Open Forum Infectious Diseases

Overestimation of SARS-CoV-2 household transmission in settings of high community transmission: insights from an informal settlement community in Salvador, Brazil

--Manuscript Draft--

Manuscript Number:	OFID-D-23-01010R1
Full Title:	Overestimation of SARS-CoV-2 household transmission in settings of high community transmission: insights from an informal settlement community in Salvador, Brazil
Article Type:	Major Article
Section/Category:	Epidemiology and Disease Surveillance
Keywords:	SARS-CoV-2; Omicron; BA.1; household transmission
Corresponding Author:	Juan Pablo Aguilar Ticona Instituto De Saude Coletiva UFBA Salvador, Bahia BRAZIL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Instituto De Saude Coletiva UFBA
Corresponding Author's Secondary Institution:	
First Author:	Juan Pablo Aguilar Ticona
First Author Secondary Information:	
Order of Authors:	Juan Pablo Aguilar Ticona
	Nivison Nery
	Matt Hitchings
	Emilia M. M. Andrade Belitardo
	Mariam O. Fofana
	Murilo Dorión
	Renato Victoriano
	Jaqueline S. Cruz
	Juliet Oliveira Santana
	Laise Eduarda Paixão de Moraes
	Cristiane W. Cardoso
	Guilherme S. Ribeiro
	Mitermayer G. Reis
	Ricardo Khouri
	Federico Costa
	Albert I. Ko
	Derek A.T. Cummings
Order of Authors Secondary Information:	
Manuscript Region of Origin:	BRAZIL
Abstract:	Background The SARS-CoV-2 Omicron variant has spread globally. However, the contribution of community versus household transmission to the overall risk of infection remains unclear.

	Methods Between November 2021, and March 2022, we conducted an active case-finding study in an urban informal settlement with biweekly visits across 1174 households with 3364 residents. Individuals displaying COVID-19-related symptoms were identified, interviewed along with household contacts, and defined as index and secondary cases based on RT-PCR and symptom onset. Results In 61 households, we detected a total of 94 RT-PCR-positive cases. Out of 69 sequenced samples, 67 cases (97.1%) were attributed to the Omicron BA.1* variant. Among 35 of their households, the secondary attack rate was 50.0%(% (95%CI 37.0–63.0%). Women (p=RR = 1.6; 95%CI = 0.079 – 2.7), older individuals (p=0.03)median difference = 15; 95%CI = 2 – 21), and those reporting symptoms (p=RR = 1.73; 95% CI = 1.0 – 3.0.04) had a significantly increased risk for SARS-CoV-2 secondary infection. Genomic analysis revealed substantial acquisition of viruses from the community even among households with other SARS-CoV-2 infections. After excluding community acquisition, we estimated a household secondary attack rate of 24.2%(% (95%CI 11.9–40.9%). Conclusions These findings underscore the ongoing risk of community acquisition of SARS-CoV-2 among households with current infections. The observed high attack rate necessitates swift booster vaccination, rapid testing availability, and therapeutic options to mitigate COVID-19's severe outcomes.
Suggested Reviewers:	Nicholas Davies nicholas.davies@lshtm.ac.uk He is a researcher focused on outbreak response and the transmission of COVID-19. He helped to produce evidence for the UK government during the pandemic as a member of SPI-M and during a secondment to the UK Cabinet Office COVID-19 Task Force
	Kylie Ainslie kylie.ainslie@rivm.nl ylie Ainslie, an infectious disease modeler at RIVM and honorary associate professor at the University of Hong Kong, employs mathematical models to assess vaccination strategies' impact on disease transmission. She also develops statistical methods to understand vaccine protection decline over time, focusing on respiratory diseases like COVID-19 and influenza at the University of Hong Kong. During the pandemic, she was part of the Imperial COVID-19 Response Team and the REal-time Assessment of Community Transmission (REACT) team, contributing to COVID-19 pandemic modeling and characterization.
	Kathy Leung ksmleung@hku.hk Kathy Leung is a Assistant Professor at HKU School of Public Health. Her research focus on mathematical modeling for a diverse spectrum of diseases, including influenza, MERS, COVID-19, hand-foot-and-mouth disease, HPV, cervical cancer, colorectal cancer, and breast cancer.
	Aine O'Toole aine.otoole@ed.ac.uk Aine O'Tooleis a researcher working in the Rambaut Group at The University of Edinburgh. Prior to the COVID-19 pandemic, Áine developed software for viral surveillance & outbreak investigations. Furthermore during the recent years, Áine has worked in the phylogenetics core of COG-UK, and developed tools such as PANGOLIN and CIVET to help researchers around the world track the spread of SARS-CoV-2.
	Leon Danon l.danon@bristol.ac.uk Leon Danon is an interdisciplinary scientist whose research spans collective behavior, complex systems, networks, and public health. He analyzes human populations and social interactions, motivated by applications in public health. His work relies on large datasets that reveal statistical patterns, explored through statistical physics, mathematical modeling, and data science. His primary focus lies in infectious disease epidemiology and the intricate relationship between human behavior and disease transmission.

Opposed Reviewers:	
Response to Reviewers:	
Additional Information:	
Question	**Response**
Are you or any co-authors a member of IDSA?	No
Manuscript Classifications:	COVID-19; Transmission
Secondary Full Title:	

August 8, 2023

Editorial Board
Open Forum Infectious Diseases

Ref: Submission of Research Article - "Overestimation of SARS-CoV-2 Household Transmission in High Community Transmission Settings: Insights from an Informal Settlement Community in Salvador, Brazil"

Dear Editorial Board Members:

We are writing to see if Open Forum Infectious Diseases would be interested in our manuscript entitled "Overestimation of SARS-CoV-2 Household Transmission in High Community Transmission Settings: Insights from an Informal Settlement Community in Salvador, Brazil" for publication as a major article.

Our manuscript delves into the transmission dynamics of the SARS-CoV-2 Omicron BA.1 variant within an urban informal settlement in Brazil. In this study, we conducted biweekly visits to 1174 households with 3364 residents, to identify COVID-19 symptoms and confirm diagnoses through RT-PCR testing. Our findings revealed a high household secondary attack rate of 50.0%, with increased vulnerability among women, older individuals, and symptomatic cases. Despite this, our genomic analysis indicated a significant influence of community transmission, leading to a refined secondary attack rate of 24.2% when community transmission was excluded. Our findings highlight the ongoing risk of community acquisition that individuals face even as other members of their household are infected, a fact that could potentially shift focus to controlling exposure within the household. We believe the high transmissibility of Omicron variants and temporally clustered outbreaks mix the timescale of household outbreaks with that of the overall community, leading to misclassification of household transmission using traditional methods that highlight the proximity of household cases in time.

We believe that our manuscript will offer valuable insights to the readership of Open Forum Infectious Diseases, who are navigating the complexities surrounding the evolving SARS-CoV-2 variants. Furthermore, our article presents distinctive epidemiological and genomic evidence elucidating the potential overestimation of secondary household transmission of the BA.1* variant. Given that many studies struggle to entirely account for infections originating outside the household, the observed household transmission rates could be overestimated.

Finally, we declare that this manuscript has not been published before, either in whole or in part, and is not presently under consideration for publication elsewhere. We affirm that all authors have made significant contributions to the creation of this manuscript, with each of them fulfilling the established criteria, and all authors have given their approval for its submission. We also declare that we did not have any writing assistance.

Finally, we recommend the following experts for potential review of our manuscript:

- Nicholas Davies
  Institution: London School of Hygiene & Tropical Medicine
  Email: nicholas.davies@lshtm.ac.uk
- Kylie Ainslie
  Institution: Centrum Infectieziektebestrijding

Email: kylie.ainslie@rivm.nl
- Kathy Leung
Institution: School of Public Health, The University of Hong Kong
Email: ksmleung@hku.hk

- Aine O'Toole
Institution: ARTIC Network & Polio Sequencing Consortium
Email: aine.otoole@ed.ac.uk

- Leon Danon
Institution: Department of Engineering Mathematics, Bristol University
Email: l.danon@bristol.ac.uk

Please feel free to contact us if you have any questions and thank you again for your consideration.

Sincerely,

Juan P. Aguilar Ticona, Universidade Federal da Bahia
Albert Ko, Yale University
Federico Costa, Universidade Federal da Bahia
Derek Cummings, University of Florida
For the authors

January 8, 2024

Editorial Board

Open Forum Infectious Diseases (OFID)

Dear Editorial Board Members:

Manuscript OFID-D-23-01010

Title: Overestimation of SARS-CoV-2 household transmission in settings of high community transmission: insights from an informal settlement community in Salvador, Brazil

Responses to Reviewers' Comments

Reviewer 1: In this study the secondary attack rate (SAR) of the SARS-CoV2, Omicron BA.1 variant was estimated in households in a densely populated area in Pau da Lima, Brazil. Household transmission was distinguished from community transmission using genomic analysis, which is a strength in this study, and has been a significant limitation in other studies, aiming to estimate SAR of the Omicron BA.1. However, the number of included households with a complete phylogenetic analysis was only 14, with a total of 47 study persons included. This is a major limitation to the power of the study, not allowing for generalizable conclusions in a broader context.

We are grateful to Reviewer #2 for highlighting the strengths of our study. While acknowledging the main limitation lies in the small number of participants, it's important to note that:

"In a systematic review of 57 studies, the majority of these (43) mainly examined the Household SAR. The authors of this review indicate that disregarding external sources of infection might lead to an overestimation of SAR within households. The absence of comparisons between secondary and community infections when estimating SAR was acknowledged as a limitation. Also, none of the reviewed studies utilized techniques like WGS to confirm genetic similarity between the strains infecting index and subsequent cases within households [27]. In contrast, our study stands out for its use of phylogenetic analysis, crucial in understanding the community and household transmissions (adjusted SAR = 24.2%) in Pau da Lima, Brazil. Analyzing genetic sequences from individuals in Pau da Lima and Salvador revealed a resemblance between the samples, suggesting multiple virus introductions into this community, making it representative of Salvador city. Despite the absence of clusters in our phylogenetic analysis, our site is representative of the transmission dynamics in Salvador, where 42% of households belong to an urban informal community. Despite limitations in our sequencing scope, we successfully identified transmission clusters within households and the community, highlighting localized virus spread. While acknowledging the need for larger-scale studies to confirm and expand our findings, previous studies utilizing WGS for transmission assessment showed similar outcomes [24, 34]." (Lines 330 – 345)

Reviewer 2: The authors present a population-based study of household transmission of the Omicron variant of SARS-CoV-2 conducted in Brazil during 2021-2022. The methods are clearly described, and the molecular analysis is very helpful in understanding patterns of community and household transmission. These data provide important insights into SARS-CoV-2 transmission in under-resourced settings which have not been well-described in the literature. The epidemiologic methods (particularly with regard to case ascertainment) could be more fully described. I have a few comments for the authors' consideration.

We are grateful to Reviewer #2 for their comments. We have prepared responses below and modified the manuscript based on their feedback.

Abstract

1. Lines 68-69: In the results section, consider presenting the risk ratios with 95% CIs rather than p-values alone.

We adjusted the text to include the RR, median difference and their 95% CIs. (Now lines 70 – 72)

Introduction

1. Lines 87-89: Please clarify what is meant here - as I'm reading it, this suggests that home isolation could increase risk of community transmission.

We have revised and highlighted the significance of this situation in urban informal settings. While home isolation functions as a preventive measure against COVID-19, its effectiveness might be compromised in communities marked by overcrowded households, inadequate infrastructure, and poor ventilation. In such contexts, relying solely on home isolation may not yield the same benefits during the pandemic, highlighting the need for public health interventions aimed at improving living conditions and overall quality of life. (Now lines 90 – 94)

Methods

1. What is the overall vaccine coverage for Pau da Lima? Have there been any mass vaccination campaigns? If so, what vaccine(s) were available (mRNA-based? bivalent?)?

Extensive vaccination campaigns were carried out across the city of Salvador. Previous studies have documented the vaccination intentions and coverage within this population (Now reference 13 and 15, lines 113 – 114). We've incorporated a table outlining the vaccine coverage and types administered up until the study period. Please refer to Supplementary Table 2 for details.

2. Lines 113-114: This is really a tremendous effort by the field teams. Could you provide more details regarding sampling strategy for the households? In particular, since households were visited every two weeks, was this a random subset of household throughout Pau da Lima, or was each 2-week sampling period performed in different geographic areas? If the sampling frame was in different geographic areas every two weeks, this could introduce bias if cases were very focally clustered.

The intention was to visit all households in the Pau da Lima community. However, it was not feasible to cover every household; instead, we managed to visit 56% to 85% of them. We have included the Supplementary table 1 with this. Alsos we add a description “Participants were included based on their availability, and multiple

attempts, including weekends, were made to limit missing data across the three valleys comprising the study area" in the methods section (now line 131 – 133) and improve the limitation base on the missing results (now lines 373 – 375).

Results

1. Lines 195-196: Consider providing median number of households visited each biweekly period.

We've added a supplementary table detailing the households visited and the number of participants for each week (Supplementary table 1).

2. What were the other diagnoses beyond SARS-CoV-2?

This study focused on identifying COVID-19 cases. Additionally, we identified cases of flu, RSV, and other respiratory viruses. The analysis of this data is currently being finalized, and we plan to include it in a forthcoming publication.

3. Lines 204-205: There is mention of a two Delta variant infections during Dec 2021 - these do not appear as a PCR+ index case in Figure 2, panel D.

We omitted the delta variants from our analysis of household secondary attack rates, which is why the delta variant was not initially considered in the figure. However, for clarity, we have updated Figure 2 to include the delta cases and previous non-Omicron PCR+ cases from 2021.

4. Lines 215-216: Is this proportion who received a vaccine higher or lower than for Pau da Lima overall?

The vaccination coverage was similar to that of the main cohort. We've included a table (Supplementary Table 2) displaying vaccine coverage until March 21st, demonstrating a comparable proportion of vaccinations (Lines 223 – 224).

5. Lines 217-222: Consider presenting these results as risk ratios.

We adjusted the text to include the RR, median difference and their 95% Cis (Lines 228 – 230).

6. Lines 231-234: I'm not sure what is meant here - my interpretation is that intrahousehold sequences were both similar to other sequences from the same household and from the surrounding community.

We've included a brief interpretation to aid readers: "Briefly, this means that there is a notable similarity in sequences among households with two or more infected participants when compared to households with a single infected participant or Pau da Lima or samples from Salvador. This similarity reinforces the likelihood of household transmission" (Lines 245 – 248).

Discussion

1. Lines 255-256: This is an excellent point and it would be nice to see more elaboration of this - particularly as many household transmission studies are in the US/Europe context, where household sizes may be smaller, and the square footage of residences is certainly larger.

We improved the text in the discussion following the reviewer's recommendation (Lines 275 – 284).

2. Lines 276-279: Are these practical recommendations for Pau da Lima? Can residents isolate at home without worries regarding job/income loss? Do the structure of residences allow for improved ventilation (e.g. power available for Cori-Rosenthal boxes, can additional windows/vents be installed)?

We improved the text in the discussion following the reviewer's recommendation. (Lines 308 -320).

Figures and Tables

1. Consider including a new table showing total number of households visited compared to those with at least one positive case. This would allow for comparison of households with at least one case to those with no cases. Variables that would be interesting to examine here include household size, age of residents, number of children, sex of residents (basically the same as presented in current Table 1).

We added a supplementary table to comprehend the distinctions among households where at least one individual tested PCR+ and those that did not. (now Supplementary Table 4 and lines 258 - 263)

Figure 1: Very helpful to understand the geography of Pau da Lima. Is this settlement divided into districts? If so this would be helpful to show.

This urban informal settlement lacks formal divisions. However, the major cohort study divided the area into blocks to facilitate standardized visits. We added the map used during the active case finding but they could change across the study and objectives in the main cohort (now supplementary Figure 1).

Figure 2: Was PCR performed prior to 27 Dec 2021? i would assume yes and these results were negative based on panel A.

We incorporated positive cases occurring before December 27th, 2021, into Figure panel D. In the WGS analysis, only two Delta cases were identified, and non-Omicron cases were also detected during this timeframe. (Figure 2 panel D).

Figure 3: Panel A is somewhat difficult to read. Would it be possible to create another tree with "House with a single..." removed? This may help improve legibility and identify clustering. Additionally, many of the colors are similar (e.g. House 10 and House 11, House 12/13/14), and the use of red/green tones may be challenging for readers with colorblindness.

We made adjustments to the figure in response to the reviewer's comments. To prevent misinterpretation between sequences from Salvador and House with a single, sequence we represented the latter as small empty circles (Figure 3 panel A).

Figure 4: Is it possible to add the house numbers to the map in Figure 1? Given the layout of the settlement, understanding geographic links between households would be very helpful. Additionally, adding some kind of symbol for each household would help the reader to rapidly identify households that exist within multiple clusters (e.g. House 14, House 5).

We add a new panel in the figure 1 following the reviewer´s suggestion (Figure 1 panel C).

Table 2: Could this be presented as risk ratios? I think that RR is an easier way to grasp what is presented here.

We modified the table to include the RR (Table 2).

Consider moving supplementary Figure 1 to the main text. If there are figure limitations, I think that current Figure 4 could be a supplementary figure.

The manuscript has reached its limit in terms of the number of figures and tables. We'll ask the editorial team if combining figures 1 and 2 into a single figure is feasible. This would allow us to include the Supplementary Figure.

Overestimation of SARS-CoV-2 household transmission in settings of high community transmission: insights from an informal settlement community in Salvador, Brazil

Running title: SARS-CoV-2 BA.1* Omicron variant household transmission

Authors: Juan P. Aguilar Ticona[1,2,3], Nivison Nery Jr.[1,2,3], Matt Hitchings[4], Emilia M. M. Andrade Belitardo[2], Mariam O. Fofana[3], Murilo Dorión[3], Renato Victoriano[2], Jaqueline S. Cruz[2], Juliet Oliveira Santana[2], Laise Eduarda Paixão de Moraes[2], Cristiane W. Cardoso[2,5], Guilherme S. Ribeiro[2,6], Mitermayer G. Reis[2,3,6], Ricardo Khouri[2], Federico Costa[1,2,3], Albert I. Ko[2,3*], Derek A.T. Cummings[7,8*]

Affiliations:

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 Biostatistics, University of Florida, Gainesville, FL, United States.
5. Secretaria Municipal de Saúde de Salvador, Salvador, Brazil.
6. Faculdade de Medicina da Bahia, Universidade Federal da Bahia, Salvador, BA, Brazil.
7. Department of Biology, University of Florida, Gainesville, FL, United States.
8. Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States.

Footnote Page

Potential conflicts 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: This work was supported by grants from the National Institutes of Health

(https://www.nih. gov; R01 ~~AI052473, U01AI088752, R01 TW009504~~AI174105 and ~~R25 TW009338~~R01 AI121207 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) , the Raj and Indra Nooyi Professorship, the Sendas Family ~~Fund~~and Beatrice Kleinberg Neuwirth Funds 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.

* Contributed equally to this work with

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

Summary

We found a high level of transmissibility of the SARS-CoV-2 BA.1* Omicron variant both within the community and households. Phylogenetic analysis suggests a diverse set of viruses were transmitted within the community and households, consistent with multiple introductions and high rates of incidence.

Abstract

Background

The SARS-CoV-2 Omicron variant has spread globally. However, the contribution of community versus household transmission to the overall risk of infection remains unclear.

Methods

Between November 2021, and March 2022, we conducted an active case-finding study in an urban informal settlement with biweekly visits across 1174 households with 3364 residents. Individuals displaying COVID-19-related symptoms were identified, interviewed along with household contacts, and defined as index and secondary cases based on RT-PCR and symptom onset.

Results

In 61 households, we detected a total of 94 RT-PCR-positive cases. Out of 69 sequenced samples, 67 cases (97.1%) were attributed to the Omicron BA.1* variant. Among 35 of their households, the secondary attack rate was 50.0%(% (95%CI 37.0–63.0%). Women (p=RR = 1.6; 95%CI = 0.079 – 2.7), older individuals (p=0.03)median difference = 15; 95%CI = 2 – 21), and those reporting symptoms (p=RR = 1.73; 95% CI = 1.0 – 3.0.04) had a significantly increased risk for SARS-CoV-2 secondary infection. Genomic analysis revealed substantial acquisition of viruses from the community even among households with other SARS-CoV-2 infections. After excluding community acquisition, we estimated a household secondary attack rate of 24.2%(% (95%CI 11.9–40.9%).

Conclusions

These findings underscore the ongoing risk of community acquisition of SARS-CoV-2 among households with current infections. The observed high attack rate necessitates swift booster vaccination, rapid testing availability, and therapeutic options to mitigate COVID-19's severe outcomes.

Keywords: SARS-CoV-2, Omicron, BA.1, Household transmission.

Introduction

The emergence of new SARS-CoV-2 variants poses a significant challenge to public health efforts to control the pandemic. Although the Omicron variant of concern (VOC) has been linked to lower disease severity [1-3], it has exhibited an unprecedented degree of immune escape, resulting in a high burden of infection even among populations with prior infections and high vaccination coverage [3-6]. Furthermore, the rapid spread of the Omicron variant suggests that it is more transmissible than previous variants, which has important implications for public health control measures [7]. For example, in densely populated settings like informal urban settlements, there is a possibility of a high proportion of secondary infections within households once one resident is infected. This situation could ~~reduce~~diminish the effectiveness of home isolation as a ~~means~~method of controlling transmission ~~control. This may be particularly problematic in densely populated settings such as informal urban settlements.~~in these settings without proper planning.

Previous studies have estimated the household secondary attack rate (SAR) of the BA.1 and BA.2 Omicron variant as ranging from 25% to 81% [4, 8-10]. However, it remains unclear to what extent multiple infections within a household are driven by transmission within the household versus high transmission in the community. Distinguishing household from community-based transmission can be particularly difficult in large outbreaks that spread rapidly among communities as cases both within and between households are clustered in time. Understanding the relative contributions of household and community transmission is crucial for providing appropriate recommendations for infection control. Therefore, we conducted a study to estimate the household secondary transmission of SARS-CoV-2 during the BA.1 Omicron wave (December 2021 to March 2022) in an urban informal settlement in Brazil. We conducted active case-finding of cases within an existing cohort and next generation sequencing (NGS) analysis to determine if pairs of cases within households and the community were consistent with transmission.

Methods

Study setting, design and participants

We conducted this study as a part of an ongoing cohort study in Pau da Lima, an urban informal settlement (in Brazil, commonly called favela) situated in Salvador, the largest city in the northeast region of Brazil. Major characteristics of the informal settlement area have been described in previous studies [11-~~13~~14] as well as the high willingness for vaccination and the social determinants of vaccine status [13, 15]. Briefly, the study area had 3,364 inhabitants residing in 1,174 households in an area of 0.35 km² comprised of 3 valleys as identified in a previous census conducted in 2021 (Figure 1A and 1B and Supplementary Figure 1). In December

2021, there was an increase in COVID-19 cases in Salvador associated with the circulation of the SARS-CoV-2 BA.1* Omicron variant (Figure 2A).

From November 11, 2021, to March 21, 2022, trained field technicians visited households in the study area every two weeks to identify and recruit eligible participants During each visit, initially, a standardized questionnaire was administered to the head of the household or any adult in the household to identify any residents showing symptoms associated with COVID-19 illness and to identify their household contacts. Symptomatic cases were defined as participants that reported fever, cough, general weakness/fatigue, headache, myalgia, sore throat, coryza, dyspnea, anorexia/nausea/vomiting, diarrhea, and/or altered mental status ~~[14].~~[16]. If any symptomatic resident was identified in the household, the study teams performed an individual interview and collection of anterior nasal swabs for all household members, including those without symptoms. The individual interview aimed to assess sociodemographic characteristics, presence and persistence of symptoms, use of health services, and vaccination status. A second visit was scheduled seven days after the initial visit to identify newly symptomatic residents and collect a second nasal swab from each household member. Participants were included based on their availability, and multiple attempts, including weekends, were made to limit missing data across the three valleys comprising the study area.

Molecular analysis

Samples collected from symptomatic and asymptomatic household members were tested by real-time reverse transcription-polymerase chain reaction (RT-PCR) to determine SARS-CoV-2 infection status. Positive RT-PCR samples were then subjected to NGS using the Illumina method to identify any variants of concern (VOCs) and/or variants of interest (VOIs). Both RT-PCR and NGS, were conducted by the COVID-19 Platform of FIOCRUZ-BA in Brazil.

To perform the phylogenetic analysis, we selected Omicron lineage sequences (BA.1*) from study participants with primer coverage greater than 90%. We compared these sequences with sequences from the city of in Salvador that were collected during September 15, 2021, and March 21, 2022, which were stored in the GISAID database. We performed a multiple sequence alignment (MSA) by using the Fast Fourier Transform (MAFFT v7.505) online alignment server. The aligned genomes were ranked based on their similarity. We used FigTree v1.4.4 to draw the tree and color the tips according to the households they belonged to. We inferred a maximum likelihood tree from the resulting alignment using the general time-reversible (GTR) substitution model. Additionally, we generated 1,000 bootstrap replicates using IQ-TREE v2.2.0.3 (see Supplementary Methods for details).

Case definitions

We defined an index symptomatic household case as the resident who reported the earliest onset of symptoms among household participants. Co-index cases were among two or more household members with symptom onset on the same date. Household contacts were individuals living in the same household as index cases during the 7 days after the onset of symptoms in the index case. After performing the PCR, index and secondary cases were confirmed. Those household contacts who also tested positive for SARS-CoV-2 were then classified as either symptomatic or asymptomatic secondary cases.

Data analysis

We analyzed data using R version 4.2.2 (https://www.r-project.org) software. We used medians and interquartile ranges (IQRs) for numeric variables and frequency and proportions for categorical variables. For bivariate analysis, we used the χ^2 or Fisher tests to compare categorical variables and t-test or Wilcoxon test to compare continuous variables. Finally, we estimated 95% confidence intervals (95% Cis) and considered a p-value <0.05 significant.

Data analysis: Secondary attack rate

The secondary attack rate (SAR) was calculated by dividing the number of secondary cases by the total number of non-index household residents. Household with co-index were excluded in the calculation of the SAR. We then stratified the SAR by age and sex to evaluate the transmission rate in different groups and identify potential risk factors associated with transmission by calculating the Relative Risk (RR) and the 95% CIs.

Data analysis: Genetic similarity analysis

To assess transmission of SARS-CoV-2 within households and the community, we used genomes obtained from whole genome sequencing analysis of the virus in the areas of Pau da Lima and Salvador. We constructed a genetic dissimilarity matrix and converted it into a similarity matrix using multidimensional scaling (by exponentiating the values) using the "smacof" package in the R software (https://www.rdocumentation.org/packages/smacof/versions/2.1-5). In this matrix, low values reflect dissimilar sequences, while high values reflect a high degree of pairwise genetic similarity (see Supplementary Methods for details~~).~~.).

To determine the threshold for pairwise genetic similarity between participants that was associated with close transmission, we analyzed three groups of sequences. The first group consisted of all individuals in the same household who had more than one confirmed case of SARS-CoV-2. We assumed that this group had a high probability of household transmission (Pau da Lima household group). The second group included one participant randomly selected from each household, or the only positive case in the household (Pau da Lima non-household group). The third group included confirmed cases from Salvador, from the same time period as our active

surveillance. We analyzed the pairwise genetic similarity within the three groups, based on their temporal opportunities for transmission. Then we calculated pairwise similarities and plotted the distribution between the groups. We identified a threshold associated with transmission as the level of genetic dissimilarity at which the cumulative distribution functions of pairwise similarities of within household pairs and non-household pairs visually departed from each other. We then plotted the results of the close transmission analysis using a network graph using Gephi software v0.9.1, to identify possible household transmission among participants~~.~~.

Ethics and patient consent Statements

The study was approved by the Ethics Committee of the Institute of Collective Health of the Federal University of Bahia (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 (no. 2000031554). Adult participants provided a signed informed consent form in the presence of a witness. For participants under 18 years of age, the consent of a parent or legal guardian was required for participation in the study. Children aged 6 years or older also provided written assent to study participation.

Results

We conducted a total of eight rounds of biweekly household visits, during which 1098 out of 1174 households (94%) participated in at least one of the visits (Figure2B and 2C). In total 56-85 % of the household were visited in each round (Supplementary table 1). Among these households, 258 (24%) had at least one symptomatic resident, and among them, at least one positive case for SARS-CoV-2 by PCR was identified in 61 (27%) households (Supplementary Figure ~~1~~2). In these households, we identified 94 individuals who tested positive for SARS-CoV-2 by RT-PCR, with 83 of them being symptomatic and 11 asymptomatic (~~Figure 2D and~~ Supplementary Figure ~~1~~2).

NGS analysis was conducted on 69 (73.4%) out of the 94 SARS-CoV-2 PCR-positive samples. The Omicron BA.1* variant was detected in 67 (97.1%) cases, all of which were linked to samples collected between January and February 2022 (Figure 2D). The remaining two cases (2.9%) were identified as the Delta variant and were linked to samples collected in December 2021.

To evaluate the SAR, we selected a subsample of 35 households with two or more residents and with at least one documented case of Omicron BA.1*. Households with residents who were infected with delta variant and households without a confirmed PCR index case or co-index cases were excluded. (Supplementary Figure ~~1~~2). In total, we identified 35 index cases, 31 secondary cases, and 31 contacts that were negative for SARS-CoV-2 among these households. no cases

were detected on day 14 visit or later. The crude secondary household attack rate was 50.0% (95%CI 37.8–62.2%). Individuals aged between 36 to 60 years old and females showed a higher SAR and risk ratio than younger individuals (<= 18 years old) and males (Table 1). ~~The~~

A description of the contacts recruited is present in Table 2. Among 62 contacts, 50 (80.7%) received at least one COVID-19 vaccine dose~~.~~, similar to the participants in the major cohort (Supplementary table 2). Out of the 40 individuals who participated in the major cohort study and had documented previous exposure, 35 (87.5%) presented a positive IgG test result. The comparison between 31 PCR-positive (secondary cases) and 31 PCR-negative household contacts revealed that individuals with secondary SARS-CoV-2 infection were more frequently female (20/31 [64.5%] female, vs. 11/31 [35.5%] male; ~~p=~~RR = 1.6; 95%CI = 0.~~074~~9 – 2.7) and older (median age of 37 years [IQR 20–43] vs. 22 years [~~13~~15–31]; ~~p=0.039~~median difference = 15; 95%CI = 2 – 21) than negative contacts (Table 2). However, the risk of secondary transmission did not vary based on vaccination status, prior infection nor other household-level factors (Table 2 and Supplementary Table ~~1~~3).

We included 62 (67.4%) confirmed SARS-CoV-2 Omicron BA.1* sequences in the phylogenetic analysis. This set comprised a subgroup of 33 sequences from 14 households with more than one PCR-positive individual, which allowed us to evaluate the frequency with which household members had virus whose sequence was consistent with transmission between pairs. Furthermore, we identified the presence of single nucleotide polymorphisms (SNPs) as R346K, I431M and L450F (Supplementary Figure ~~2~~3). When comparing the sequences from Pau da Lima to 742 sequences from Salvador, all of them belonged to the same genetic clusters (Figure 3A). The similarity metric traversed values of 1 (the largest distance) and a value of 29 as a low distance in pairwise genomic comparisons. The sequences from household pairs in Pau da Lima demonstrated high similarity when compared to sequences from Salvador or the non-household sequences (one sequence selected per household) from Pau da Lima (Figure 3B). In contrast, the genomic similarity between non-household sequences and Salvador city was similar (Figure 3B). Briefly, this means that there is a notable similarity in sequences among households with two or more infected participants when compared to sequences from households with a single infected participant or Pau da Lima or samples from Salvador. This similarity provides evidence supporting household transmission. Finally, we defined a threshold of similarity of ≥2, based on where the cumulative distribution of pairwise differences among pairs departed among household pairs in Pau da Lima compared to non-household pairs and pairs from Salvador (Figure 3C).

We identified high similarity and interrelation between the viral sequences from this community, leading to the identification of seven clusters of SARS-CoV-2 community transmission (Figure 4). Within these clusters, we found 14 households with more than one PCR-positive individual

and we identified 14 index cases, along with 19 PCR-positive contacts and 14 PCR-negative contacts. However, only 8 secondary transmissions could be confirmed by the similarity analysis as resulting from household transmission. The estimated secondary attack rate using the definition based on phylogenetic data was 24.2% (95%CI 11.9 – 40.9%) (Supplementary Figure ~~3~~4). Finally, we performed a sensitivity analysis comparing households with at least one positive PCR against those that reported no symptoms or tested negative. No differences were observed in terms of sex, age, and the mean number of participants under 18 years old, demonstrating the representativeness of the participants. However, there was a difference in the number of residents reported by the head of the household, especially in houses with more than seven residents (Supplementary Table 4).

Discussion

Our findings show that the SARS-CoV-2 Omicron variant was highly transmissible in a community that had near-universal previous exposure to SARS-CoV-2 infection and/or vaccination. The secondary household attack rate was 50%, and there was no difference in the risk of secondary transmission based on vaccination status or prior infection. During the same period, we compared cases from our community to those from the city of Salvador and determined that rapid transmission and multiple introductions contributed to the high attack rate within our community. Furthermore, we found that a high proportion of infections that were identified as secondary cases in the household investigation could be attributed to community transmission based on the genomic similarity analysis.

Like other informal settlements, Pau da Lima community is characterized by poverty, overcrowding, and poor sanitation ~~[11]~~[11]. ~~These structural factors~~Previous studies in developing countries have highlighted that household overcrowding significantly increases the risk of COVID-19 mortality, primarily affecting older individuals residing in crowded households [17, 18]. Although guidelines suggest a two meters of distance among household members and avoid crowded and inadequately ventilated spaces to limit airborne transmission [19, 20], it is challenging in crowded homes. This scenario is representative of urban formal settlements, as showed by research conducted in India, assessing living conditions in large communities [21]. This research identified overcrowding, unsanitary environments, and restricted access to essential services as primary contributors to the rapid spread of COVID-19 [21]. These structural factors were similar in Pau da Lima where they were associated with a high seroprevalence (48%) during the first wave of SARS-CoV-2 transmission in Brazil [~~15~~22]. During the initial period of the Omicron wave associated with the BA.1 variant, we observed an elevated secondary household attack rate (50%) compared to previous variants ~~[16].~~[23]. This is in line with the literature that shows a high transmissibility of the Omicron variant in diverse settings, including high-income

countries ~~[17].~~[24]. Two previous studies in South Korea reported secondary attack rates exceeding 50% [~~18, 19~~25, 26], and a U.S. study reported household transmission ranging from 40.9% among individuals with previous infections to 59.8% among those without [10]. To date, evidence from low- and middle-income countries has been scarce ~~[20].~~[27]. It is important to determine the main transmission patterns of COVID-19 in communities in order to develop effective preventive strategies. Typically, the household SAR is used to estimate the transmissibility of respiratory viruses such as influenza, but this method may overestimate transmissibility if outside sources of infection are not taken into account [~~21, 22~~28, 29], particularly if outbreaks in communities are temporally clustered, driving the time scales of household outbreaks and the overall community outbreak to overlap. Our study found evidence of significant community transmission by analyzing the genomic similarities between household members and confirmed cases in the community study site and in Salvador city. By conducting detailed contact tracing and analyzing genomic data, we were able to identify genetically similar viruses within households and better understand transmission patterns. In this analysis, roughly half of putative household transmission pairs were genetically inconsistent with transmission, substantially revising the risk of household acquisition versus community acquisition.

Given the high rate of household transmission and in the community, it may be necessary to recommend additional protective measures, improved ventilation in households and reevaluated the home isolation during the infectious period in urban informal settlements. Despite the unexpected catastrophic nature of the COVID-19 pandemic, our results emphasize the urgent need for health policies that prioritize equity, especially those supporting urban informal settlements. This demands active engagement from both government and the community as described by Corburn et. al. [30] and Nix E. et al [31]. Government and communities should provide support to elevate living standards, upgrade water, sanitation, and hygiene, alongside improved home ventilation which can also impact other infectious diseases. Community mobilization is also crucial for effective intervention. For instance, community involvement in contact tracing efforts becomes pivotal in identifying potential cases within households and the broader community. Another approach involves immediate and small-scale interventions, such as providing air filters, cooling systems, subsidies for electricity, or access to cooler spaces like community centers. These immediate interventions aim to address the pressing needs and improve conditions swiftly. However, there's a long-term need to address the poor housing conditions in these settlements [31].

The high transmission of the BA.1* Omicron variant observed in our study population emphasizes the level of immune evasion by the new variants and the resulting challenges for transmission control. In our study population, 81% of participants had received at least one vaccine dose, and at least 50% had a previous SARS-CoV-2 infection during the first wave of the pandemic in Brazil

[11]. These findings are in line with the literature, where the effectiveness of vaccination decreased since the old variants until Omicron [4, ~~16~~23]. Furthermore, several SNPs identified in the isolates from our study were associated with high immune evasion, including the R346K mutation in the RBD, which is associated with weakened neutralizing antibody response [~~23-25~~32-34].

In a systematic review of 57 studies, 43 mainly examined the Household SAR. The authors of this review indicate that disregarding external sources of infection might lead to an overestimation of SAR within households. The absence of comparisons between secondary and community infections when estimating SAR was acknowledged as a limitation. Also, none of the reviewed studies utilized techniques like WGS to confirm genetic similarity between the strains infecting index and subsequent cases within households [28]. In contrast, our study stands out for its use of phylogenetic analysis, crucial in understanding the community and household transmissions (adjusted SAR = 24.2%) in Pau da Lima, Brazil. Analyzing genetic sequences from individuals in Pau da Lima and Salvador revealed a resemblance between the samples, suggesting multiple virus introductions into this community, making it representative of Salvador city. Despite the absence of clusters in our phylogenetic analysis, our site is representative of the transmission dynamics in Salvador, where 42% of households belong to an urban informal community. Despite limitations in our sequencing scope, we successfully identified transmission clusters within households and the community, highlighting localized virus spread. While acknowledging the need for larger-scale studies to confirm and expand our findings, previous studies utilizing WGS for transmission assessment showed similar outcomes [25, 35].

Our study found that older age and female gender were associated with risk of infection among household contacts. While initial studies conducted prior to the emergence of the Omicron variant showed low prevalence in children and adolescents, as well as low incidence of severe cases and deaths [~~26~~36], the increased number of infections among children in South Africa ~~[27] and the U.K.~~ [37] and the U.K. [~~28, 29~~38, 39] during the beginning of the Omicron wave raised concerns for health authorities. A systematic review on SARS-CoV-2 household transmission found a lower secondary transmission to child contacts compared to adults. Interestingly, individuals older than 60 years were identified as the most susceptible to infection ~~[16].~~[23]. Furthermore, studies Denmark and the UK observed an increased susceptibility with age and that that the transmission and the SAR were higher for the Omicron variant than previous Variants across all age groups [8, ~~30~~40]. The pattern of household risk may reflect which family members are mostly likely to spend time at home, in contact with other family members and potentially in contact with ill household members. Furthermore, unlike previous COVID-19 waves, the reduction in risk perception, the return to normal activities, and the sense of security following vaccination may have led to an increase in risky behaviors, leaving this population more vulnerable when the

Omicron variant emerged. Female participants were also found to be at a higher risk of secondary transmission than male participants, which could be due to social vulnerability factors in urban informal communities [~~15~~22]. For instance, due to their role as primary family caregivers, women may experience a higher intensity of exposure to infections. This increased exposure can be attributed to factors such as longer duration and closer contact while caring for other sick household members[~~31, 32~~41, 42].

There are some potential limitations in this study. First, the sample size in this population study was limited, affecting the study's statistical power, as reflected in the wide ranges in the confidence intervals. Second, whole-genome sequencing was not complete for 18 participants with PCR-confirmed infection. However, all these cases were reported between January and February 2022, and the Omicron variant accounted for more than 95% of the cases in the region during that period; thus, it is plausible that these 18 cases were attributable to the Omicron variant. Thirdly, during the visits, 56-85% of the households were visited every two weeks, based on the availability of the participants. The field team made multiple visits to each household across the three valleys comprising the study area, aiming to minimize losses. Finally, the screening protocol was paused from December 21st, 2021 to January 10^{th,} 2022. It is possible that transmission in the community began during this period, and that these early cases were not included in this study.

The high attack rate observed in this study underscores the urgent need to implement prevention measures. This includes reinforcing preventive practices such as handwashing, and mask use not only outside the household but also when symptomatic household members are identified. Improving structural housing and health conditions in urban informal settlements (e.g., improving ventilation) may also be an important intervention. Our findings demonstrate the need for continued genomic surveillance to not only ~~–~~identify variants and subvariants that represent a hazard to public health, but also for accurate estimation of community and household transmission. ~~–~~Finally, although our results are consistent with existing data on immune evasion of the Omicron variant, it remains crucial to offer booster vaccination and provide access to rapid testing and therapeutics to mitigate the severe outcomes of COVID-19 for vulnerable urban informal residents.

Acknowledgments

The authors would like to thank the residents 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, University of Florida, and Yale University.

References

1. Nyberg T, Ferguson NM, Nash SG, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. Lancet **2022**; 399:1303-12.
2. Maslo C, Friedland R, Toubkin M, Laubscher A, Akaloo T, Kama B. Characteristics and Outcomes of Hospitalized Patients in South Africa During the COVID-19 Omicron Wave Compared With Previous Waves. Jama **2022**; 327:583-4.
3. Silva SJRd, Kohl A, Pena L, Pardee K. Recent insights into SARS-CoV-2 omicron variant. Reviews in Medical Virology **2023**; 33:e2373.
4. López-Muñoz I, Torrella A, Pérez-Quílez O, et al. SARS-CoV-2 Secondary Attack Rates in Vaccinated and Unvaccinated Household Contacts during Replacement of Delta with Omicron Variant, Spain. Emerging Infectious Disease journal **2022**; 28:1999.
5. Pearson CAB, Silal SP, Li MWZ, et al. Bounding the levels of transmissibility & immune evasion of the Omicron variant in South Africa. medRxiv **2021**:2021.12.19.21268038.
6. Madhi SA, Kwatra G, Myers JE, et al. Population immunity and Covid-19 severity with Omicron variant in South Africa. New England Journal of Medicine **2022**.
7. Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature **2022**.
8. Lyngse FP, Mortensen LH, Denwood MJ, et al. Household transmission of the SARS-CoV-2 Omicron variant in Denmark. Nature Communications **2022**; 13:5573.
9. 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**; 327:1610-1.
10. Baker JM, Nakayama JY, O'Hegarty M, 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**; 71:341-6.
11. Hagan JE, Moraga P, Costa F, et al. Spatiotemporal determinants of urban leptospirosis transmission: four-year prospective cohort study of slum residents in Brazil. PLoS neglected tropical diseases **2016**; 10:e0004275.
12. Rodriguez-Barraquer I, Costa F, Nascimento EJM, et al. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. Science **2019**; 363:607-10.
13. Aguilar Ticona JP, Nery N, Jr., Victoriano R, et al. Willingness to Get the COVID-19 Vaccine among Residents of Slum Settlements. Vaccines (Basel) **2021**; 9.
14. Aguilar Ticona JP, Xiao M, Li D, et al. Extensive transmission of SARS-CoV-2 BQ.1* variant in a population with high levels of hybrid immunity: A prevalence survey. International Journal of Infectious Diseases **2024**; 139:159-67.
15. Dorión M, Ticona JPA, Fofana MO, et al. Social determinants of adult COVID-19 vaccine acceptance and uptake in a Brazilian urban informal community: a longitudinal time-to-event study. medRxiv **2023**:2023.11.28.23298927.
16. Organization WH. WHO COVID-19 case definition: World Health Organization, **2020**.
17. Varshney K, Glodjo T, Adalbert J. Overcrowded housing increases risk for COVID-19 mortality: an ecological study. BMC Research Notes **2022**; 15:126.
18. Ghosh AK, Venkatraman S, Soroka O, et al. Association between overcrowded households, multigenerational households, and COVID-19: a cohort study. Public Health **2021**; 198:273-9.
19. Organization WH. Coronavirus disease (COVID-19) advice for the public. 2020, **2020**.
20. CDC. COVID Data Tracker. U.S. Department of Health and Human Services. Available at: https://covid.cdc.gov/covid-data-tracker/#datatracker-home. Accessed Jan 05th 2024.
21. Das M, Das A, Giri B, Sarkar R, Saha S. Habitat vulnerability in slum areas of India – What we learnt from COVID-19? International Journal of Disaster Risk Reduction **2021**; 65:102553.
22. Fofana MO, Nery N, Jr., Aguilar Ticona JP, 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:e1004093.
23. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Household secondary attack rates of SARS-CoV-2 by variant and vaccination status: an updated systematic review and meta-analysis. JAMA network open **2022**; 5:e229317-e.

24. Madewell ZJ, Yang Y, Longini IM, Jr., Halloran ME, Dean NE. Household secondary attack rates of SARS-CoV-2 by variant and vaccination status: an updated systematic review and meta-analysis. medRxiv **2022**.
25. Kim EY, Choe YJ, Park H, et al. Community Transmission of SARS-CoV-2 Omicron Variant, South Korea, 2021. Emerg Infect Dis **2022**; 28:898-900.
26. Song JS, Lee J, Kim M, et al. Serial Intervals and Household Transmission of SARS-CoV-2 Omicron Variant, South Korea, 2021. Emerg Infect Dis **2022**; 28:756-9.
27. Madewell ZJ, Yang Y, Longini IM, Jr., Halloran ME, Dean NE. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis. JAMA Netw Open **2022**; 5:e229317.
28. Koh WC, Naing L, Chaw L, et al. What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. PLOS ONE **2020**; 15:e0240205.
29. Hsu C-Y, Wang J-T, Huang K-C, Fan AC-H, Yeh Y-P, Chen SL-S. Household transmission but without the community-acquired outbreak of COVID-19 in Taiwan. Journal of the Formosan Medical Association **2021**; 120:S38-S45.
30. Know WYNT. People with certain medical conditions.
31. Nix E, Paulose J, Lakhanpaul M, et al. COVID-19 & informal settlements: is 'Stay Home' safe? UCL Open Environ **2022**; 4:e038.
32. Ou J, Lan W, Wu X, et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. Signal Transduction and Targeted Therapy **2022**; 7:138.
33. Liu H, Wei P, Kappler JW, Marrack P, Zhang G. SARS-CoV-2 variants of concern and variants of interest receptor binding domain mutations and virus infectivity. Frontiers in immunology **2022**; 13:50.
34. Mohandas S, Yadav PD, Sapkal G, et al. Pathogenicity of SARS-CoV-2 Omicron (R346K) variant in Syrian hamsters and its cross-neutralization with different variants of concern. eBioMedicine **2022**; 79.
35. Chow K, Aslam A, McClure T, et al. Risk of Healthcare-Associated Transmission of Sever Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Hospitalized Cancer Patients. Clin Infect Dis **2022**; 74:1579-85.
36. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. JAMA Pediatr **2021**; 175:143-56.
37. Cloete J, Kruger A, Masha M, et al. Paediatric hospitalisations due to COVID-19 during the first SARS-CoV-2 omicron (B.1.1.529) variant wave in South Africa: a multicentre observational study. The Lancet Child & Adolescent Health **2022**; 6:294-302.
38. Sheikh A, Kerr S, Woolhouse M, McMenamin J, Robertson C. Severity of omicron variant of concern and effectiveness of vaccine boosters against symptomatic disease in Scotland (EAVE II): a national cohort study with nested test-negative design. Lancet Infect Dis **2022**; 22:959-66.
39. Ferguson N, Ghani A, Hinsley W, Volz E. Report 50: Hospitalisation risk for Omicron cases in England. **2021**.
40. Frederik Plesner L, Kåre M, Kristina Træholt F, et al. Association between SARS-CoV-2 Transmissibility, Viral Load, and Age in Households. medRxiv **2021**:2021.02.28.21252608.
41. Ogata T, Tanaka H, Tanaka E, et al. Increased Secondary Attack Rates among the Household Contacts of Patients with the Omicron Variant of the Coronavirus Disease 2019 in Japan. Int J Environ Res Public Health **2022**; 19.
42. Wei Z, Ma W, Wang Z, et al. Household transmission of SARS-CoV-2 during the Omicron wave in Shanghai, China: A case-ascertained study. Influenza and Other Respiratory Viruses **2023**; 17:e13097.

Figures legends

Figure 1. Study setting. A) Image of the study area, with inset depicting the location of Salvador and Bahia state within Brazil; B) Location of households in the study area with no symptomatic resident (blue dots), no PCR+ resident (gray dots), or at least one PCR+ resident (red dots~~.~~; and C) yellow dots represent the 14 households with > 1 resident included in the phylogentic analysis

Figure 2. Study period and visits A) weekly new cases of COVID-19 in Salvador, Brazil; B) Number of participants screened and proportion with symptoms C) Number of participants tested classified as contacts and symptomatic index cases; and D) Number of participants in households with >1 PCR+ resident.

* No Omicron variants were detected in November and December 2021. Only two Delta cases were confirmed, these PCR+ results were not included in the SAR analysis.

Figure 3. A) Time-resolved maximum likelihood (ML) phylogenetic tree of the SARS-CoV-2 Omicron BA.1 in Salvador including 62 Omicron BA.1 isolates obtained in this study and an additional 742 representative BA.1 genomes collected throughout the city of Salvador up to March 21st, 2022. Colored circles represent participants from 14 households with > 1 resident included in the analysis and ~~grey~~small white circles represent households with a single participant. Branches with no circles represent the genomes collected from GISAID. B) Genomic similarity among groups, C) Proportion of pairs identified at varying genetic similarity thresholds

Figure 4. Genetic similarity network of SARS-CoV-2 isolates among study households. Nodes represent individual SARS-CoV-2 sequences and edge weights represent the dissimilarity values between each pair of sequences. The colored nodes on the plot represent sequences from the Pau da Lima community, which are distributed across six transmission clusters indicated by the color of the nodes. Sequences with labels belong to households with more than one individual included in this analysis. Red labels indicate potential household transmission based on several household members belonging to the same cluster. The nodes without labels represent sequences from households with a single participant included in the analysis. The lines on the plot indicate genomic similarity (the threshold for genomic similarity is set at >2), with thicker lines representing higher degrees of similarity between sequence pairs. Node size represents the value calculated for betweenness centrality, indicating the amount of influence a node has over the flow of information in the graph.

Formatted: Left

Formatted: Font color: Auto

Overestimation of SARS-CoV-2 household transmission in settings of high community transmission: insights from an informal settlement community in Salvador, Brazil

Running title: SARS-CoV-2 BA.1* Omicron variant household transmission

Authors: Juan P. Aguilar Ticona[1,2,3], Nivison Nery Jr.[1,2,3], Matt Hitchings[4], Emilia M. M. Andrade Belitardo[2], Mariam O. Fofana[3], Murilo Dorión[3], Renato Victoriano[2], Jaqueline S. Cruz[2], Juliet Oliveira Santana[2], Laise Eduarda Paixão de Moraes[2], Cristiane W. Cardoso[2,5], Guilherme S. Ribeiro[2,6], Mitermayer G. Reis[2,3,6], Ricardo Khouri[2], Federico Costa[1,2,3], Albert I. Ko[2,3*], Derek A.T. Cummings[7,8*]

Affiliations:

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 Biostatistics, University of Florida, Gainesville, FL, United States.
5. Secretaria Municipal de Saúde de Salvador, Salvador, Brazil.
6. Faculdade de Medicina da Bahia, Universidade Federal da Bahia, Salvador, BA, Brazil.
7. Department of Biology, University of Florida, Gainesville, FL, United States.
8. Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States.

Footnote Page

Potential conflicts 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: This work was supported by grants from the National Institutes of Health

(https://www.nih. gov; R01 AI174105 and R01 AI121207 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) , the Raj and Indra Nooyi Professorship, the Sendas Family and Beatrice Kleinberg Neuwirth Funds 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.

* Contributed equally to this work with

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

Summary

We found a high level of transmissibility of the SARS-CoV-2 BA.1* Omicron variant both within the community and households. Phylogenetic analysis suggests a diverse set of viruses were transmitted within the community and households, consistent with multiple introductions and high rates of incidence.

Abstract

Background

The SARS-CoV-2 Omicron variant has spread globally. However, the contribution of community versus household transmission to the overall risk of infection remains unclear.

Methods

Between November 2021, and March 2022, we conducted an active case-finding study in an urban informal settlement with biweekly visits across 1174 households with 3364 residents. Individuals displaying COVID-19-related symptoms were identified, interviewed along with household contacts, and defined as index and secondary cases based on RT-PCR and symptom onset.

Results

In 61 households, we detected a total of 94 RT-PCR-positive cases. Out of 69 sequenced samples, 67 cases (97.1%) were attributed to the Omicron BA.1* variant. Among 35 of their households, the secondary attack rate was 50.0% (95%CI 37.0–63.0%). Women (RR = 1.6; 95%CI = 0.9 – 2.7), older individuals (median difference = 15; 95%CI = 2 – 21), and those reporting symptoms (RR = 1.73; 95% CI = 1.0 – 3.0) had a significantly increased risk for SARS-CoV-2 secondary infection. Genomic analysis revealed substantial acquisition of viruses from the community even among households with other SARS-CoV-2 infections. After excluding community acquisition, we estimated a household secondary attack rate of 24.2% (95%CI 11.9–40.9%).

Conclusions

These findings underscore the ongoing risk of community acquisition of SARS-CoV-2 among households with current infections. The observed high attack rate necessitates swift booster vaccination, rapid testing availability, and therapeutic options to mitigate COVID-19's severe outcomes.

Keywords: SARS-CoV-2, Omicron, BA.1, Household transmission.

Introduction

The emergence of new SARS-CoV-2 variants poses a significant challenge to public health efforts to control the pandemic. Although the Omicron variant of concern (VOC) has been linked to lower disease severity [1-3], it has exhibited an unprecedented degree of immune escape, resulting in a high burden of infection even among populations with prior infections and high vaccination coverage [3-6]. Furthermore, the rapid spread of the Omicron variant suggests that it is more transmissible than previous variants, which has important implications for public health control measures [7]. For example, in densely populated settings like informal urban settlements, there is a possibility of a high proportion of secondary infections within households once one resident is infected. This situation could diminish the effectiveness of home isolation as a method of controlling transmission in these settings without proper planning.

Previous studies have estimated the household secondary attack rate (SAR) of the BA.1 and BA.2 Omicron variant as ranging from 25% to 81% [4, 8-10]. However, it remains unclear to what extent multiple infections within a household are driven by transmission within the household versus high transmission in the community. Distinguishing household from community-based transmission can be particularly difficult in large outbreaks that spread rapidly among communities as cases both within and between households are clustered in time. Understanding the relative contributions of household and community transmission is crucial for providing appropriate recommendations for infection control. Therefore, we conducted a study to estimate the household secondary transmission of SARS-CoV-2 during the BA.1 Omicron wave (December 2021 to March 2022) in an urban informal settlement in Brazil. We conducted active case-finding of cases within an existing cohort and next generation sequencing (NGS) analysis to determine if pairs of cases within households and the community were consistent with transmission.

Methods

Study setting, design and participants

We conducted this study as a part of an ongoing cohort study in Pau da Lima, an urban informal settlement (in Brazil, commonly called favela) situated in Salvador, the largest city in the northeast region of Brazil. Major characteristics of the informal settlement area have been described in previous studies [11-14] as well as the high willingness for vaccination and the social determinants of vaccine status [13, 15]. Briefly, the study area had 3,364 inhabitants residing in 1,174 households in an area of 0.35 km² comprised of 3 valleys as identified in a previous census conducted in 2021 (Figure 1A and 1B and Supplementary Figure 1). In December 2021, there was an increase in COVID-19 cases in Salvador associated with the circulation of the SARS-CoV-2 BA.1* Omicron variant (Figure 2A).

From November 11, 2021, to March 21, 2022, trained field technicians visited households in the study area every two weeks to identify and recruit eligible participants During each visit, initially, a standardized questionnaire was administered to the head of the household or any adult in the household to identify any residents showing symptoms associated with COVID-19 illness and to identify their household contacts. Symptomatic cases were defined as participants that reported fever, cough, general weakness/fatigue, headache, myalgia, sore throat, coryza, dyspnea, anorexia/nausea/vomiting, diarrhea, and/or altered mental status [16]. If any symptomatic resident was identified in the household, the study teams performed an individual interview and collection of anterior nasal swabs for all household members, including those without symptoms. The individual interview aimed to assess sociodemographic characteristics, presence and persistence of symptoms, use of health services, and vaccination status. A second visit was scheduled seven days after the initial visit to identify newly symptomatic residents and collect a second nasal swab from each household member. Participants were included based on their availability, and multiple attempts, including weekends, were made to limit missing data across the three valleys comprising the study area.

Molecular analysis

Samples collected from symptomatic and asymptomatic household members were tested by real-time reverse transcription-polymerase chain reaction (RT-PCR) to determine SARS-CoV-2 infection status. Positive RT-PCR samples were then subjected to NGS using the Illumina method to identify any variants of concern (VOCs) and/or variants of interest (VOIs). Both RT-PCR and NGS, were conducted by the COVID-19 Platform of FIOCRUZ-BA in Brazil.

To perform the phylogenetic analysis, we selected Omicron lineage sequences (BA.1*) from study participants with primer coverage greater than 90%. We compared these sequences with sequences from the city of in Salvador that were collected during September 15, 2021, and March 21, 2022, which were stored in the GISAID database. We performed a multiple sequence alignment (MSA) by using the Fast Fourier Transform (MAFFT v7.505) online alignment server. The aligned genomes were ranked based on their similarity. We used FigTree v1.4.4 to draw the tree and color the tips according to the households they belonged to. We inferred a maximum likelihood tree from the resulting alignment using the general time-reversible (GTR) substitution model. Additionally, we generated 1,000 bootstrap replicates using IQ-TREE v2.2.0.3 (see Supplementary Methods for details).

Case definitions

We defined an index symptomatic household case as the resident who reported the earliest onset of symptoms among household participants. Co-index cases were among two or more household members with symptom onset on the same date. Household contacts were individuals living in

the same household as index cases during the 7 days after the onset of symptoms in the index case. After performing the PCR, index and secondary cases were confirmed. Those household contacts who also tested positive for SARS-CoV-2 were then classified as either symptomatic or asymptomatic secondary cases.

Data analysis

We analyzed data using R version 4.2.2 (https://www.r-project.org) software. We used medians and interquartile ranges (IQRs) for numeric variables and frequency and proportions for categorical variables. For bivariate analysis, we used the χ^2 or Fisher tests to compare categorical variables and t-test or Wilcoxon test to compare continuous variables. Finally, we estimated 95% confidence intervals (95% Cis) and considered a p-value <0.05 significant.

Data analysis: Secondary attack rate

The secondary attack rate (SAR) was calculated by dividing the number of secondary cases by the total number of non-index household residents. Household with co-index were excluded in the calculation of the SAR. We then stratified the SAR by age and sex to evaluate the transmission rate in different groups and identify potential risk factors associated with transmission by calculating the Relative Risk (RR) and the 95% CIs.

Data analysis: Genetic similarity analysis

To assess transmission of SARS-CoV-2 within households and the community, we used genomes obtained from whole genome sequencing analysis of the virus in the areas of Pau da Lima and Salvador. We constructed a genetic dissimilarity matrix and converted it into a similarity matrix using multidimensional scaling (by exponentiating the values) using the "smacof" package in the R software (https://www.rdocumentation.org/packages/smacof/versions/2.1-5). In this matrix, low values reflect dissimilar sequences, while high values reflect a high degree of pairwise genetic similarity (see Supplementary Methods for details).

To determine the threshold for pairwise genetic similarity between participants that was associated with close transmission, we analyzed three groups of sequences. The first group consisted of all individuals in the same household who had more than one confirmed case of SARS-CoV-2. We assumed that this group had a high probability of household transmission (Pau da Lima household group). The second group included one participant randomly selected from each household, or the only positive case in the household (Pau da Lima non-household group). The third group included confirmed cases from Salvador, from the same time period as our active surveillance. We analyzed the pairwise genetic similarity within the three groups, based on their temporal opportunities for transmission. Then we calculated pairwise similarities and plotted the distribution between the groups. We identified a threshold associated with transmission as the

level of genetic dissimilarity at which the cumulative distribution functions of pairwise similarities of within household pairs and non-household pairs visually departed from each other. We then plotted the results of the close transmission analysis using a network graph using Gephi software v0.9.1, to identify possible household transmission among participants.

Ethics and patient consent Statements

The study was approved by the Ethics Committee of the Institute of Collective Health of the Federal University of Bahia (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 (no. 2000031554). Adult participants provided a signed informed consent form in the presence of a witness. For participants under 18 years of age, the consent of a parent or legal guardian was required for participation in the study. Children aged 6 years or older also provided written assent to study participation.

Results

We conducted a total of eight rounds of biweekly household visits, during which 1098 out of 1174 households (94%) participated in at least one of the visits (Figure2B and 2C). In total 56-85 % of the household were visited in each round (Supplementary table 1). Among these households, 258 (24%) had at least one symptomatic resident, and among them, at least one positive case for SARS-CoV-2 by PCR was identified in 61 (27%) households (Supplementary Figure 2). In these households, we identified 94 individuals who tested positive for SARS-CoV-2 by RT-PCR, with 83 of them being symptomatic and 11 asymptomatic (Supplementary Figure 2).

NGS analysis was conducted on 69 (73.4%) out of the 94 SARS-CoV-2 PCR-positive samples. The Omicron BA.1* variant was detected in 67 (97.1%) cases, all of which were linked to samples collected between January and February 2022 (Figure 2D). The remaining two cases (2.9%) were identified as the Delta variant and were linked to samples collected in December 2021.

To evaluate the SAR, we selected a subsample of 35 households with two or more residents and with at least one documented case of Omicron BA.1*. Households with residents who were infected with delta variant and households without a confirmed PCR index case or co-index cases were excluded. (Supplementary Figure 2). In total, we identified 35 index cases, 31 secondary cases, and 31 contacts that were negative for SARS-CoV-2 among these households. no cases were detected on day 14 visit or later. The crude secondary household attack rate was 50.0% (95%CI 37.8–62.2%). Individuals aged between 36 to 60 years old and females showed a higher SAR and risk ratio than younger individuals (<= 18 years old) and males (Table 1).

A description of the contacts recruited is present in Table 2. Among 62 contacts, 50 (80.7%) received at least one COVID-19 vaccine dose, similar to the participants in the major cohort (Supplementary table 2). Out of the 40 individuals who participated in the major cohort study and had documented previous exposure, 35 (87.5%) presented a positive IgG test result. The comparison between 31 PCR-positive (secondary cases) and 31 PCR-negative household contacts revealed that individuals with secondary SARS-CoV-2 infection were more frequently female (20/31 [64.5%] female, vs. 11/31 [35.5%] male; RR = 1.6; 95%CI = 0.9 – 2.7) and older (median age of 37 years [IQR 20–43] vs. 22 years [15–31]; median difference = 15; 95%CI = 2 – 21) than negative contacts (Table 2). However, the risk of secondary transmission did not vary based on vaccination status, prior infection nor other household-level factors (Table 2 and Supplementary Table 3).

We included 62 (67.4%) confirmed SARS-CoV-2 Omicron BA.1* sequences in the phylogenetic analysis. This set comprised a subgroup of 33 sequences from 14 households with more than one PCR-positive individual, which allowed us to evaluate the frequency with which household members had virus whose sequence was consistent with transmission between pairs. Furthermore, we identified the presence of single nucleotide polymorphisms (SNPs) as R346K, I431M and L450F (Supplementary Figure 3). When comparing the sequences from Pau da Lima to 742 sequences from Salvador, all of them belonged to the same genetic clusters (Figure 3A). The similarity metric traversed values of 1 (the largest distance) and a value of 29 as a low distance in pairwise genomic comparisons. The sequences from household pairs in Pau da Lima demonstrated high similarity when compared to sequences from Salvador or the non-household sequences (one sequence selected per household) from Pau da Lima (Figure 3B). In contrast, the genomic similarity between non-household sequences and Salvador city was similar (Figure 3B). Briefly, this means that there is a notable similarity in sequences among households with two or more infected participants when compared to sequences from households with a single infected participant or Pau da Lima or samples from Salvador. This similarity provides evidence supporting household transmission. Finally, we defined a threshold of similarity of ≥2, based on where the cumulative distribution of pairwise differences among pairs departed among household pairs in Pau da Lima compared to non-household pairs and pairs from Salvador (Figure 3C).

We identified high similarity and interrelation between the viral sequences from this community, leading to the identification of seven clusters of SARS-CoV-2 community transmission (Figure 4). Within these clusters, we found 14 households with more than one PCR-positive individual and we identified 14 index cases, along with 19 PCR-positive contacts and 14 PCR-negative contacts. However, only 8 secondary transmissions could be confirmed by the similarity analysis as resulting from household transmission. The estimated secondary attack rate using the definition based on phylogenetic data was 24.2% (95%CI 11.9 – 40.9%) (Supplementary Figure 4). Finally,

we performed a sensitivity analysis comparing households with at least one positive PCR against those that reported no symptoms or tested negative. No differences were observed in terms of sex, age, and the mean number of participants under 18 years old, demonstrating the representativeness of the participants. However, there was a difference in the number of residents reported by the head of the household, especially in houses with more than seven residents (Supplementary Table 4).

Discussion

Our findings show that the SARS-CoV-2 Omicron variant was highly transmissible in a community that had near-universal previous exposure to SARS-CoV-2 infection and/or vaccination. The secondary household attack rate was 50%, and there was no difference in the risk of secondary transmission based on vaccination status or prior infection. During the same period, we compared cases from our community to those from the city of Salvador and determined that rapid transmission and multiple introductions contributed to the high attack rate within our community. Furthermore, we found that a high proportion of infections that were identified as secondary cases in the household investigation could be attributed to community transmission based on the genomic similarity analysis.

Like other informal settlements, Pau da Lima community is characterized by poverty, overcrowding, and poor sanitation [11]. Previous studies in developing countries have highlighted that household overcrowding significantly increases the risk of COVID-19 mortality, primarily affecting older individuals residing in crowded households [17, 18]. Although guidelines suggest a two meters of distance among household members and avoid crowded and inadequately ventilated spaces to limit airborne transmission [19, 20], it is challenging in crowded homes. This scenario is representative of urban formal settlements, as showed by research conducted in India, assessing living conditions in large communities [21]. This research identified overcrowding, unsanitary environments, and restricted access to essential services as primary contributors to the rapid spread of COVID-19 [21]. These structural factors were similar in Pau da Lima where they were associated with a high seroprevalence (48%) during the first wave of SARS-CoV-2 transmission in Brazil [22]. During the initial period of the Omicron wave associated with the BA.1 variant, we observed an elevated secondary household attack rate (50%) compared to previous variants [23]. This is in line with the literature that shows a high transmissibility of the Omicron variant in diverse settings, including high-income countries [24]. Two previous studies in South Korea reported secondary attack rates exceeding 50% [25, 26], and a U.S. study reported household transmission ranging from 40.9% among individuals with previous infections to 59.8% among those without [10]. To date, evidence from low- and middle-income countries has been scarce [27]. It is important to determine the main transmission patterns of COVID-19 in

communities in order to develop effective preventive strategies. Typically, the household SAR is used to estimate the transmissibility of respiratory viruses such as influenza, but this method may overestimate transmissibility if outside sources of infection are not taken into account [28, 29], particularly if outbreaks in communities are temporally clustered, driving the time scales of household outbreaks and the overall community outbreak to overlap. Our study found evidence of significant community transmission by analyzing the genomic similarities between household members and confirmed cases in the community study site and in Salvador city. By conducting detailed contact tracing and analyzing genomic data, we were able to identify genetically similar viruses within households and better understand transmission patterns. In this analysis, roughly half of putative household transmission pairs were genetically inconsistent with transmission, substantially revising the risk of household acquisition versus community acquisition.

Given the high rate of household transmission and in the community, it may be necessary to recommend additional protective measures, improved ventilation in households and reevaluated the home isolation during the infectious period in urban informal settlements. Despite the unexpected catastrophic nature of the COVID-19 pandemic, our results emphasize the urgent need for health policies that prioritize equity, especially those supporting urban informal settlements. This demands active engagement from both government and the community as described by Corburn et. al. [30] and Nix E. et al [31]. Government and communities should provide support to elevate living standards, upgrade water, sanitation, and hygiene, alongside improved home ventilation which can also impact other infectious diseases. Community mobilization is also crucial for effective intervention. For instance, community involvement in contact tracing efforts becomes pivotal in identifying potential cases within households and the broader community. Another approach involves immediate and small-scale interventions, such as providing air filters, cooling systems, subsidies for electricity, or access to cooler spaces like community centers. These immediate interventions aim to address the pressing needs and improve conditions swiftly. However, there's a long-term need to address the poor housing conditions in these settlements [31].

The high transmission of the BA.1* Omicron variant observed in our study population emphasizes the level of immune evasion by the new variants and the resulting challenges for transmission control. In our study population, 81% of participants had received at least one vaccine dose, and at least 50% had a previous SARS-CoV-2 infection during the first wave of the pandemic in Brazil [11]. These findings are in line with the literature, where the effectiveness of vaccination decreased since the old variants until Omicron [4, 23]. Furthermore, several SNPs identified in the isolates from our study were associated with high immune evasion, including the R346K mutation in the RBD, which is associated with weakened neutralizing antibody response [32-34].

In a systematic review of 57 studies, 43 mainly examined the Household SAR. The authors of this review indicate that disregarding external sources of infection might lead to an overestimation of SAR within households. The absence of comparisons between secondary and community infections when estimating SAR was acknowledged as a limitation. Also, none of the reviewed studies utilized techniques like WGS to confirm genetic similarity between the strains infecting index and subsequent cases within households [28]. In contrast, our study stands out for its use of phylogenetic analysis, crucial in understanding the community and household transmissions (adjusted SAR = 24.2%) in Pau da Lima, Brazil. Analyzing genetic sequences from individuals in Pau da Lima and Salvador revealed a resemblance between the samples, suggesting multiple virus introductions into this community, making it representative of Salvador city. Despite the absence of clusters in our phylogenetic analysis, our site is representative of the transmission dynamics in Salvador, where 42% of households belong to an urban informal community. Despite limitations in our sequencing scope, we successfully identified transmission clusters within households and the community, highlighting localized virus spread. While acknowledging the need for larger-scale studies to confirm and expand our findings, previous studies utilizing WGS for transmission assessment showed similar outcomes [25, 35].

Our study found that older age and female gender were associated with risk of infection among household contacts. While initial studies conducted prior to the emergence of the Omicron variant showed low prevalence in children and adolescents, as well as low incidence of severe cases and deaths [36], the increased number of infections among children in South Africa [37] and the U.K. [38, 39] during the beginning of the Omicron wave raised concerns for health authorities. A systematic review on SARS-CoV-2 household transmission found a lower secondary transmission to child contacts compared to adults. Interestingly, individuals older than 60 years were identified as the most susceptible to infection [23]. Furthermore, studies Denmark and the UK observed an increased susceptibility with age and that that the transmission and the SAR were higher for the Omicron variant than previous Variants across all age groups [8, 40]. The pattern of household risk may reflect which family members are mostly likely to spend time at home, in contact with other family members and potentially in contact with ill household members. Furthermore, unlike previous COVID-19 waves, the reduction in risk perception, the return to normal activities, and the sense of security following vaccination may have led to an increase in risky behaviors, leaving this population more vulnerable when the Omicron variant emerged. Female participants were also found to be at a higher risk of secondary transmission than male participants, which could be due to social vulnerability factors in urban informal communities [22]. For instance, due to their role as primary family caregivers, women may experience a higher intensity of exposure to infections. This increased exposure can be attributed to factors such as longer duration and closer contact while caring for other sick household members[41, 42].

There are some potential limitations in this study. First, the sample size in this population study was limited, affecting the study's statistical power, as reflected in the wide ranges in the confidence intervals. Second, whole-genome sequencing was not complete for 18 participants with PCR-confirmed infection. However, all these cases were reported between January and February 2022, and the Omicron variant accounted for more than 95% of the cases in the region during that period; thus, it is plausible that these 18 cases were attributable to the Omicron variant. Thirdly, during the visits, 56-85% of the households were visited every two weeks, based on the availability of the participants. The field team made multiple visits to each householde across the three valleys comprising the study area, aiming to minimize losses. Finally, the screening protocol was paused from December 21st, 2021 to January 10^{th,} 2022. It is possible that transmission in the community began during this period, and that these early cases were not included in this study.

The high attack rate observed in this study underscores the urgent need to implement prevention measures. This includes reinforcing preventive practices such as handwashing, and mask use not only outside the household but also when symptomatic household members are identified. Improving structural housing and health conditions in urban informal settlements (e.g., improving ventilation) may also be an important intervention. Our findings demonstrate the need for continued genomic surveillance to not only identify variants and subvariants that represent a hazard to public health, but also for accurate estimation of community and household transmission. Finally, although our results are consistent with existing data on immune evasion of the Omicron variant, it remains crucial to offer booster vaccination and provide access to rapid testing and therapeutics to mitigate the severe outcomes of COVID-19 for vulnerable urban informal residents.

Acknowledgments

The authors would like to thank the residents 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, University of Florida, and Yale University.

References

1. Nyberg T, Ferguson NM, Nash SG, et al. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. Lancet **2022**; 399:1303-12.
2. Maslo C, Friedland R, Toubkin M, Laubscher A, Akaloo T, Kama B. Characteristics and Outcomes of Hospitalized Patients in South Africa During the COVID-19 Omicron Wave Compared With Previous Waves. Jama **2022**; 327:583-4.
3. Silva SJRd, Kohl A, Pena L, Pardee K. Recent insights into SARS-CoV-2 omicron variant. Reviews in Medical Virology **2023**; 33:e2373.

4. López-Muñoz I, Torrella A, Pérez-Quílez O, et al. SARS-CoV-2 Secondary Attack Rates in Vaccinated and Unvaccinated Household Contacts during Replacement of Delta with Omicron Variant, Spain. Emerging Infectious Disease journal **2022**; 28:1999.
5. Pearson CAB, Silal SP, Li MWZ, et al. Bounding the levels of transmissibility & immune evasion of the Omicron variant in South Africa. medRxiv **2021**:2021.12.19.21268038.
6. Madhi SA, Kwatra G, Myers JE, et al. Population immunity and Covid-19 severity with Omicron variant in South Africa. New England Journal of Medicine **2022**.
7. Viana R, Moyo S, Amoako DG, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature **2022**.
8. Lyngse FP, Mortensen LH, Denwood MJ, et al. Household transmission of the SARS-CoV-2 Omicron variant in Denmark. Nature Communications **2022**; 13:5573.
9. 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**; 327:1610-1.
10. Baker JM, Nakayama JY, O'Hegarty M, 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**; 71:341-6.
11. Hagan JE, Moraga P, Costa F, et al. Spatiotemporal determinants of urban leptospirosis transmission: four-year prospective cohort study of slum residents in Brazil. PLoS neglected tropical diseases **2016**; 10:e0004275.
12. Rodriguez-Barraquer I, Costa F, Nascimento EJM, et al. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. Science **2019**; 363:607-10.
13. Aguilar Ticona JP, Nery N, Jr., Victoriano R, et al. Willingness to Get the COVID-19 Vaccine among Residents of Slum Settlements. Vaccines (Basel) **2021**; 9.
14. Aguilar Ticona JP, Xiao M, Li D, et al. Extensive transmission of SARS-CoV-2 BQ.1* variant in a population with high levels of hybrid immunity: A prevalence survey. International Journal of Infectious Diseases **2024**; 139:159-67.
15. Dorión M, Ticona JPA, Fofana MO, et al. Social determinants of adult COVID-19 vaccine acceptance and uptake in a Brazilian urban informal community: a longitudinal time-to-event study. medRxiv **2023**:2023.11.28.23298927.
16. Organization WH. WHO COVID-19 case definition: World Health Organization, **2020**.
17. Varshney K, Glodjo T, Adalbert J. Overcrowded housing increases risk for COVID-19 mortality: an ecological study. BMC Research Notes **2022**; 15:126.
18. Ghosh AK, Venkatraman S, Soroka O, et al. Association between overcrowded households, multigenerational households, and COVID-19: a cohort study. Public Health **2021**; 198:273-9.
19. Organization WH. Coronavirus disease (COVID-19) advice for the public. 2020, **2020**.
20. CDC. COVID Data Tracker. U.S. Department of Health and Human Services. Available at: https://covid.cdc.gov/covid-data-tracker/#datatracker-home. Accessed Jan 05th 2024.
21. Das M, Das A, Giri B, Sarkar R, Saha S. Habitat vulnerability in slum areas of India – What we learnt from COVID-19? International Journal of Disaster Risk Reduction **2021**; 65:102553.
22. Fofana MO, Nery N, Jr., Aguilar Ticona JP, 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:e1004093.
23. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Household secondary attack rates of SARS-CoV-2 by variant and vaccination status: an updated systematic review and meta-analysis. JAMA network open **2022**; 5:e229317-e.
24. Madewell ZJ, Yang Y, Longini IM, Jr., Halloran ME, Dean NE. Household secondary attack rates of SARS-CoV-2 by variant and vaccination status: an updated systematic review and meta-analysis. medRxiv **2022**.
25. Kim EY, Choe YJ, Park H, et al. Community Transmission of SARS-CoV-2 Omicron Variant, South Korea, 2021. Emerg Infect Dis **2022**; 28:898-900.
26. Song JS, Lee J, Kim M, et al. Serial Intervals and Household Transmission of SARS-CoV-2 Omicron Variant, South Korea, 2021. Emerg Infect Dis **2022**; 28:756-9.
27. Madewell ZJ, Yang Y, Longini IM, Jr., Halloran ME, Dean NE. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis. JAMA Netw Open **2022**; 5:e229317.

28. Koh WC, Naing L, Chaw L, et al. What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. PLOS ONE **2020**; 15:e0240205.
29. Hsu C-Y, Wang J-T, Huang K-C, Fan AC-H, Yeh Y-P, Chen SL-S. Household transmission but without the community-acquired outbreak of COVID-19 in Taiwan. Journal of the Formosan Medical Association **2021**; 120:S38-S45.
30. Know WYNT. People with certain medical conditions.
31. Nix E, Paulose J, Lakhanpaul M, et al. COVID-19 & informal settlements: is 'Stay Home' safe? UCL Open Environ **2022**; 4:e038.
32. Ou J, Lan W, Wu X, et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. Signal Transduction and Targeted Therapy **2022**; 7:138.
33. Liu H, Wei P, Kappler JW, Marrack P, Zhang G. SARS-CoV-2 variants of concern and variants of interest receptor binding domain mutations and virus infectivity. Frontiers in immunology **2022**; 13:50.
34. Mohandas S, Yadav PD, Sapkal G, et al. Pathogenicity of SARS-CoV-2 Omicron (R346K) variant in Syrian hamsters and its cross-neutralization with different variants of concern. eBioMedicine **2022**; 79.
35. Chow K, Aslam A, McClure T, et al. Risk of Healthcare-Associated Transmission of Sever Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Hospitalized Cancer Patients. Clin Infect Dis **2022**; 74:1579-85.
36. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic Review and Meta-analysis. JAMA Pediatr **2021**; 175:143-56.
37. Cloete J, Kruger A, Masha M, et al. Paediatric hospitalisations due to COVID-19 during the first SARS-CoV-2 omicron (B.1.1.529) variant wave in South Africa: a multicentre observational study. The Lancet Child & Adolescent Health **2022**; 6:294-302.
38. Sheikh A, Kerr S, Woolhouse M, McMenamin J, Robertson C. Severity of omicron variant of concern and effectiveness of vaccine boosters against symptomatic disease in Scotland (EAVE II): a national cohort study with nested test-negative design. Lancet Infect Dis **2022**; 22:959-66.
39. Ferguson N, Ghani A, Hinsley W, Volz E. Report 50: Hospitalisation risk for Omicron cases in England. **2021**.
40. Frederik Plesner L, Kåre M, Kristina Træholt F, et al. Association between SARS-CoV-2 Transmissibility, Viral Load, and Age in Households. medRxiv **2021**:2021.02.28.21252608.
41. Ogata T, Tanaka H, Tanaka E, et al. Increased Secondary Attack Rates among the Household Contacts of Patients with the Omicron Variant of the Coronavirus Disease 2019 in Japan. Int J Environ Res Public Health **2022**; 19.
42. Wei Z, Ma W, Wang Z, et al. Household transmission of SARS-CoV-2 during the Omicron wave in Shanghai, China: A case-ascertained study. Influenza and Other Respiratory Viruses **2023**; 17:e13097.

Figures legends

Figure 1. Study setting. A) Image of the study area, with inset depicting the location of Salvador and Bahia state within Brazil; B) Location of households in the study area with no symptomatic resident (blue dots), no PCR+ resident (gray dots), or at least one PCR+ resident (red dots; and C) yellow dots represent the 14 households with > 1 resident included in the phylogentic analysis

Figure 2. Study period and visits A) weekly new cases of COVID-19 in Salvador, Brazil; B) Number of participants screened and proportion with symptoms C) Number of participants tested classified as contacts and symptomatic index cases; and D) Number of participants in households with >1 PCR+ resident.

* No Omicron variants were detected in November and December 2021. Only two Delta cases were confirmed, these PCR+ results were not included in the SAR analysis.

Figure 3. A) Time-resolved maximum likelihood (ML) phylogenetic tree of the SARS-CoV-2 Omicron BA.1 in Salvador including 62 Omicron BA.1 isolates obtained in this study and an additional 742 representative BA.1 genomes collected throughout the city of Salvador up to March 21st, 2022. Colored circles represent participants from 14 households with > 1 resident included in the analysis and small white circles represent households with a single participant. Branches with no circles represent the genomes collected from GISAID. B) Genomic similarity among groups, C) Proportion of pairs identified at varying genetic similarity thresholds

Figure 4. Genetic similarity network of SARS-CoV-2 isolates among study households. Nodes represent individual SARS-CoV-2 sequences and edge weights represent the dissimilarity values between each pair of sequences. The colored nodes on the plot represent sequences from the Pau da Lima community, which are distributed across six transmission clusters indicated by the color of the nodes. Sequences with labels belong to households with more than one individual included in this analysis. Red labels indicate potential household transmission based on several household members belonging to the same cluster. The nodes without labels represent sequences from households with a single participant included in the analysis. The lines on the plot indicate genomic similarity (the threshold for genomic similarity is set at >2), with thicker lines representing higher degrees of similarity between sequence pairs. Node size represents the value calculated for betweenness centrality, indicating the amount of influence a node has over the flow of information in the graph.[31]

Figure_4

Click here to access/download;Figure;Fig4.tif

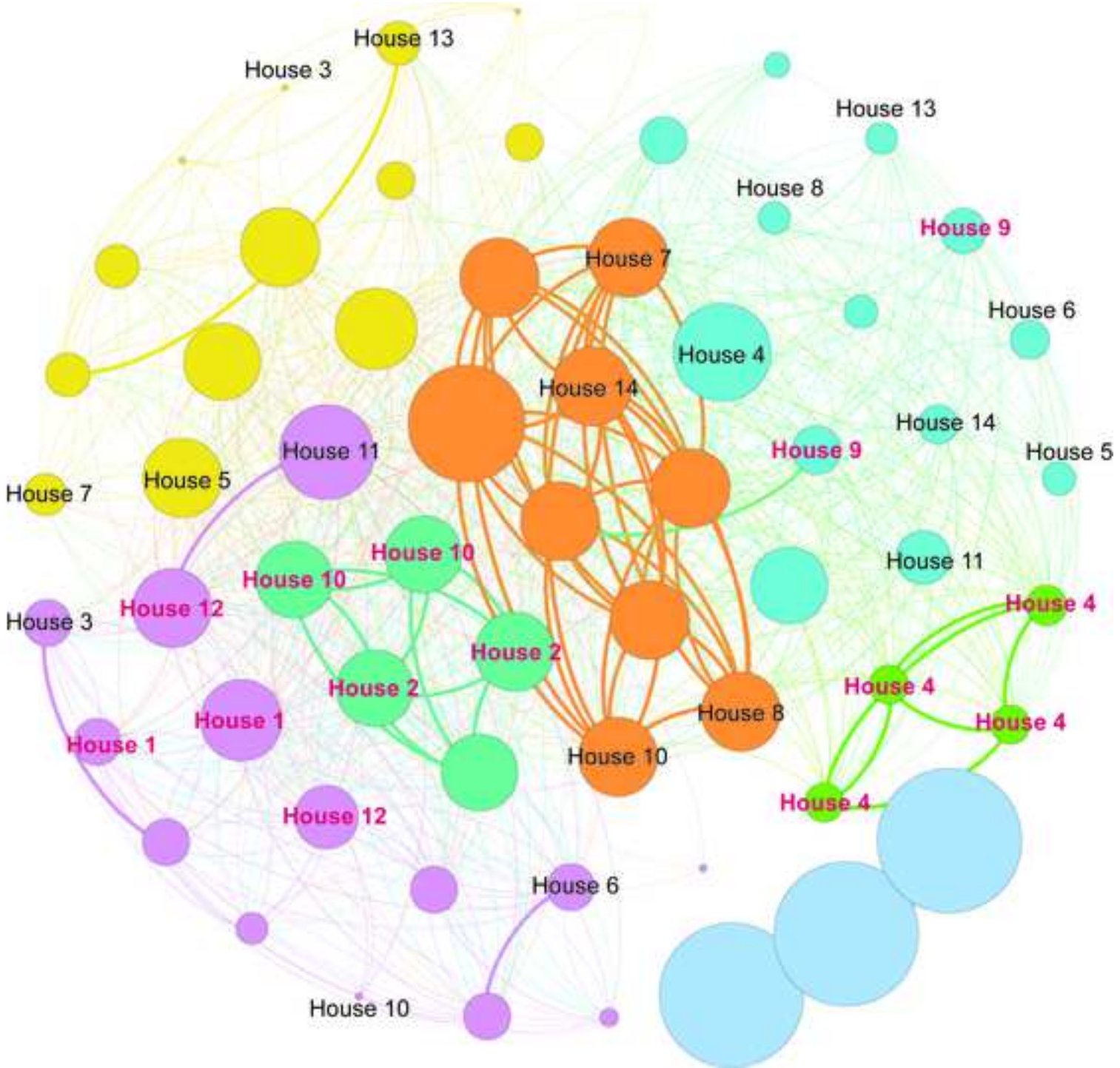


Figure 1. Study setting

Click here to access/download;Figure;Fig 1.tif

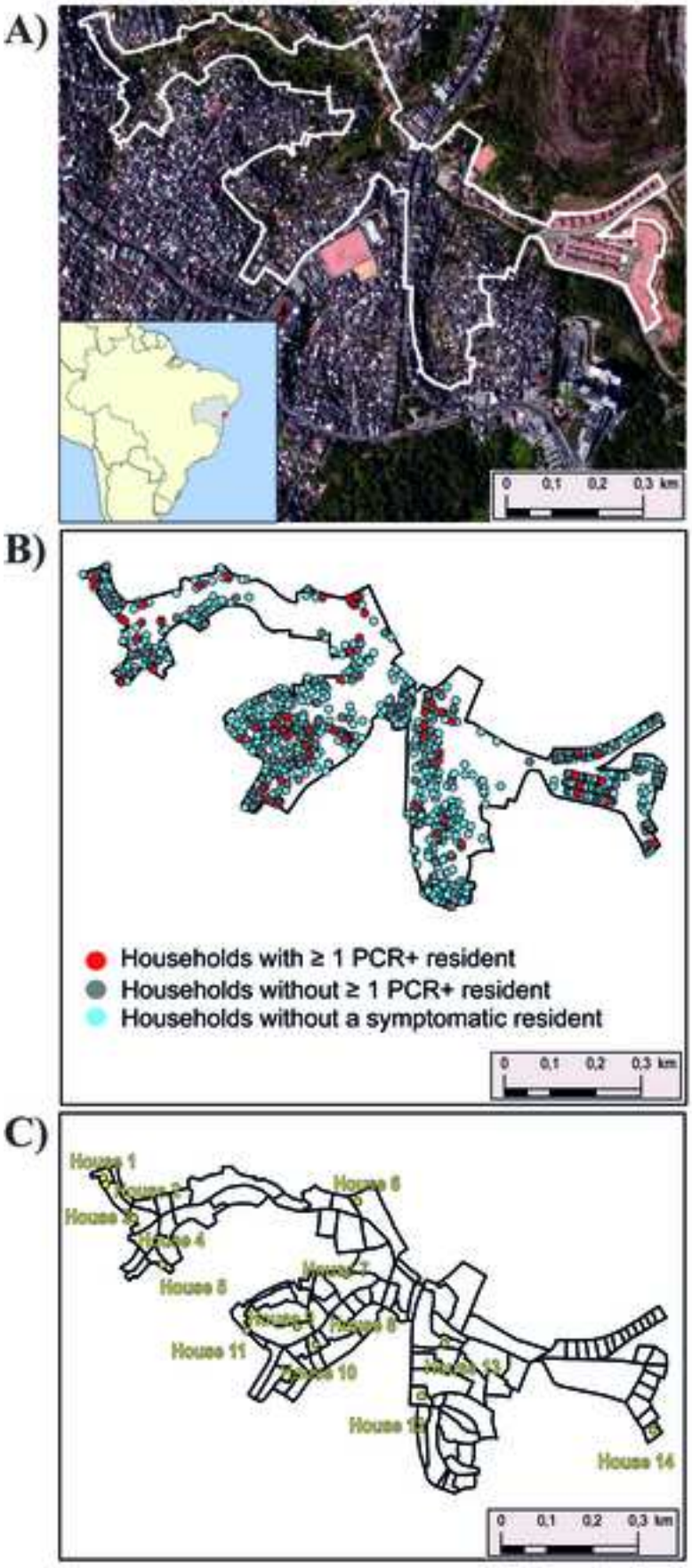


Figure 2. Study period and visits

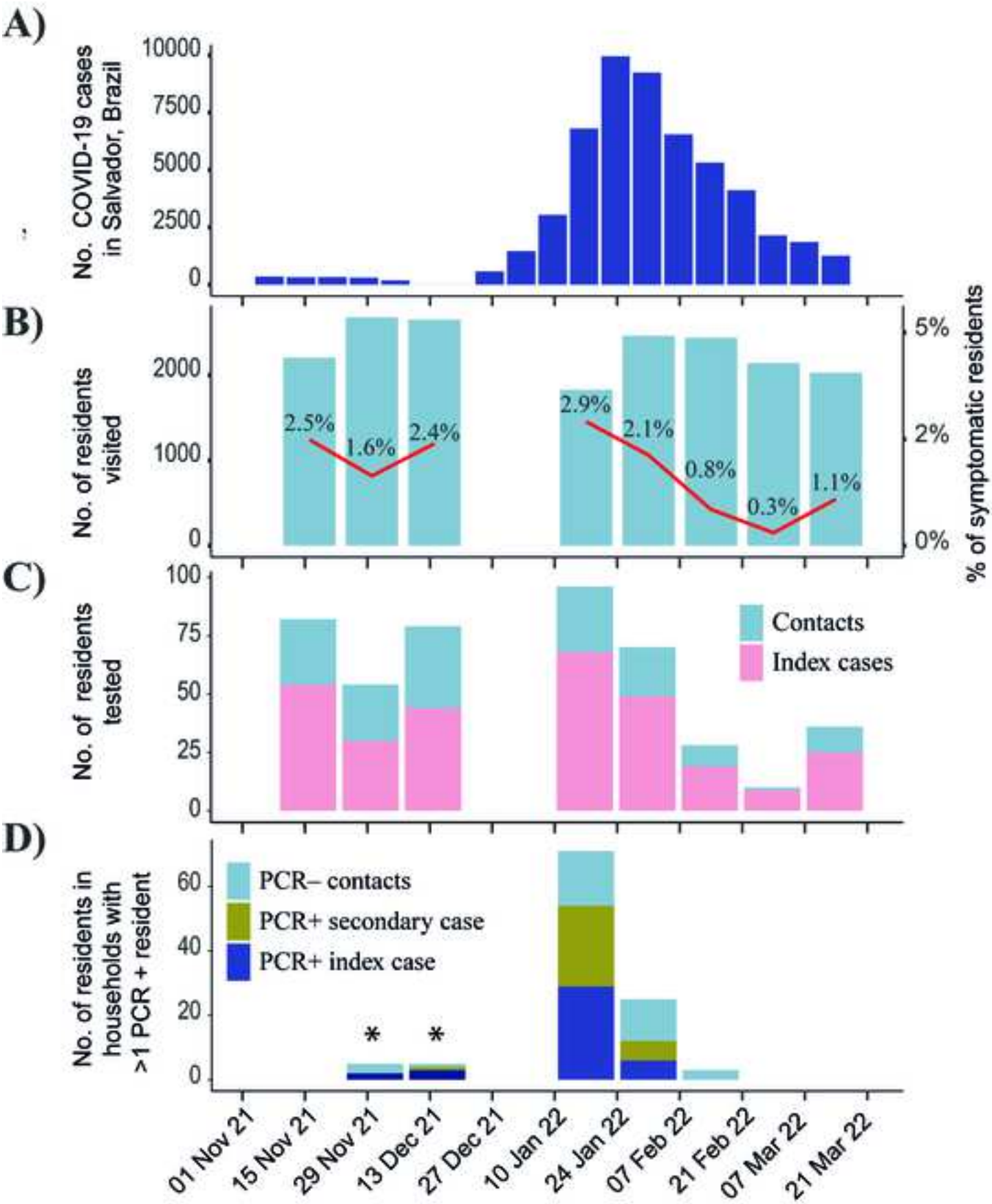


Figure 3. A) Time-resolved maximum likelihood (ML) phylogenetic tree of the SARS-CoV-2 Omicron BA.1 in Salvador including 62 Omicron BA.1 isolates obtained in this

Click here to access/download;Figure;Fig 3.tif

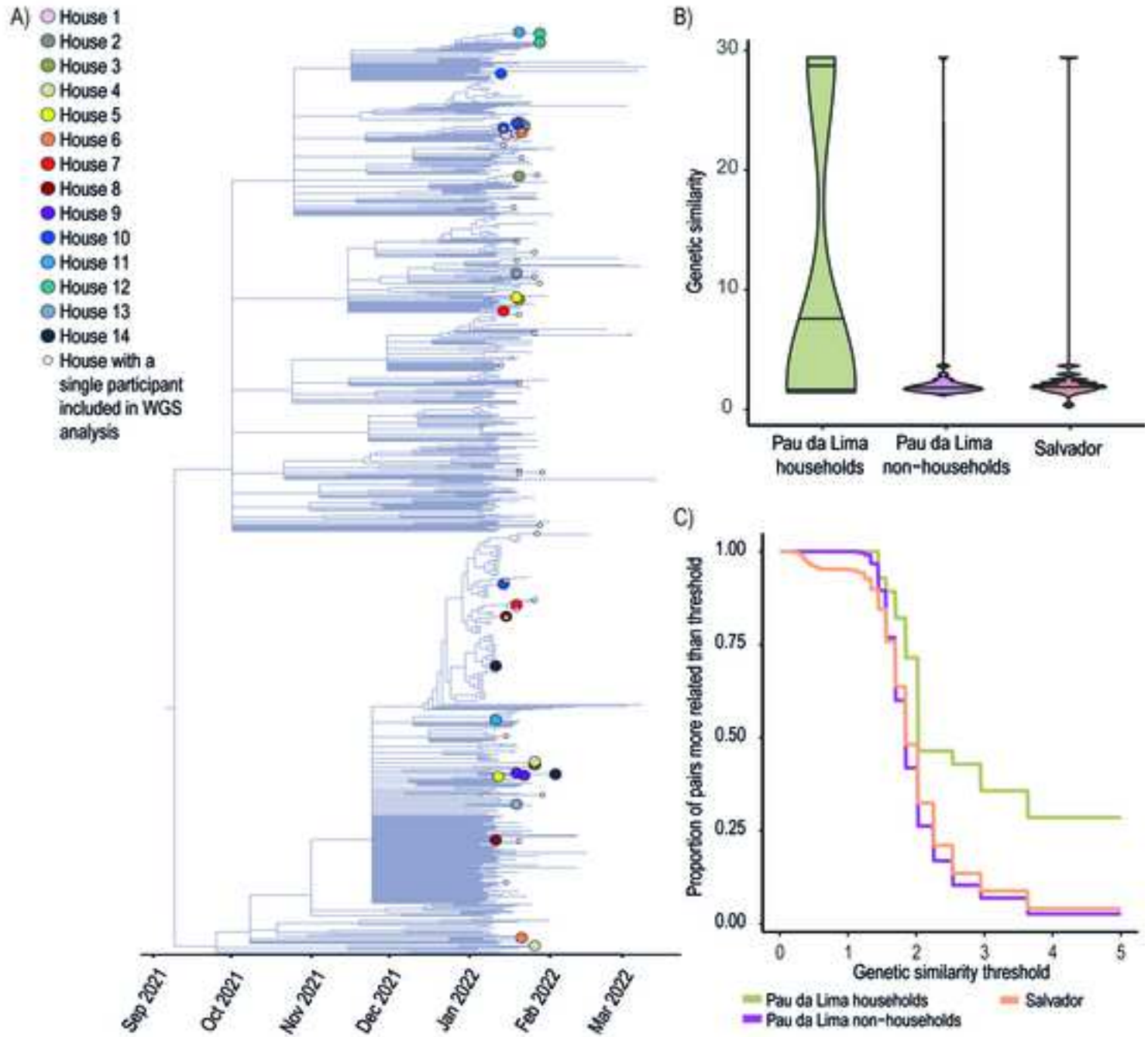


Table 1. Crude household secondary attack rate with Omicron BA.1 variant, by age and sex.

	Secondary cases / Total number of contacts	Secondary attack rate (95% CI)	RR (95% CI)
Crude secondary attack rate	31/62	50.0% (37.0–63.0%)	
Secondary attack rate confirmed by genomic similarity	8/25	32% (13.7 – 50.3%)	
Age groups			
<= 18	8/20	40.0% (21.8 – 61.3%)	ref
19 – 35	6/19	31.6% (15.4 – 54.0%)	0.74 (0.30 – 1.83)
36 – 60	15/20	75.0% (53.1 – 88.8%)	1.82 (1.00 – 3.30)
61+	2/3	66.7% (20.7 – 93.8%)	1.67 (0.64 – 4.37)
Sex			
Female	20/33	60.6% (43.7 – 75.3%)	1.60 (0.93 – 2.74)
Male	11/29	37.9% (22.6 – 56.0%)	ref

Table 2. Risk factors associated with household secondary transmission of Omicron BA.1 variant

Characteristics	SARS-CoV-2 (+) Household contacts (N=31)	SARS-CoV-2 (-) Household contacts (N=31)	RR (95% CI) or median difference (95% CI)
Sex, n (%)			
Female	20 (64.5%)	13 (41.9%)	1.60 (0.93 – 2.74)
Male	11 (35.5%)	18 (58.1%)	Ref
Median age in years, (IQR)	37.0 (20 – 43)	22.0 (15 – 31.0)	15.0 (2.0 – 21.0)
Age groups, n (%)			
≤ 18	8 (25.8%)	12 (38.7%)	Ref
19 – 35	5 (16.1%)	12 (38.7%)	0.74 (0.30 – 1.83)
36 – 60	16 (51.6%)	6 (19.4%)	1.82 (1.00 – 3.30)
≥ 61	2 (6.4%)	1 (3.2%)	1.67 (0.64 – 4.37)
Reported symptoms, n (%)			
Symptomatic	21 (67.7%)	13 (41.9%)	1.73 (1.00 – 3.04)
Asymptomatic	10 (32.3%)	18 (58.1%)	Ref
Vaccination status, n (%)[a]			
Vaccinated	25 (80.6%)	25 (80.6%)	1 (0.53 – 1.88)
Non-vaccinated	6 (19.4%)	6 (19.4%)	Ref
Prior SARS-CoV-2 exposure, n (%)[b]			
Prior exposure	10 (62.5%)	11 (68.8%)	0.63 (0.31 – 1.31)
No prior exposure	3 (18.8%)	1 (6.3%)	Ref
Prior SARS-CoV-2 exposure and vaccination status, n (%)			
Prior exposure and vaccinated	3 (18.8%)	2 (12.5%)	Ref
Prior exposure and unvaccinated	0	2 (12.5%)	0.33 (0.03 – 4.4)
No prior exposure and vaccinated	20 (64.5%)	13 (41.9%)	1.01 (0.47 – 2.17)
No prior exposure and unvaccinated	11 (35.5%)	18 (58.1%)	0.63 (0.27 – 1.49)

[a] Individuals who received at least one COVID-19 vaccine dose

[b] Positive for SARS-CoV-2 anti-S during previous serosurveys studies in the study site (between July 2021 to September 2022). Only 42 individuals participated in previous serosurveys and had a prior exposure documented

Household transmission of the SARS-CoV-2 BA.1* Omicron variant in an urban slum settlement in Salvador, Brazil.

Supplemental Material

Supplemental Methods

RNA Extraction and RT-qPCR

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

SARS-CoV-2 RNA detection was performed by RT-qPCR using the BIOMOL-OneStep/COVID-19 Kit (Instituto de Biologia Molecular do Paraná, ANVISA no. 80780040004), the Molecular SARS-CoV-2 Kit EDx (Bio-Manguinhos, ANVISA no. 80142170045), or the 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 the manufacturer's instructions.

NGS Library Preparation and Sequencing

Libraries were prepared using the COVIDSeq Test (Illumina, Cat. no. 20043675 and 20043137) with the ARTIC V4 or V4.1 primer set as they become available. All libraries were pooled together in equimolar amounts. Fragment length distribution was assessed using the Agilent Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, Cat. no. 5067-4626) on the Agilent 2100 Bioanalyzer (Agilent Technologies, Cat. no. G2939BA). Concentration was assessed using the Qubit 1X dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Cat. no. Q33230 or Q33231) on the Qubit 3 Fluorometer (Thermo Fisher Scientific, Cat. no. Q33216). The library pool was denatured and diluted to a final loading concentration of 8 pM, then loaded into the 300-cycle MiSeq Reagent Kit v2 (Illumina, Cat. no. MS-102-2002) or 600-cycle MiSeq Reagent Kit v3 (Illumina, Cat. no. MS-102-3003). Paired-end sequencing was performed using Illumina MiSeq (Illumina, Cat. no. SY-410-1003) with a 150 bp read length. All protocols followed the manufacturer's instructions.

Genome Assembly

The FASTQ files were processed using the pipeline described by Dezordi et al. (2) with minor modifications. Briefly, reads were trimmed to remove low-quality base pairs and primers using fastp v.0.22.0 (3). Assembly was performed by Burrows-Wheeler Aligner (BWA) v.0.7.17 (4) using NCBI GenBank accession no. MN908947.3 as the genome reference. The consensus sequence was then masked with "N" at regions with coverage depth <10, and variant candidates were incorporated into the consensus genome using iVAR v1.3.1 (5). Assembly statistics were calculated with SAMtools v1.15.1 (using HTSlib v1.15.1) (6) and Seqtk v1.3-r106 (https://github.com/lh3/seqtk). The sequences generated in this study are available via the GISAID Epi Set identifier EPI_SET_230417xo (doi: 10.55876/gis8.230417xo).

Variant Assignment and Mutation Calling

The lineage assignment was conducted using the Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) v4.1.2 (7). Mutation calling was performed by Nextclade v2.5.0 (8). The mutation profile was illustrated using an UpSet plot, produced with R v4.2.2 (9) and the following packages: ggplot2 (10), ComplexHeatmap (11), and UpSetR (12). The plot was further processed using Adobe Illustrator CC 2022 (http://www.adobe.com).

Phylogenetic Analysis

We retrieved data for the SARS-CoV-2 BA.1 Omicron variant from Salvador (Northeast Brazil, Bahia) available in the GISAID database (13) between September 15, 2021, and March 21, 2022. To ensure the quality of the data analyzed in this study, only genomes >29,000pb and with a variant assignment provided by the PANGOLIN (7) were considered (n = 742). The complete set of sequences used in the analysis is available via the GISAID Epi Set identifier EPI_SET_230417ns (doi: 10.55876/gis8.230417ns). Multiple sequence alignment was performed using MAFFT v7.505 with --6merpair and --addfragments (14,15). The alignment was masked with "N" at all problematic sites (16) and manually inspected using AliView v1.28 (17). The maximum likelihood (ML) phylogenetic analyses were performed using IQ-TREE v2.2.0.3 (18) under the transition model 2 (TIM2) of nucleotide substitution with empirical base frequencies (+F) and a proportion of invariant sites, with 1,000 replicates of ultrafast bootstrapping (--B 1000) and SH-aLRT branch test (--alrt 1000) (19). The best-fitting model was chosen according to the Bayesian Information Criterion inferred by ModelFinder (20) implemented in IQ-TREE. The ML tree topology was transformed into a time-scaled tree using TreeTime v0.9.3 (21). Visualizations of the ML time-scaled tree were produced using R v4.2.2 (9) and the following packages: ggtree (22–24), ggplot2 (10), treeio (25), phangorn (26), readxl (27), svglite (28); and further processed using Adobe Illustrator CC 2022 (http://www.adobe.com).

Genetic Distance Analysis and Network Graph Construction

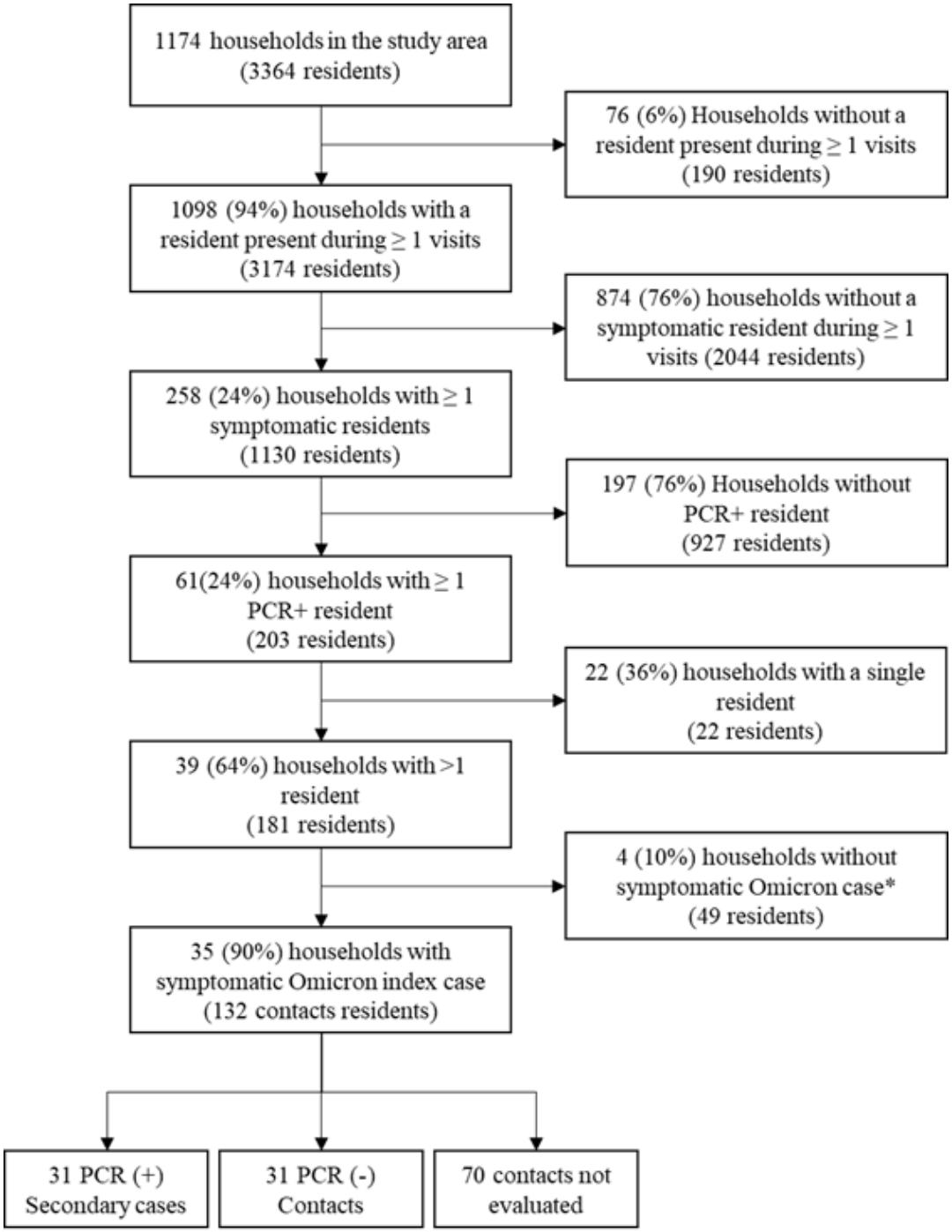
To evaluate transmission dynamics within households and the community, we constructed a distance matrix using the alignment previously described to allow us to investigate the genetic variations and similarities between all sequences under investigation. The distance matrix parameter settings included terminal gaps and penalized gap-letter matches. Then, we convert the distance matrix into a dissimilarity matrix using the exponential negative transformation method as follows: dissimilarity(i, j) = exp(-distance(i, j)). By applying this transformation, we mapped smaller distances to higher dissimilarity values and larger distances to lower dissimilarity values. The matrixes were produced using R v4.2.2 (9) and the following packages: DECIPHER (29) and smacof (30,31).

We constructed a network graph utilizing the Gephi software v0.9.1. This graph was based on the dissimilarity matrix, with nodes representing the SARS-CoV-2 sequences and edge weights representing the dissimilarity values between corresponding sequences. Self-loops were omitted for clarity, and a threshold-based subgraph was generated, incorporating only edges with weights exceeding 2, predicated on the similarity threshold for transmission. The community structure within the network was determined using modularity analysis (32,33), targeting communities comprising three or more households. The modularity parameters were based on Randomization, edge weights, and a resolution of 0.6. From this analysis, seven distinct communities were identified. To effectively visualize and interpret the graph's structure, we employed the Fruchterman-Reingold layout algorithm (34) for optimal positioning of vertices. Node size represents the value calculated for betweenness centrality, representing the amount of influence a node has over the flow of information in a graph.

Reference

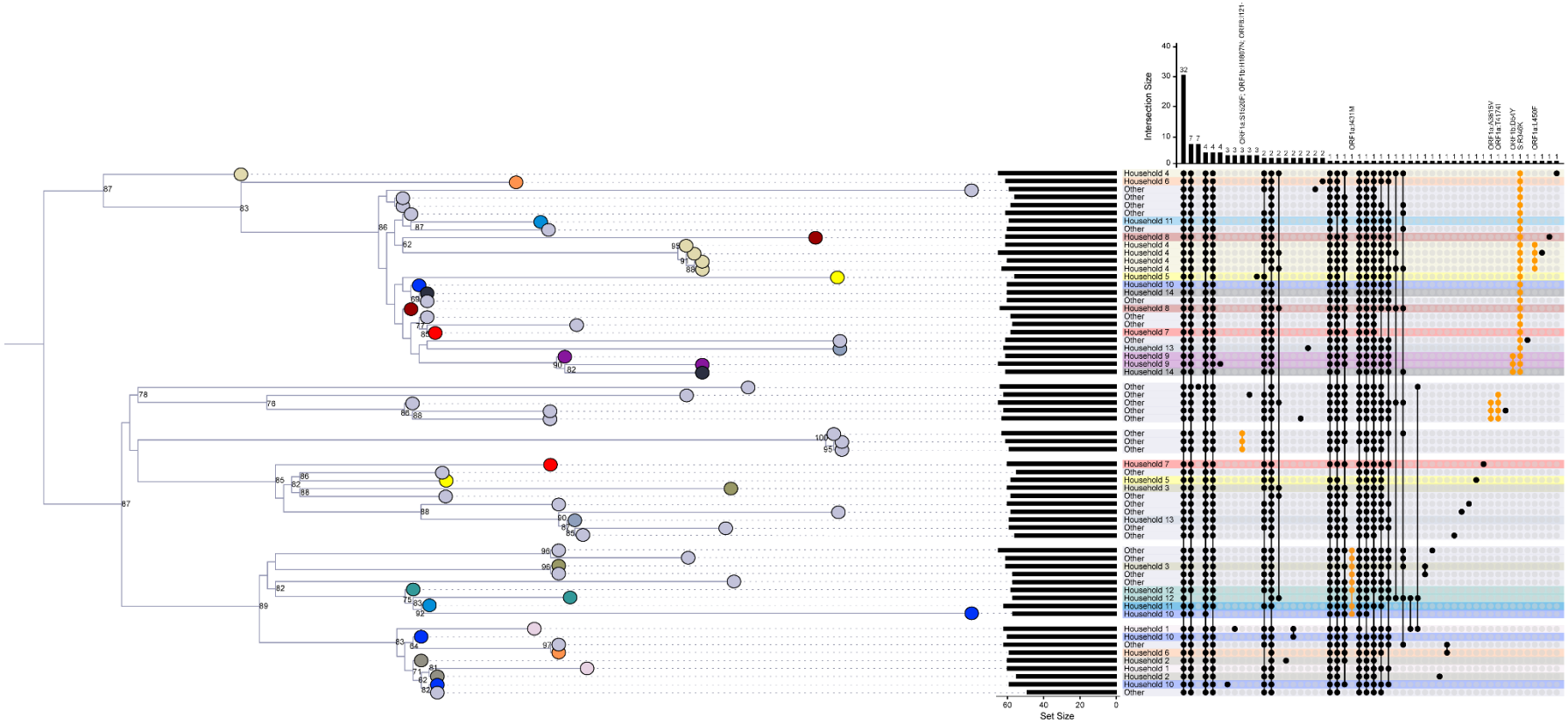
1. CDC C for DC and P. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel [Internet]. Atlanta; 2020. Available from: https://www.fda.gov/media/134922/download
2. Dezordi FZ, Neto AM da S, Campos T de L, Jeronimo PMC, Aksenen CF, Almeida SP, et al. ViralFlow: A Versatile Automated Workflow for SARS-CoV-2 Genome Assembly, Lineage Assignment, Mutations and Intrahost Variant Detection. Viruses. 2022;14(2):217.
3. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
4. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–95.
5. Castellano S, Cestari F, Faglioni G, Tenedini E, Marino M, Artuso L, et al. iVar, an Interpretation-Oriented Tool to Manage the Update and Revision of Variant Annotation and Classification. Genes (Basel). 2021;12(3):384.
6. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Jan 29;10(2):1–4.
7. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the PangoliO'TOOLE, Áine et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. Virus Evolution, [s. l.], v. 7, n. 2, 2021. n Tool. Virus Evol. 2021;7(2).
8. Aksamentov I, Roemer C, Hodcroft E, Neher R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw. 2021;6(67):3773.
9. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. 2016. Available from: https://www.r-project.org
10. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. New York: Springer-Verlag; 2016. 276 p.
11. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.
12. Conway JR, Lex A, Gehlenborg N. UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938–40.
13. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID's Role in Pandemic Response. China CDC Wkly. 2021;3(49):1049–51.
14. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.
15. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.
16. Turakhia Y, De Maio N, Thornlow B, Gozashti L, Lanfear R, Walker CR, et al. Stability of SARS-CoV-2 phylogenies. Barsh GS, editor. PLOS Genet. 2020;16(11):e1009175.
17. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30(22):3276–8.
18. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.
19. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018;35(2):518–22.
20. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
21. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 2018 Jan 1;4(1):1–9.
22. 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 Ecol Evol. 2017;8(1):28–36.
23. Yu G, Lam TTY, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using GGTree. Mol Biol Evol. 2018;35(12):3041–3.

24. Yu G. Using ggtree to Visualize Data on Tree- Like Structures. Curr Protoc Bioinforma. 2020;69(1).
25. Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, et al. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Kumar S, editor. Mol Biol Evol. 2020 Feb 1;37(2):599–603.
26. Schliep KP. phangorn: Phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–3.
27. Wickham H, Bryan J. readxl: Read Excel Files [Internet]. 2023. Available from: https://readxl.tidyverse.org
28. Wickham H, Henry L, Pedersen TL, Luciani TJ, Decorde M, Lise V. svglite: An "SVG" Graphics Device [Internet]. 2023. Available from: https://svglite.r-lib.org
29. Wright, Erik S. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. R J. 2016;8(1):352.
30. Leeuw J de, Mair P. Multidimensional Scaling Using Majorization: SMACOF in R. J Stat Softw. 2009;31(3).
31. Mair P, Groenen PJF, de Leeuw J. More on Multidimensional Scaling and Unfolding in R: smacof Version 2. J Stat Softw. 2022;102(10):1–47.
32. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. 2008 Oct 9;2008(10):P10008.
33. Lambiotte R, Delvenne J-C, Barahona M. Laplacian Dynamics and Multiscale Modular Structure in Networks. IEEE Trans Netw Sci Eng. 2008;1(2):76–90.
34. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. Softw Pract Exp. 1991 Nov;21(11):1129–64.

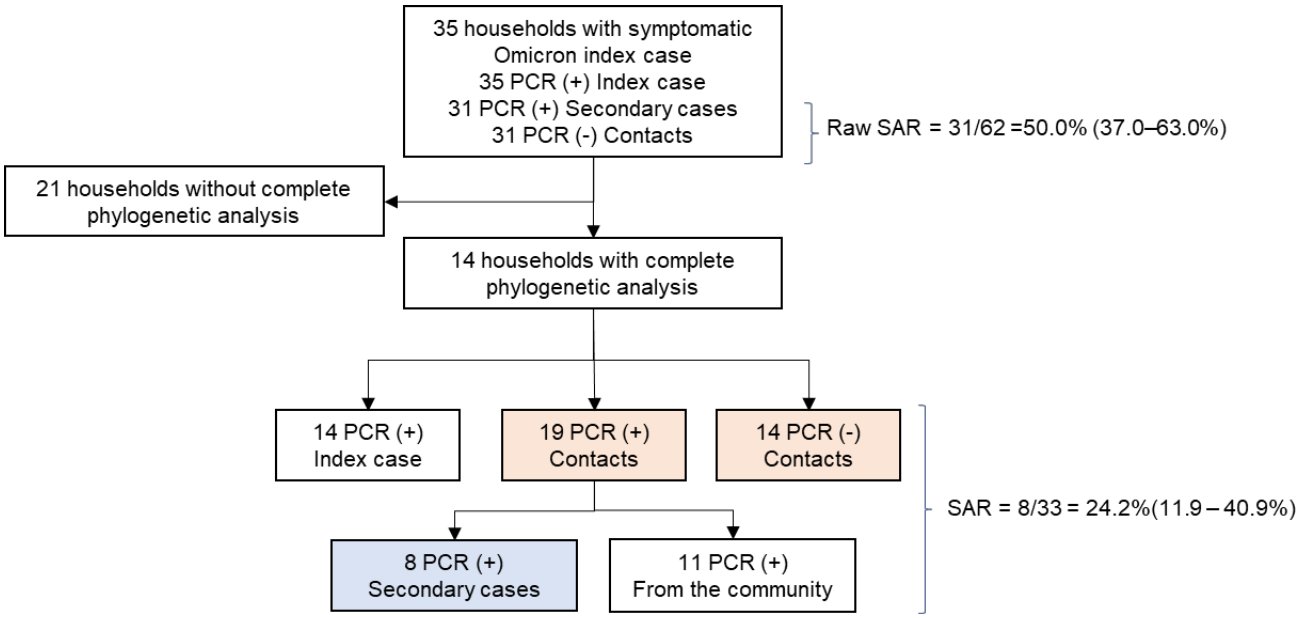


Supplementary Figure 2. Study flowchart.

* Four households included two PCR(+) confirmed Delta variants two PCR(+) confirmed secondary cases and one PCR(+) confirmed that did not have an index case defined.



Supplementary Figure 3. Phylogenetic tree and SNPs among participants from the community of Pau da Lima. Colored circles in the phylogenetic tree (left side) and colored lines in the matrix (right side) represent households with more than one PCR+ resident. Yellow and black dots represent specific mutations identified, with the yellow dot used to highlight specific clusters among the samples.



Supplementary Figure 4. Household secondary attack rate based on genomic similarity analysis.

Supplementary Table 1. Households and residents visited during the COVID-19 active case finding in the Pau da Lima community.

Round	Date	Household (HH) visited (n=1098)		number of residents in the household reported by head of the HH (n=3174)	
1	Nov 10 to Nov 23, 2021	757	68,94%	2204	69,44%
2	Nov 24 to Dic 07, 2021	904	82,33%	2679	84,40%
3	Dic 08 to Dic 21, 2021	938	85,43%	2649	83,46%
break	Dic 22, 2021 to Jan 11, 2022				
4	Jan 11 to Jan 24, 2022	616	56,10%	1828	57,59%
5	Jan 25 to Feb 7, 2022	852	77,60%	2462	77,57%
6	Feb 8 to Feb 21, 2022	831	75,68%	2438	76,81%
7	Feb 22 to Mar 07, 2022	738	67,21%	2142	67,49%
8	Mar 08 to Mar 21, 2022	705	64,21%	2027	63,86%

Supplementary Table 2: COVID-19 Vaccine coverage in the Pau da Lima Cohort until March 21st, 2022

COVID-19 Vaccine coverage in the Pau da Lima Cohort*	n (%)
Vaccination - first dose	
Yes	1193 (80,6)
No	288 (19,4)
Vaccine type	
Pfizer	597 (50)
Coronavac	322 (27)
AstraZeneca - Fiocruz - Oxford	241 (20,2)
Johnson & Johnson	29 (2,4)
NA	4 (0,3)
Vaccination - second dose	
Yes	989 (66,8)
No	492 (33,2)
Vaccine type	
Pfizer	461 (46,6)
Coronavac	273 (27,6)
AstraZeneca - Fiocruz - Oxford	229 (23,2)
Johnson & Johnson	23 (2,3)
NA	3 (0,3)
Vaccination - third dose	
Yes	346 (23,4)
No	1135 (76,6)
Vaccine type	
Pfizer	234 (67,6)
AstraZeneca - Fiocruz - Oxford	79 (22,8)
Coronavac	11 (3,2)
Johnson & Johnson	16 (4,6)
NA	6 (1,7)

* Data based on the survey conducted between October 2022 and March 2023 in the cohort of Pua da Lima

Supplementary Table 3. Household-level factors associated with secondary transmission

	SARS-CoV-2 (+) Household contacts Individual (N=31)	**SARS-CoV-2 (-) Household contacts** Individual (N=31)	p-value
Number of household contacts			0.128
Median [IQR]	4.00 [2.50, 4.50]	3 [2.0 - 4.0]	
Sex of the index case, n (%)			0.290
Female	22 (71.0%)	18 (58.1%)	
Male	9 (29.0%)	13 (41.9%)	
Age groups of the index case, n (%)			0.890
≤ 18	7 (22.6%)	8 (25.8%)	
19 - 35	8 (25.8%)	9 (29.0%)	
36 - 60	15 (48.4%)	11 (35.5%)	
≥ 61	1 (3.2%)	3 (9.7%)	
Index case Ct value			0.072
Median [IQR]	24.3 [21.4 - 26.0]	25.9 [22.6 - 27.7]	
Vaccination status of the index case, n (%)			1
Vaccinated	22 (71.0%)	22 (71.0%)	
Non-vaccinated	9 (29.0%)	9 (29.0%)	

Supplementary Table 4. Comparison of households with ≥ 1 PCR+ resident and households without any PCR+ resident

	Overall (household = 1014 participants = 2964)	**Households with ≥ 1 PCR+ resident (household = 61 participants = 213)**	**Households without 1 PCR+ resident* (household = 953 participants = 2751)**	**p-value**
individual factors				
Sex n (%), n = 2961				
Female	1644 (55.5%)	109 (51.2%)	1535 (55.8%)	0.242
Male	1317 (44.4%)	102 (47.9%)	1215 (44.2%)	
Age n (%), n = 2947				
≤ 18	839 (28.3%)	56 (26.3%)	783 (28.5%)	0.187
19 – 35	963 (32.5%)	57 (26.8%)	906 (32.9%)	
36 – 60	934 (31.5%)	68 (31.9%)	866 (31.5%)	
≥ 61	211 (7.1%)	21 (9.9%)	190 (6.9%)	
Household factors				
Median of No. of residents	3.00 [1.0 - 9.0]	3.00 [1.0 - 7.0]	3.00 [1.0, 9.0]	<0.001
Median of No. of residents < 10 years old	0 [0, 5.0]	0 [0, 2.0]	0 [0, 5.0]	0.118
Median of No. of residents between 10 to 17 years old	0 [0, 4.0]	0 [0, 2.0]	0 [0, 4.0]	0.692

* Data based on the survey conducted between October 2022 and March 2023 in the cohort of Pua da Lima