our study , it is included within the expected values for a representative sample of the population, with an expected proportion of 13.9% (95%CI 11.0% - 16.9%) (Supplementary Table 6).

In conclusion, although the number of residents is higher in the selected sample, it is not associated with the results of the PCR test. The sample proved to be robust in determining positivity during the study period in this community.

Estimation of cumulative incidence

To inform the estimate of daily risk of infection using Ct values and PCR positivity prevalence, the model of Hay et al. [21] assumes a distribution of Ct over time following infection, and a corresponding probability of being detectable. To account for differential dynamics of viral load following Omicron infection compared to earlier variants, we updated the parameters used in the original paper using the results of longitudinal studies conducted among Omicron patients. Specifically Hay et al. [21] found that peak Ct of between 25.0 and 26.2 among individuals infected with Omicron, proliferation time (i.e. time to peak viral load) ranging between 3.3 and 4.3 days, and clearance time (i.e. time to maximum Ct following the peak) rangin gfrom 5.8 to 8.7 days. Therefore, we set

the peak Ct at 25, time to peak at 3.8 days, and time to clearance at 7 days. To allow for quicker loss of detection following clearance, we set the probability of loss of detectability to 0.2 per day. With these parameters, the model produced close to 0% PCR positivity at 28 days following infection, consistent with Hay et al. [21] and Boucau et al. [22], and a median duration of PCR positivity of 15 days, consistant with 14.3 days from Kojima et al [23]. Finally, to allow the epidemic curve to have a sharper peak, we relaxed the autocorrelation used in the Gaussian Process model describing incidence over time, using ρ = 0.06.

Assessment of the robustness of the incidence estimation.

In brief, the method of Hay et al assumes a distribution of CT values and probability of PCR positivity over time since infection, and uses individual-level PCR results to infer the likely time of infection among PCR positive individuals, and the proportion of individuals ever infected based on the prevalence of PCR positivity. The likelihood of observing a given CT *y* value in a PCR positive individual sampled on day *t* is the probability that they were infected some days *d* previously, multiplied by the probability of having a detectable CT value and of having CT value *y d* days later, summed over all possible values of *d*. If π_{t-d} is the probability of being infected *d* days before *t*, φ_d is the probability of detectable CT *a* days after infection, and $p_d(y_i)$ is the probability of having CT values y_i *d* days after infection for an individual with detectable CT, then the likelihood of observing a given CT in a PCR-positive individual is

$$\Pr\left(Y_i = y_i \middle| \pi_{t-D_{max}}, \dots, \pi_{t-1}\right) = \sum_{d=1}^{D_{max}} p_d(y_i)\phi_d\pi_{t-d}$$

where D_{max} is the maximum duration of PCR positivity. Similarly, the probability of an individual not having detectable CT (i.e. being PCR negative or having $Y=C_{LOD}$, the limit of detection of CT) is simply one minus the probability of an individual having detectable CT on the day of sampling, i.e.

$$\Pr\left(Y_i = C_{LOD} \middle| \pi_{t-D_{max}}, \dots, \pi_{t-1}\right) = 1 - \sum_{d=1}^{D_{max}} \phi_d\pi_{t-d}$$

The likelihood for the whole sample can be constructed by multiplying together the individual likelihood contributions. The daily probability of infection is modeled using a Gaussian process, such that daily infection probabilities are correlated (see Hay et al Supplementary Material for details).

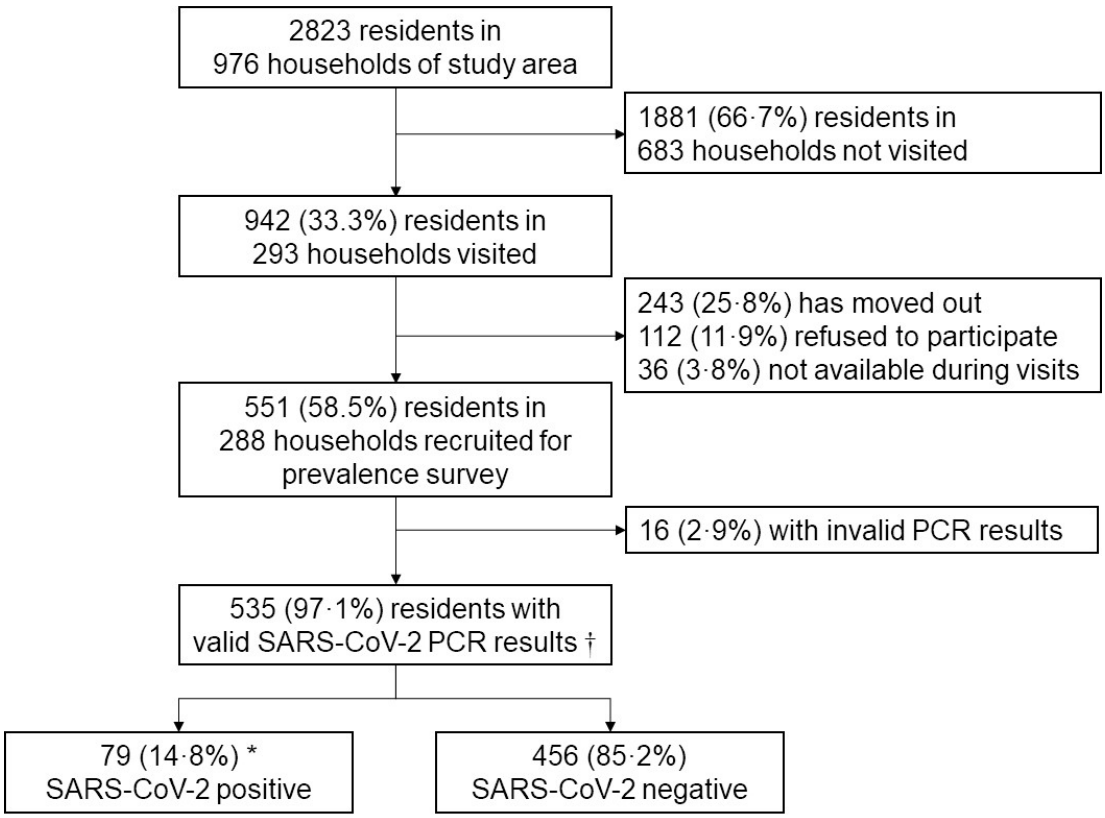
We performed sensitivity analyses to check the robustness of the cumulative incidence estimate to changes in the CT distribution and PCR positivity probability over time. Specifically, we assumed: longer time to loss of detectability following clearance (consistent with Saade et al. [24] and Luna-Muschi et al. [25]); faster time from peak viral load to viral clearance (4 days vs 7 days); higher peak viral load (peak CT 20 vs. 25); and higher viral load following viral clearance (CT following clearance 35 vs 38) (Supplementary Table 7).

Analysis code for the incidence estimation are available on Github (https://github.com/mhitchings/BQ1_PCR_CumulativeIncidence).

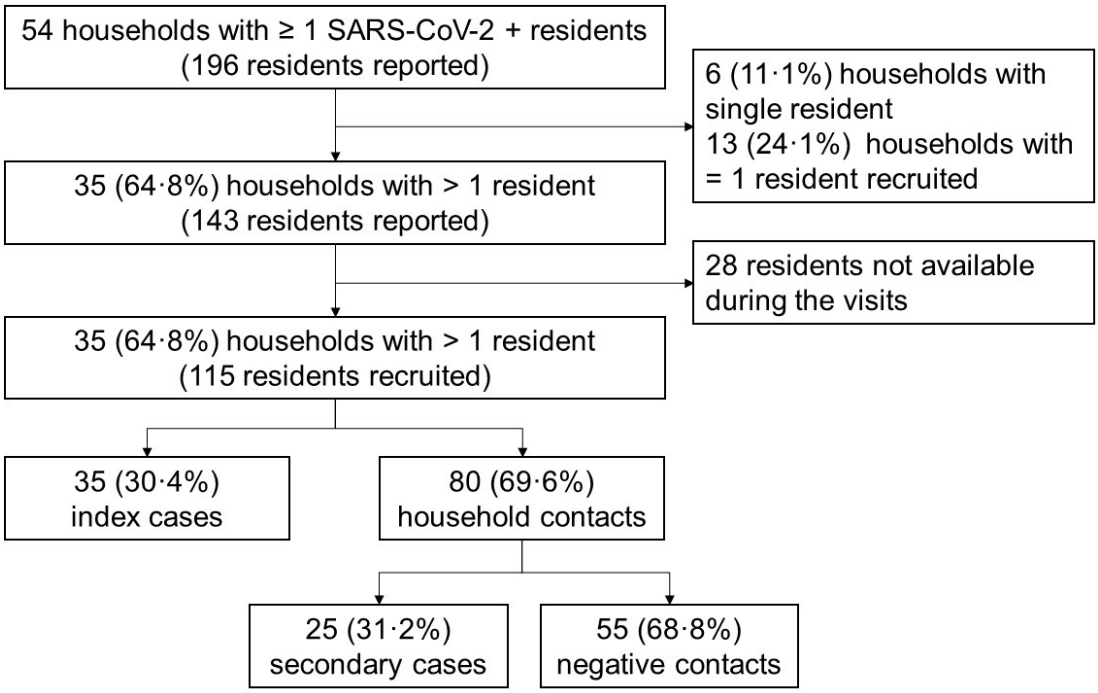
References

1. Centers for Disease Control and Prevention. CDC 2019-novel coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel. 2020.
2. Dezordi FZ, Neto A, Campos TL, et al. ViralFlow: A Versatile Automated Workflow for SARS-CoV-2 Genome Assembly, Lineage Assignment, Mutations and Intrahost

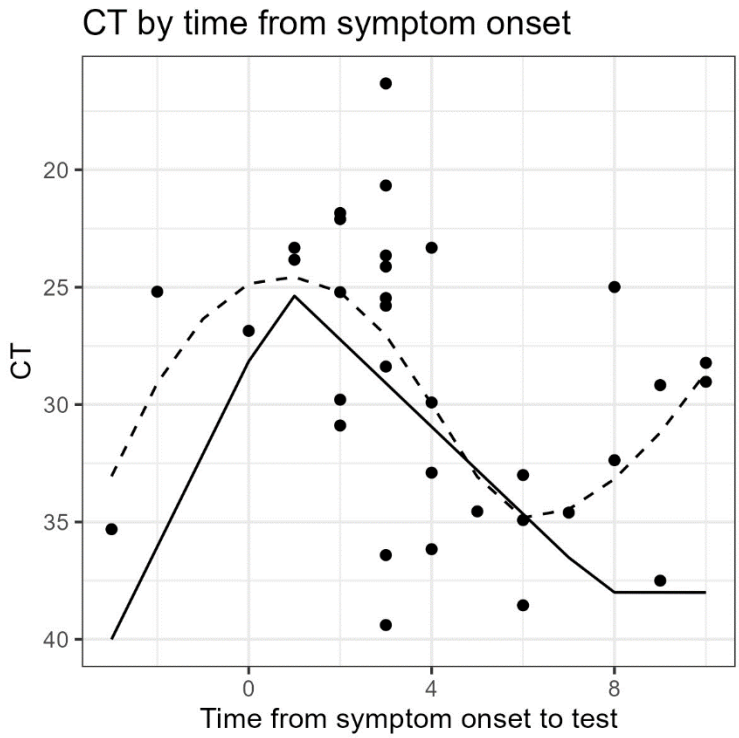
Variant Detection. Viruses **2022**; 14.
3. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **2018**; 34:i884-i90.
4. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **2010**; 26:589-95.
5. Castellano S, Cestari F, Faglioni G, et al. iVar, an Interpretation-Oriented Tool to Manage the Update and Revision of Variant Annotation and Classification. Genes (Basel) **2021**; 12.
6. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. Gigascience **2021**; 10.
7. Khare S, Gurry C, Freitas L, et al. GISAID's Role in Pandemic Response. China CDC Wkly **2021**; 3:1049-51.
8. R Core Team R. R: A language and environment for statistical computing. **2013**.
9. Wirth W, Duchene S. GISAIDR. Zenodo; 2022.
10. O'Toole Á, Scher E, Underwood A, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol **2021**; 7:veab064.
11. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **2002**; 30:3059-66.
12. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **2013**; 30:772-80.
13. Turakhia Y, De Maio N, Thornlow B, et al. Stability of SARS-CoV-2 phylogenies. PLoS Genet **2020**; 16:e1009175.
14. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics **2014**; 30:3276-8.
15. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol **2020**; 37:1530-4.
16. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods **2017**; 14:587-9.
17. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol **2018**; 35:518-22.
18. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution **2017**; 8:28-36.
19. Yu G, Lam TT, Zhu H, Guan Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. Mol Biol Evol **2018**; 35:3041-3.
20. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics **2020**; 69:e96.
21. Hay JA, Kissler SM, Fauver JR, et al. Quantifying the impact of immune history and variant on SARS-CoV-2 viral kinetics and infection rebound: A retrospective cohort study. eLife **2022**; 11:e81849.
22. Boucau J, Marino C, Regan J, et al. Duration of Shedding of Culturable Virus in SARS-CoV-2 Omicron (BA.1) Infection. New England Journal of Medicine **2022**; 387:275-7.
23. Kojima N, Roshani A, Klausner JD. Duration of COVID-19 PCR positivity for Omicron vs earlier variants. J Clin Virol Plus **2022**; 2:100085.
24. Saade C, Brengel-Pesce K, Gaymard A, et al. Dynamics of viral shedding during ancestral or Omicron BA.1 SARS-CoV-2 infection and enhancement of pre-existing immunity during breakthrough infections. Emerg Microbes Infect **2022**; 11:2423-32.
25. Luna-Muschi A, Noguera SV, Borges IC, et al. Characterization of Severe Acute Respiratory Syndrome Coronavirus 2 Omicron Variant Shedding and Predictors of Viral Culture Positivity on Vaccinated Healthcare Workers With Mild Coronavirus Disease 2019. The Journal of Infectious Diseases **2022**; 226:1726-30.



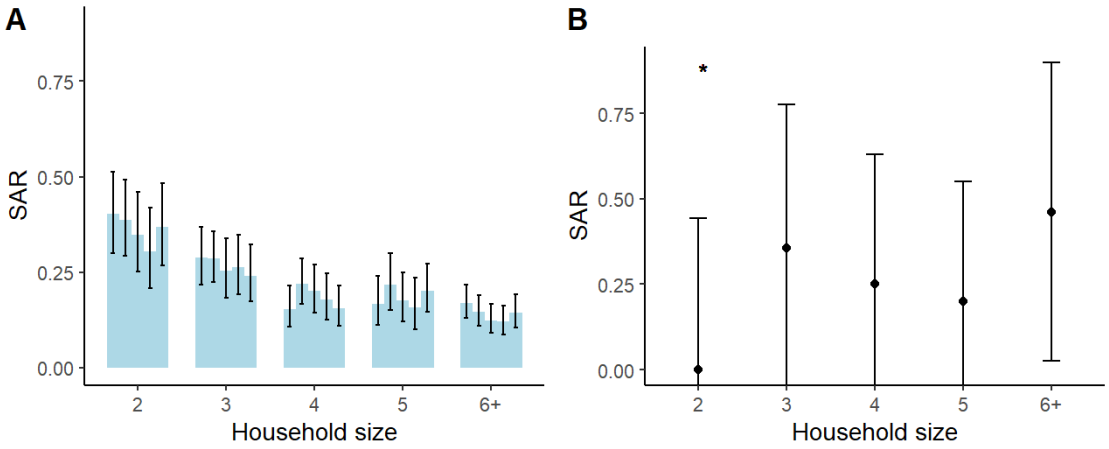
Supplementary Figure 1. Study flowchart. * Four only defined by rapid antigen test (RAT), among them two were RT-PCR negative, one RT-PCR intermediate and one RT-PCR not completed. † Invalid PCR results indicated samples failed to amplify internal controls.



Supplementary Figure 2. Flowchart of households selected for estimating secondary attack rate. The index case was defined as the individual with the earliest positive test or onset date for symptoms in the household. If two members had the same earliest date for positive test and symptoms, they were considered co-index cases. If infection order could not be determined for any household member, that household was excluded from the SAR analyses (co-index cases).



Supplementary Figure 3. Ct value over time from symptom onset for individuals with positive PCR test and recorded symptom onset date, with a LOESS smoother fit to the data (dotted line) and the assumed Ct distribution over time assumed by the model (solid line), assuming three days from infection to symptom onset



Supplementary Figure 4. Secondary Attack Rate distributed by number of residents in the household A) data obtained by multiple imputation and B) sample study
* 95%CI for a binomial proportion using the Agresti-Coull method

Supplementary Table 1. Household secondary attack rate (SAR) and risk factors.

Secondary cases characteristic	No. secondary cases /No. contacts	SAR (%) (95% CI)	RR (95%CI)
All	25/80	31.3 (22.2-42.1)	
Age group			
<18	15/34	44.1 (28.9-60.6)	2.03 (1.04-3.95)
≥18	10/46	21.7 (12.3-35.6)	1 [Reference]
18 - 35	6/24	25.0 (12.0-44.9)	
36 - 59	2/18	11.1 (3.1-32.8)	
≥60	2/4	50.0 (15.0-85.0)	
Sex			
Female	15/45	33.3 (21.4-47.9)	1.16 (0.60-2.27)
Male	10/35	28.6 (16.3-45.1)	1 [Reference]
Vaccine dose(s) received			
≥3 doses	10/26	38.5 (22.4-57.5)	0.87 (0.36-2.08)
2 doses	4/27	14.8 (5.9-32.5)	0.33 (0.10-1.07)
1 dose	7/18	38.9 (20.3-61.4)	0.63 (0.24-1.65)
0 dose	4/9	44.4 (18.9-73.3)	1 [Reference]

Supplementary Table 2. Prior SARS-CoV-2 exposure during the major cohort study.

Prior exposure	No. (%)		p-value
	SARS-CoV-2 positive n = 79	SARS-CoV-2 negative n = 456	
Seropositive in Survey 1[a]			0.157
Yes	17 (10.4%)	146 (89.6%)	
No	36 (17.4%)	171 (82.6%)	
Incomplete follow-up[b]	26 (15.8%)	139 (84.2%)	
Seroconversion in Survey 2 and not previously vaccinated[c]			0.047
Yes	5 (8.1%)	57 (91.9%)	
No	17 (17.7%)	79 (82.3%)	
Unknown	31 (12.3%)	222 (87.7%)	
Incomplete follow-up[b]	26 (21.0%)	98 (79.0%)	
Seroconversion in Survey 3 and not previously vaccinated[c]			0.083
Yes	4 (11.1%)	32 (88.9%)	
No	0 (0.0%)	6 (100.0%)	
Unknown	55 (13.4%)	354 (86.6%)	
Incomplete follow-up[b]	20 (23.8%)	64 (76.2%)	
PCR-confirmed infection during active case finding			0.957
Yes	7 (15.9%)	37 (84.1%)	
No	70 (14.6%)	409 (85.4%)	
Incomplete follow-up[b]	2 (16.7%)	10 (83.3%)	
Prior documented SARS-CoV-2 exposure[c]			0.018
Yes	31 (10.7%)	258 (89.3%)	
No	0 (0.0%)	6 (100.0%)	
Unknown	27 (21.3%)	100 (78.7%)	
Incomplete follow-up	21 (18.6%)	92 (81.4%)	
Prior documented SARS-CoV-2 exposure or vaccination[c]			>0.999
Yes	77 (14.9%)	441 (85.1%)	
No	2 (11.8%)	15 (88.2%)	

[a] As SARS-CoV-2 vaccines were not available to public until the end of Survey 1, any seropositive observed in Survey 1 was considered due to SARS-CoV-2 infection.
[b] Incomplete follow-up in Surveys 1-3 indicates individuals were not recruited in corresponding surveys. Incomplete follow-up in active case finding indicates individuals were not visited at least once during the complete study period.
[c] See definitions of prior SARS-CoV-2 exposure classifications in Supplementary Material 1.

Supplementary Table 3. Factors associated with SARS-CoV-2 positive result.

	n	Positive	Negative	Odds Ratio	95%CI	p
Model 1						
Seropositive in Survey 1, n (%)	350	17 (21.5%)	146 (32.0%)	0.50	0.25 – 0.97	0.045
Model 2						
SARS-CoV-2 IgG OD value in Survey 3, mean (SD)	416	2.79 (0.622)	2.76 (0.760)	1.15	0.71 – 1.91	0.571
Model 3						
Seropositive in Survey 1, n (%)	308	16 (36.4%)	131 (49.6%)	0.52	0.24 – 1.07	0.080
SARS-CoV-2 IgG OD value in Survey 3, mean (SD)	308	2.89 (0.503)	2.82 (0.714)	1.36	0.75 – 2.53	0.317
Model 4						
Seropositive in Survey 1, n (%)	350	17 (33.3%)	146 (48.8%)	0.50	0.25 – 0.98	0.046
PCR-confirmed infection in active case finding, n (%)	350	5 (9.8%)	27 (9.0%)	1.06	0.32 – 3.03	0.915

All models were adjusted by age, the number of vaccine doses and the week of sample collection.

Supplementary Table 4. Characteristics of recruited and non-recruited Pau da Lima residents

Characteristic	Responses	Recruited n = 942	Non-recruited n = 1881	p-value
Available information, n (%)*	2,823			**<0.001**
Yes		920 (97.7)	1740 (92.9)	
No		22 (2.3)	141 (8.1)	
Age in years	2,658	32 (18)	32 (19)	0.74
Sex, n (%)	2,660			0.31
Female		494 (54)	976 (56)	
Male		423 (46)	767 (44)	
Number of residents	2,660	3.94 (1.99)	3.39 (1.87)	**<0.001**
Vaccination, n (%)	2,105			0.60
Yes		647 (85)	1,160 (86)	
No		112 (15)	186 (14)	
Participate in previous survey L47, n (%)	2,655			**<0.001**
Yes		685 (75)	908 (52)	
No		229 (25)	833 (48)	
IgG SARS-CoV-2 result in the previous survey, n (%)	1,760			0.25
Positive		605 (98)	1,130 (99)	
Negative		12 (1.9)	13 (1.1)	
BQ1 survey PCR result, n (%)	2,660			**<0.001**
Positive		79 (8.6)	0 (0)	
Negative		456 (49)	0 (0)	
Invalid PCR		16 (1.7)	0 (0)	
Non recruited		369 (40)	1,740 (100)	

* Data obtained through the population census, previous surveys, active case finding, and during the completed survey in the cohort study

Supplementary Table 5. Logistic regression to evaluate the effect of the number of residents in the PCR SARS-CoV-2 positive result.

Characteristics	Odds Ratios	95%CI	p
Intercept	0.19	0.06 – 0.60	0.006
Number of residents	1.09	0.96 – 1.22	0.184
Age in years	1.00	0.99 – 1.01	0.957
Male	0.82	0.49 – 1.35	0.445
Vaccinated	1.33	0.66 – 2.89	0.450
Participate in previous survey L47	0.49	0.27 – 0.92	0.021
Observations		535	

Supplementary Table 6. Sensitivity analysis employing multiple imputation techniques to assess the sample's robustness in evaluating the proportion of PCR SARS-CoV-2 positive result among a sample of 535 participants

	Proportion of PCR SARS-CoV-2 +	95%CI	
dataset 1	13.5%	10.5%	16.4%
dataset 2	16.7%	13.6%	19.8%
dataset 3	12.7%	9.7%	15.5%
dataset 4	13.2%	10.3%	16.1%
dataset 5	13.7%	10.7%	16.5%
Pooled proportion	13.9%	11.0%	16.9%

Supplementary Table 7. Estimated cumulative incidence from October 19 to December 22, 2022, adjusting parameters relating to PCR CT distribution and probability of PCR positivity over time following infection

Model	Estimated CI (95% CrI)	Estimated day of peak incidence (95% CrI)
Base model	56% (36%, 88%)	Nov 17 (Nov 9, Nov 21)
Longer time to loss of detectability (daily probability of loss of detectability following clearance = 0.1)	**49%** (33%, 76%)	Nov 16 (Nov 9, Nov 21)
Faster time to clearance (time from peak to maximum CT = 4 days)	**62%** (40%, 94%)	Nov 17 (Nov 6, Nov 21)
Lower peak CT (peak CT = 20)	54% (34%, 86%)	Nov 17 (Nov 5, Nov 21)
Lower CT after clearance (CT=35)	53% (34%, 85%)	Nov 16 (Nov 4, Nov 20)

In bold, the maximum and minimum values.

Supplementary Table 8. Sensitivity analysis employing multiple imputation techniques to assess the household secondary attack rate

	SAR	95%CI	
dataset 1	20.9%	13.9%	28.7%
dataset 2	22.4%	14.8%	31.3%
dataset 3	19.1%	13.0%	27.0%
dataset 4	17.7%	11.3%	26.1%
dataset 5	19.3%	13.0%	25.2%
Pooled	20.5%	13.2%	27.7%