Genomic epidemiology of SARS-CoV-2 within households in coastal Kenya: a case ascertained cohort study

Authors
Charles N. Agoti[1,2,#], Katherine E. Gallagher[1,3], Joyce Nyiro[1], Arnold W. Lambisia[1], Nickson Murunga[1], Khadija Said Mohammed[1], Leonard Ndwiga[1], John M. Morobe[1], Maureen W. Mburu[1], Edidah M. Ongera[1], Timothy O. Makori[1], My Phan[4,5], Matthew Cotten[4,5], Lynette Isabella Ochola-Oyier[1], Simon Dellicour[6,7], Philip Bejon[1,8], George Githinji[1,2], and D. James Nokes[1,9].

Affiliations
[1]Kenya Medical Research Institute (KEMRI)-Wellcome Trust Research Programme (KWTRP), Kilifi, Kenya.
[2]Pwani University, Kilifi, Kenya.
[3]London School of Hygiene & Tropical Medicine, United Kingdom.
[4]Medical Research Centre (MRC)/ Uganda Virus Research Institute, Entebbe, Uganda.
[5]MRC-University of Glasgow, Centre for Virus Research Glasgow, United Kingdom.
[6]Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Bruxelles, Belgium.
[7]Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory for Clinical and Epidemiological Virology, KU Leuven, University of Leuven, Leuven, Belgium.
[8]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.
[9]School of Life Sciences and Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, United Kingdom.

Correspondence
Dr Charles N. Agoti,
Epidemiology and Demography Department,
KEMRI-Wellcome Trust Research Programme,
P.O. BoX 230, Kilifi – 80108,
Kenya.
Email: cnyaigoti@kemri-wellcome.org

Article type
Research Article

Key words
Household, Transmission, SARS-CoV-2, COVID-19, genomics, Kenya

Abstract

Background

Analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genomic sequence data from household infections should aid its detailed epidemiological understanding. Using viral genomic sequence data, we investigated household SARS-CoV-2 transmission and evolution in coastal Kenya households.

Methods

We conducted a case-ascertained cohort study between December 2020 and February 2022 whereby 573 members of 158 households were prospectively monitored for SARS-CoV-2 infection. Households were invited to participate if a member tested SARS-CoV-2 positive or was a contact of a confirmed case. Follow-up visits collected a nasopharyngeal/oropharyngeal (NP/OP) swab on days 1, 4 and 7 for RT-PCR diagnosis. If any of these were positive, further swabs were collected on days 10, 14, 21 and 28. Positive samples with an RT-PCR cycle threshold of <33.0 were subjected to whole genome sequencing followed by phylogenetic analysis. Ancestral state reconstruction was used to determine if multiple viruses had entered households.

Results

Of 2,091 NP/OP swabs that were collected, 375 (17.9%) tested SARS-CoV-2 positive. Viral genome sequences (>80% coverage) were obtained from 208 (55%) positive samples obtained from 61 study households. These genomes fell within 11 Pango lineages and four variants of concern (Alpha, Beta, Delta and Omicron). We estimated 163 putative transmission events involving members of the sequenced households, 40 (25%) of which were intra-household transmission events while 123 (75%) were infections that likely occurred outside the households. Multiple virus introductions (up-to-5) were observed in 28 (47%) households with the 1-month follow-up period.

Conclusions

We show that a considerable proportion of SARS-CoV-2 infections in coastal Kenya occurred outside the household setting. Multiple virus introductions frequently occurred into households within the same infection wave in contrast to observations from high income settings, where single introduction appears to be the norm. Our findings suggests that control of SARS-CoV-2 transmission by household member isolation may be impractical in this setting.

Introduction

Households are a fundamental unit of social structure and the frequent locale of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission [1,2]. The household secondary attack rate for SARS-CoV-2 has been estimated to be about 21.1% (95%CI: 17.4-24.8%) with considerable heterogeneity observed over geographic regions and time periods [3-6]. Improved understanding SARS-CoV-2 household transmission dynamics including the frequency of virus transmitting from within a household compared to from outside the household, may help refine local control measures. However, to date, such data are limited for sub-Saharan Africa.

SARS-CoV-2 genomic analysis has played a key role throughout the coronavirus disease 2019 (COVID-19) pandemic in elucidating its transmission dynamics [7-11]. Genomic analysis has helped uncover multiple virus introductions into close living environments e.g., hospitals [12,13], prisons[14], cruise ship [15], long-term care facilities [16], and learning institutions [17] and has also uncovered superspreading events [13,18]. It is however still unclear if analysing SARS-CoV-2 genomic data from household clusters can delineate transmission chains [19,20]. Unlike many RNA viruses, SARS-CoV-2 replication is believed to be under some level of proof-reading[21], limiting its substitution rate (9.90×10^{-4} substitutions/site/year; 95% Bayesian credible interval: 6.29×10^{-4} to 1.35×10^{-3}) [22]. A previous genomic analysis of a family infection cluster in Ireland, found only a limited number of mutations between family members testing positive [20].

In the present study, we sought to document SARS-CoV-2 transmission patterns within households in coastal Kenya by analysis of infections identified in a case-ascertained cohort during the local waves of infection. Until September 2022, Kenya had experienced six major waves of SARS-CoV-2 infections[23]. The current study coincided with national waves three, four, and five, a period during which Alpha (B.1.1.7) , Delta (B.1.617.2) and Omicron (B.1.1.529) variants of concern (VOC) predominated, respectively [24]. We undertook detailed genomic analysis to identify independent SARS-CoV-2 introductions into households during clustered infections, and understand frequency of infection spread within households in coastal Kenya.

Methods

Study design and recruitment

We conducted a case-ascertained study in coastal Kenya, where new households were recruited via five local health facilities or County Department of Health rapid response team (RRTs). Households were defined as dwellings or groups of dwellings that share the same kitchen or cooking space. Many of the recruited households were from within the Kilifi Health and Demographic Surveillance System (KHDSS) area located in Kilifi, Coastal Kenya [25]. To get enrolled, a household needed to have at least two occupants be accessible by road and permission obtained from the household head. In the initial study period, only households whose members were contacts of confirmed cases within 2-5 days were recruited and but due to slow enrolment, this was revised to include households with confirmed cases. A household was exempted if at the time of recruitment: two or more members had already developed COVID-19 symptoms (e.g. fever, sore throat, cough etc), a member had been hospitalized due to COVID-19, or the household had been enrolled in a trial of therapeutic COVID-19 product.

Follow-up

During each household visit, a nasopharyngeal and/or oropharyngeal (NP/OP) swab was obtained for real-time RT-PCR testing. The study had two follow-up arms: "reduced follow-up" and "intense follow-up". Households in the "reduced follow-up" arm were those where all the members tested SARS-CoV-2 negative at day 1, 4 and 7; therefore, and follow-up was discontinued henceforth. The "intense follow-up arm" was activated when a household member tested positive on day 1, 4, or 7, and the household was sampled again on day 10, 14, 21 and 28. Data on baseline household and demographic characteristics were collected by the study team at enrolment. During all households' visits, data on presence of acute respiratory illness (ARI) symptoms (e.g., fever, cough, runny nose, sore throat, headache) were collected.

Laboratory procedures

SARS-CoV-2 diagnosis

SARS-CoV-2 testing of study samples was undertaken alongside samples collected in six coastal counties of Kenya as part of the national COVID-19 tests as previously described [26]. Four different viral RNA extraction kits were deployed in combination with five different RT-PCR kits/protocol namely, Da An Gene Co. detection Kit, European Virus Archive-Global (EVAg) E gene protocol, Standard M Kit, Sansure Biotech Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic Real-time RT-PCR kit[26]. Positives were determined using the kit/protocol-defined cycle thresholds (Ct). In kits where multiple SARS-CoV-2 genomic regions were targeted, the average cycle threshold (Ct) was calculated from the individual Cts.

Genome sequencing

We aimed to whole genome sequence all the RT-PCR positive samples with a cycle threshold of < 33.0. Viral RNA was re-extracted from the specimens using QIAamp viral RNA mini-Kit following the manufacturer's instructions and converted to cDNA using Lunascript kit with ARTIC protocol primers [27]. Genome amplification was conducted using Q5 PCR kit and ARTIC protocol primers (initially v3 and then v4). Sequencing

libraries were prepared using Oxford Nanopore Technologies (ONT) ligation sequencing kit SQK-LSK109 and the ONT Native Barcoding Expansion kit as described in the ARTIC protocol [27]. Sequencing was performed on Oxford Nanopore Technologies' MinION or GridION devices using R9.4.1 flow cells.

Bioinformatic analysis

Genome assembly and lineage assignment

The raw sequencing reads (FAST5) were base-called and demultiplexed using ONT's Guppy v3.5-4.2. The resultant files (FASTQ) were assembled into consensus genomes using ARTIC bioinformatic pipeline reference-based approach (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html; last accessed 2022-09-17). Only nucleotides with a read depth of more than × 20 were included into the consensus sequence. High-quality genomes were assigned Pango lineages by using pangolin v4.0.5, PUSHER-v1.3, scorpio v0.3.16 and constellation v0.1.6 [28,29].

Phylogenetic analysis

Multiple sequence alignments were generated using Nextalign v.1.10.1 referenced-based aligner within the Nextclade tool v0.14.2 [30]. Alignments were visualized using a custom Python script and "snipit" tool (https://github.com/aineniamh/snipit; last accessed 2022-05-20). Pairwise distances were calculated using pairsnp.py (https://github.com/gtonkinhill/pairsnp/; last accessed 2022-05-20). Phylogenetic relationships between all recovered genomes and between viruses classified under the same VOC were inferred using maximum likelihood (ML) methods in IQTREE v2.1.3 under the general time reversible (GTR) substitution model. We included contemporaneous genomes from the six coastal Kenya counties (Mombasa, Kilifi, Kwale, Taita Taveta, Tana River and Lamu) that were sequenced as part of the national SARS-CoV-2 genomic surveillance to provide phylogenetic context to the household study genomes. The phylogenetic trees were combined with metadata and visualized with the R package "ggtree" v2.4.2.

Virus introductions

The number of independent virus introductions into the households was inferred using two approaches; (i) comparing observed nucleotide differences between pairs of household genomes with the number of mutations expected over the time interval between the two sampling dates, and (ii) using ancestral state reconstruction (ASR) to count the transitions into a household [10].

Statistical analysis

Summary statistics were computed for key demographic characteristics including mean, median, standard deviation as appropriate. Infection prevalence was expressed using proportions and comparison between groups included appropriate statistical tests (e.g., chi-square or Fisher's exact). All statistical analyses were performed in R packages.

Ethical consideration

The study protocol was reviewed and approved by both the Scientific and Ethics Research Unit (SERU) at Kenya Medical Research Institute (KEMRI), Nairobi, Kenya

(SERU protocol # 4077) and the University of Warwick, Biomedical and Scientific Research Ethics Committee, Coventry, United Kingdom (REF: BSREC 150/19-20 AM01). Prior to data and sample collection, written informed consent was obtained from all participants aged 18 years or older, while for participants aged less than 18 years consent was obtained from their parents or legal guardians. Assent was also sought for adolescents (11-17 years of age).

Role of funding source
The funders of the study had no role in study design, data collection, data analysis , data interpretation or writing of the report.

Results

Baseline characteristics

Of 2,091 nasopharyngeal/oropharyngeal (NP/OP) swabs collected from 573 participants from 158 households between 10th December 2020 and 22nd February 2022, 375 (17.9%) samples tested SARS-CoV-2 positive (**Fig. 1)**). The positives arose from 171 infected participants in 80 households with temporal distribution as shown in **S1 Fig**. The positive cases had a median age of 27 years (IQR: 13.0-46.0; **S1 Table)**, with 104 (60.8%) being females. Compared to participants who remained SARS-CoV-2 negative during the follow-up period; positive cases were more likely to report at least one ARI symptom (63.2% vs 22.6%; $p < 0.001$). The bulk of household recruitments coincided with the national waves 3 and 4 (**Fig. 2A & B**) with only one household recruited during wave 2. The Kenyan government COVID-19 counter-measures during the study period fluctuated as depicted by the Oxford stringency index (**Fig. 2C**) [31].

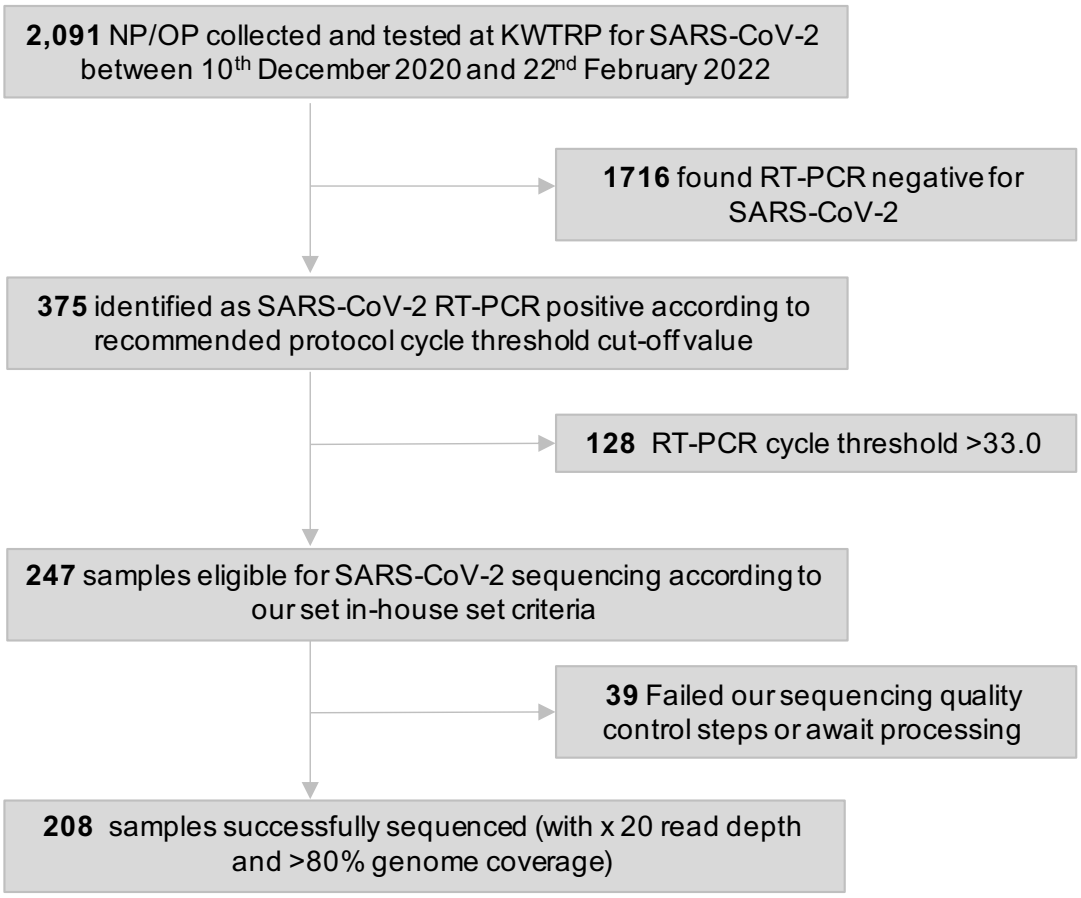

Fig. 1. *Sample flow in the study.*

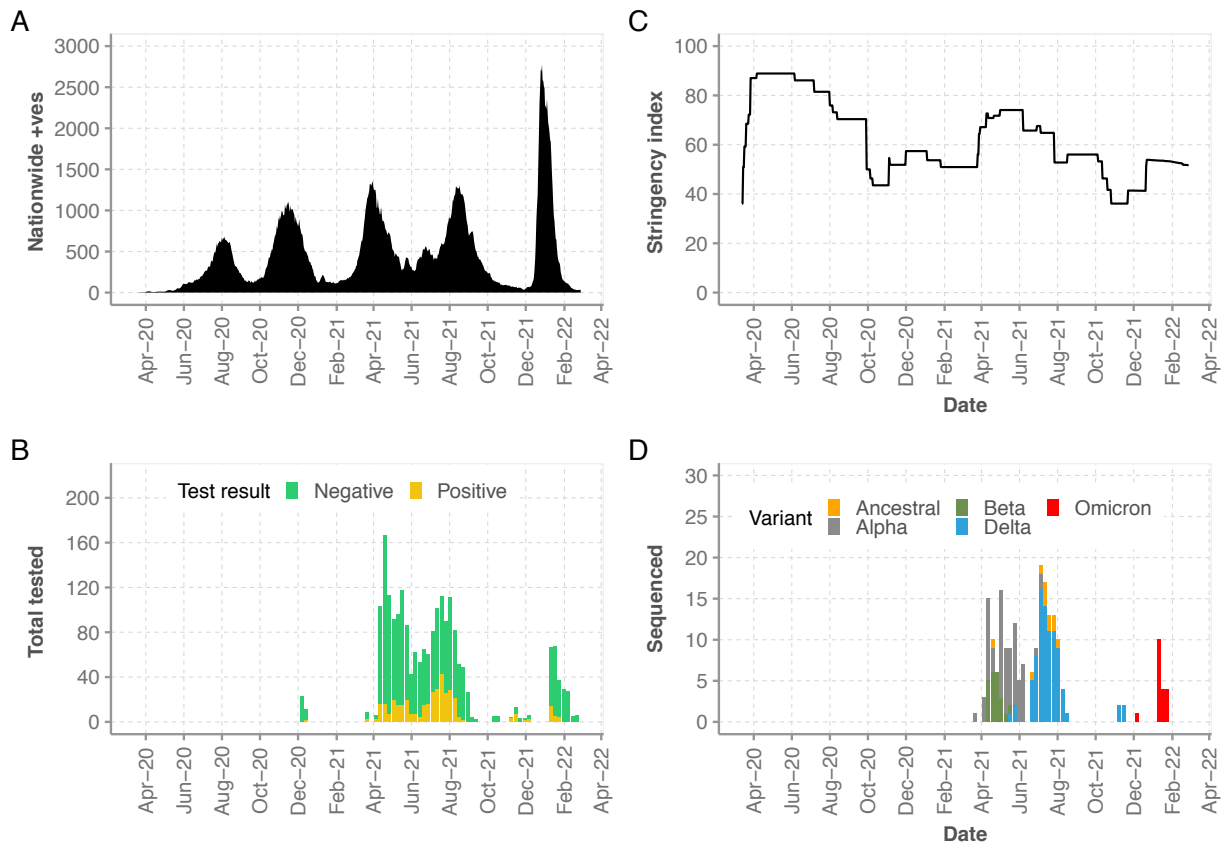

Fig. 2. The timeline of the household study and genomic sequencing results. *Panel A: Reported SARS-CoV-2 infections observed in Kenya between March 2020 and March 2022. Panel B: Temporal distribution of the collected NP/OPs and their RT-PCR diagnostic results. Panel C: The level of government restrictions. Panel D: Temporal distribution of the SARS-CoV-2 VOCs detected in the household study.*

Genomic sequencing and lineage/VOC classification

We recovered near complete genomes (over 80% coverage) from 208 (55.4% of positive samples) from 111 participants from 61 households (**Fig. 2D)**. The samples that failed sequencing (n = 167) had either high Ct values on re-extraction (>33.0; **S2 Fig**.) or yielded poor quality PCR products during library preparation. The recovered genomes were classified into Pango lineage B.1 (n = 11), Alpha variant of concern (VOC; n = 70), Beta VOC (n = 22), Delta VOC (n = 86) and Omicron VOC (n = 19). Within the Delta VOC, five Pango lineages were identified namely, B.1.617.2 (n = 16), AY.16 (n = 5), AY46 (n = 3), AY.116 (n = 58) and AY.122 (n = 4) while within the Omicron VOC three Pango lineages were identified, namely, BA.1.1 (n = 14), BA.1.1.1 (n = 4) and BA.1.9. (n = 1**)**. A summary of the distribution 12 Pango lineages that were identified across the households and sequenced cases and their history is presented in **(S2 Fig and S2 Table**).

Phylogenetic clustering of the household study genomes

To investigate the genetic diversity in the household study genome sequences, we reconstructed a ML phylogeny that included background coastal Kenya co-circulating viruses (n = 2,382**)**. As expected, the genome sequences clustered by VOC and Pango lineages (**S3 Fig)**. Notably, lineage B.1 sequences were found in multiple branches of the phylogeny, including some at the base of branches leading to Beta and Delta VOCs. To assess the genetic relatedness of the recovered genomes within and between various households, we reconstructed VOC-specific phylogenies with tips coloured by the household of sampling (**Fig. 3**). Here we observed both intra- and inter-household clustering. For a few households, tip nodes corresponding to genomic sequences were inferred in distinct clades, already indicating multiple introductions into the same household (**Fig. 3**).

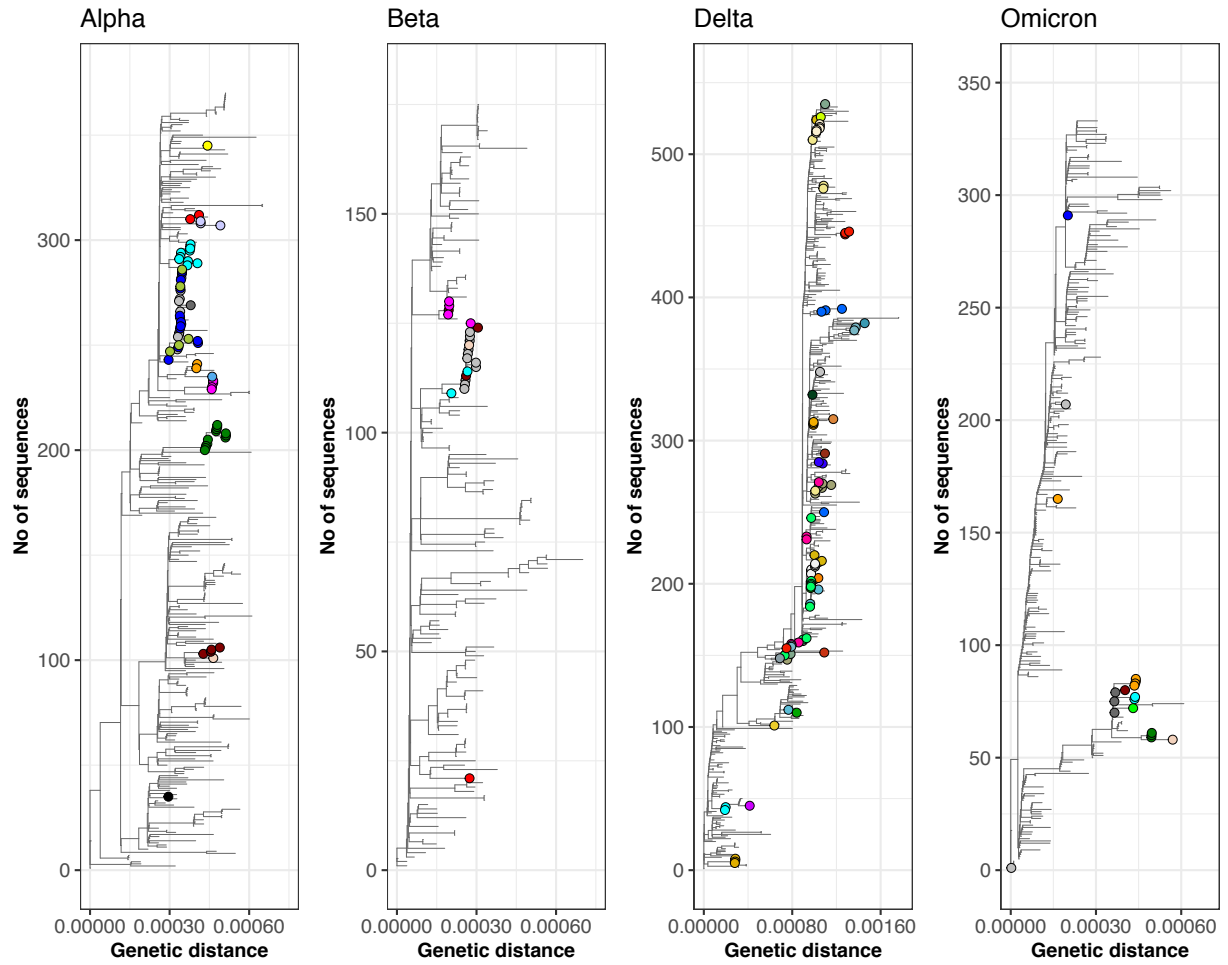

Fig. 3. Phylogenetic patterns of variants of concern (VOC) in the household study using maximum-likelihood methods. *These household study VOC phylogenies include other SARS-CoV-2 genome sequences generated from samples collected in six Kenyan coastal counties (Mombasa, Kwale, Kilifi, Taita Taveta, Tana River and Lamu) during the study period as background diversity (tips without symbols). On the phylogenies, household sample derived sequences are displayed as filled circles, colored distinctly by household. In the Alpha, Beta, Delta, and Omicron phylogenies, 370, 175, 535 and 333 genome sequences were included, respectively.*

Estimating the number of introductions into the households

SARS-CoV-2 has been reported to have an evolutionary rate of ~2 substitutions per genome per month. A heterogenous distribution of the pairwise nucleotide differences of specimens identified in the same household was observed (**S4 Fig.**). More than two nucleotide differences were however seen in 17 households, implying multiple introductions. We investigated the potential number of virus introductions into the households using the ASR approach performed along the dated ML phylogeny . A total of 113 virus introductions were predicted into the 61 households where we recovered sequence data. On classifying the introduction events by origin ("non-household" events - those from populations that are not part of the household study - and "household"

events - those from recruited households) we found that most introductions came from non-household populations (75.2% vs. 24.8%; **Fig. 4A**). Overall, we estimated that a single introduction occurred for 33 households (54%), two introductions for 15 households (25%), three introductions for six households (10%), four introductions for three households (5%), and five introductions for four households (7%) (**Fig. 4B**).

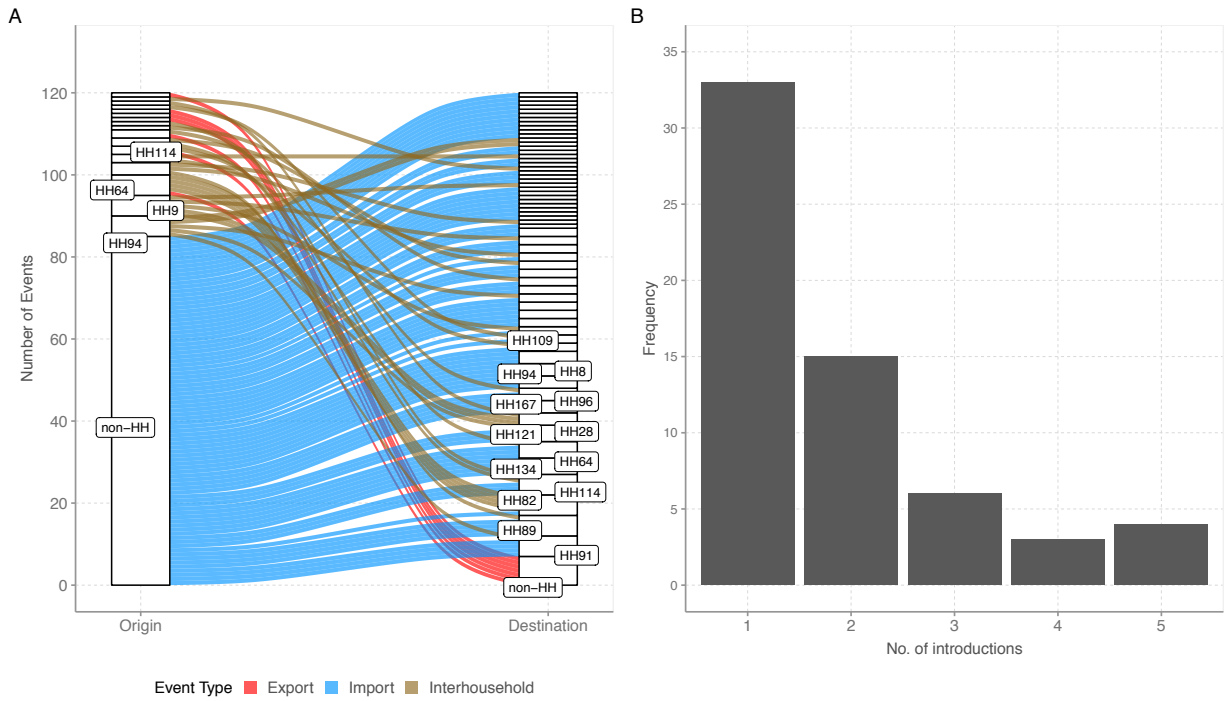

Fig. 4. Patterns of SARS-CoV-2 introductions into study households as determined by ancestral state reconstruction. *Panel A: Alluvial plot showing the number of viral imports and exports into and out of study households. Panel B: Frequency of distinct introductions into the study households.*

Discussion

We provide evidence of frequent multiple SARS-CoV-2 introductions to rural coastal Kenyan households, a finding that was unexpected. The conventional view has been that households with concurrently infected members acquired the infection from one index case. This assumption has repeatedly supported by a number of genomic studies, for instance, by a Dutch study following 85 households, where phylogenetic analysis showed a single introduction into all study households [4]. However in this study, only about half of infected households (54%) had a single introduction.

A variety of factors may explain the differences in household virus introduction patterns in our study from previous observations. First, in our setting, multiple families may live in one compound and eat together in one kitchen. Second, the larger household size increases the chances of multiple viruses being introduced, especially at the height of epidemic waves. Third, the dominance of informal jobs in this setting where effective contacts outside household set-ups might give as much chance to infection transmission as within household.

Our study followed-up participants for a period of up to 1 month with serial sample collection and recovered genomes were analysed in the context of contemporaneous locally circulating diversity in coastal Kenya [32]. Despite observing minimal nucleotide variation between samples from members of the same household infection clusters, when we incorporated sampling dates through the ASR analysis, we were able to partially reconstruct potential within HH transmission events. This allowed identification of virus multiple introductions into the households of closely related viruses, observation of within household transmission and detection of potential short-interval reinfections.

Few studies have examined SARS-CoV-2 households transmission dynamics within Africa [33-35], and these have resulted in diverse findings. In rural Egypt, a 6-month study reported a SAR of 89.8% [33], in South Africa a 13-month study reported a 25% infection rate among vulnerable household contacts [34], in Madagascar, a SAR of 38.8% (CI:19.5-57.2) [36] was reported. None of these studies included genome analysis to confirm that the inferred household transmission clusters were epidemiologically linked within the household.

The Kenyan government countermeasures in place during the study period may have had an impact on the way SARS-COV-2 spread within the study households. In June 2020, the Kenyan government announced guidelines for home-based care for asymptomatic or mildly symptomatic patients without co-morbidities. Kenya started immunizing its population in March 2021, but the coverage was low (<10%) during the study period, and it is unlikely that it affected transmission during our study. The stringency index in the country during the study period fluctuated from 35% to 75%. However, we did not detect variation in the pattern of introductions over time, which could suggest that the various restrictions had minimal impact at the household level. However, concluding on this aspect would likely require more advanced investigations.

This study presents some limitations. First, our sampling interval, especially after week two, may have missed persons who had been positive for less than the 7 days sample collection interval. High density sampling has previously been associated with a higher attack rate[4]. Second, several positive NP/OP samples (44.5%) failed to sequence or had large gaps due to PCR amplicon drop-offs. With this data missingness, overall phylogenetic signal was reduced in trying to establish who infected whom or directionality of transmission. Third, we cannot rule out that a few of the sequence changes could be sequencing or assembly artifacts. Forth, the case-ascertained study design we used had the drawback that by the time of the first sample collection, multiple positive cases had already occurred in households. Most of the index cases were recruited following presentation to a health facility with ARI. This complicated our effort of fully working out who infected whom back in the household. To overcome this challenge, future studies should observe members before entry of the virus into households and genomic data co-analyzed with other relevant epidemiological data [37]. Fifth, intra-patient minority variants may also be examined to provide insights of potential transmission linkages through examination of shared intra-host variation with caveats[38].

In conclusion, our study highlights the importance of examining genomic data for accurate estimation and interpretation of SARS-CoV-2 household epidemiological parameters in these settings. We identified unusually high number of independent virus introductions into households in coastal Kenya during clustered infections. Our findings suggests that control of SARS-CoV-2 transmission by household member isolation alone may not stop community transmission in this setting.

Author's contributions

The project was conceived and designed by CNA, KEG, JUN, MC, and DJN; Laboratory processing of specimens was conducted by JUN, LIO, NM, AWL, KSM, LN, JMM, MWM, EMO, and TOM; Management and analysis of data were handled by AWL, CNA, NM, SD and GG. CNA wrote the first draft; MP, MC, LIO, SD, PB, and DJN critically reviewed the manuscript to produce the final draft.

Data availability

The consensus genome sequences obtained in this study that passed our quality control filters have been submitted to GISAID database (accession numbers available in appendix pages of the supplementary material). The code for the analyses presented in this manuscript is available from the corresponding author upon request. For more detailed information beyond the metadata used in the paper, there is a process of managed access requiring submission of a request form for consideration by our Data Governance Committee (http://kemri-wellcome.org/about-us/#ChildVerticalTab_15).

Acknowledgements

We thank (a) the members of the Kilifi County rapid response team who worked with our field study team in collecting the samples analysed here; (b) the members of the COVID-19 KWTRP Testing Team who undertook real-time RT-PCR processing of the samples received at KWTRP to identify positives (see full list of members below).

This work was supported by the National Institute for Health and Care Research (NIHR) (project reference 17/63/82) using UK aid from the UK Government to support global health research, The UK Foreign, Commonwealth and Development Office and Wellcome Trust (grant# 220985). Members of COVID-19 Testing Team at KWTRP are supported by multiple funding sources including UNITAD (BOHEMIA study received by Dr Marta Maia funded UNITAID), EDCTP (Senior Fellowship and Research and Innovation Action (RIA) grants received by Dr Francis Ndungu), GAVI (PCIVS grant received by Prof. Anthony Scott). Dr Simon Dellicour acknowledges support from the *Fonds National de la Recherche Scientifique* (F.R.S.-FNRS, Belgium; grant n°F.4515.22), from the Research Foundation - Flanders (*Fonds voor Wetenschappelijk Onderzoek-Vlaanderen*, FWO, Belgium; grant n°G098321N), and from the European Union Horizon 2020 project MOOD (grant agreement n°874850).The views expressed in this publication are those of the author (s) and not necessarily those of NIHR, the Department of Health and Social Care, Foreign Commonwealth and Development Office, Wellcome Trust or the UK government.

List of members of the COVID-19 Testing Team at KWTRP

Agnes Mutiso, Alfred Mwanzu, Angela Karani, Bonface M. Gichuki, Boniface Kaaria, Brian Bartilol, Brian Tawa, Calleb Odundo, Caroline Ngetsa, Clement Lewa, Daisy Mugo, David Amadi, David Ireri, Debra Riako, Domtila Kimani, Edwin Machanja, Elijah Gicheru, Elisha Omer, Faith Gambo, Horace Gumba, Isaac Musungu, James Chemweno, Janet Thoya, Jedida Mwacharo, John Gitonga, Johnstone Makale, Justine Getonto, Kelly Ominde, Kelvias Keter, Lydia Nyamako, Margaret Nunah, Martin

Mutunga, Metrine Tendwa, Moses Mosobo, Nelson Ouma, Nicole Achieng, Patience Kiyuka, Perpetual Wanjiku, Peter Mwaura, Rita Warui, Robinson Cheruiyot, Salim Mwarumba, Shaban Mwangi, Shadrack Mutua, Susan Njuguna, Victor Osoti, Wesley Cheruiyot, Wilfred Nyamu, Wilson Gumbi and Yiakon Sein.

References

1 Lee, E. C., Wada, N. I., Grabowski, M. K., Gurley, E. S. & Lessler, J. The engines of SARS-CoV-2 spread. *Science* **370**, 406-407, doi:doi:10.1126/science.abd8755 (2020).

2 Park, Y. J. *et al.* Contact Tracing during Coronavirus Disease Outbreak, South Korea, 2020. *Emerging Infectious Disease journal* **26**, 2465, doi:10.3201/eid2610.201315 (2020).

3 Thompson, H. A. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Setting-specific Transmission Rates: A Systematic Review and Meta-analysis. *Clin Infect Dis* **73**, e754-e764, doi:10.1093/cid/ciab100 (2021).

4 Kolodziej, L. M. *et al.* High SARS-CoV-2 household transmission rates detected by dense saliva sampling. *Clinical Infectious Diseases*, doi:10.1093/cid/ciac261 (2022).

5 Madewell, Z. J., Yang, Y., Longini, I. M., Jr, Halloran, M. E. & Dean, N. E. Household Secondary Attack Rates of SARS-CoV-2 by Variant and Vaccination Status: An Updated Systematic Review and Meta-analysis. *JAMA Network Open* **5**, e229317-e229317, doi:10.1001/jamanetworkopen.2022.9317 (2022).

6 Jørgensen, S. B., Nygård, K., Kacelnik, O. & Telle, K. Secondary Attack Rates for Omicron and Delta Variants of SARS-CoV-2 in Norwegian Households. *JAMA* **327**, 1610-1611, doi:10.1001/jama.2022.3780 (2022).

7 Li, J., Lai, S., Gao, G. F. & Shi, W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **600**, 408-418, doi:10.1038/s41586-021-04188-6 (2021).

8 Bugembe, D. L. *et al.* Main Routes of Entry and Genomic Diversity of SARS-CoV-2, Uganda. *Emerg Infect Dis* **26**, 2411-2415, doi:10.3201/eid2610.202575 (2020).

9 Githinji, G. *et al.* Tracking the introduction and spread of SARS-CoV-2 in coastal Kenya. *Nat Commun* **12**, 4809, doi:10.1038/s41467-021-25137-x (2021).

10 Wilkinson, E. *et al.* A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* **374**, 423-431, doi:10.1126/science.abj4336 (2021).

11 Vöhringer, H. S. *et al.* Genomic reconstruction of the SARS-CoV-2 epidemic in England. *Nature* **600**, 506-511, doi:10.1038/s41586-021-04069-y (2021).

12 Ellingford, J. M. *et al.* Genomic and healthcare dynamics of nosocomial SARS-CoV-2 transmission. *Elife* **10**, doi:10.7554/eLife.65453 (2021).

13 Illingworth, C. J. *et al.* Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections. *Elife* **10**, doi:10.7554/eLife.67308 (2021).

14 Hershow, R. B. *et al.* Rapid Spread of SARS-CoV-2 in a State Prison After Introduction by Newly Transferred Incarcerated Persons - Wisconsin, August 14-October 22, 2020. *MMWR Morb Mortal Wkly Rep* **70**, 478-482, doi:10.15585/mmwr.mm7013a4 (2021).

15 Hoshino, K. *et al.* Transmission dynamics of SARS-CoV-2 on the Diamond Princess uncovered using viral genome sequence analysis. *Gene* **779**, 145496, doi:10.1016/j.gene.2021.145496 (2021).

16 Aggarwal, D. *et al.* The role of viral genomics in understanding COVID-19 outbreaks in long-term care facilities. *Lancet Microbe* **3**, e151-e158, doi:10.1016/s2666-5247(21)00208-1 (2022).

17 Baumgarte, S. *et al.* Investigation of a Limited but Explosive COVID-19 Outbreak in a German Secondary School. *Viruses* **14**, doi:10.3390/v14010087 (2022).

18 Popa, A. *et al.* Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. *Sci Transl Med* **12**, doi:10.1126/scitranslmed.abe2555 (2020).
19 Manuto, L. *et al.* Rapid SARS-CoV-2 Intra-Host and Within-Household Emergence of Novel Haplotypes. *Viruses* **14**, doi:10.3390/v14020399 (2022).
20 Hare, D. *et al.* Genomic epidemiological analysis of SARS-CoV-2 household transmission. *Access Microbiol* **3**, 000252, doi:10.1099/acmi.0.000252 (2021).
21 Smith, E. C., Blanc, H., Surdel, M. C., Vignuzzi, M. & Denison, M. R. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog* **9**, e1003565, doi:10.1371/journal.ppat.1003565 (2013).
22 Nie, Q. *et al.* Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Res* **287**, 198098, doi:10.1016/j.virusres.2020.198098 (2020).
23 Tegally, H. *et al.* The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science*, eabq5358, doi:10.1126/science.abq5358 (2022).
24 Nasimiyu, C. *et al.* Imported SARS-CoV-2 Variants of Concern Drove Spread of Infections across Kenya during the Second Year of the Pandemic. *COVID* **2**, 586-598 (2022).
25 Scott, J. A. *et al.* Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *Int J Epidemiol* **41**, 650-657, doi:dys062 [pii] 10.1093/ije/dys062 (2012).
26 Nyagwange, J. *et al.* Epidemiology of COVID-19 infections on routine polymerase chain reaction (PCR) and serology testing in Coastal Kenya. *Wellcome Open Res* **7**, 69, doi:10.12688/wellcomeopenres.17661.1 (2022).
27 Tyson, J. R. *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*, doi:10.1101/2020.09.04.283077 (2020).
28 O'Toole, Á. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* **7**, veab064, doi:10.1093/ve/veab064 (2021).
29 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).
30 Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121-4123, doi:10.1093/bioinformatics/bty407 (2018).
31 Hale, T. *et al.* A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat Hum Behav* **5**, 529-538, doi:10.1038/s41562-021-01079-8 (2021).
32 Agoti, C. N. *et al.* Transmission networks of SARS-CoV-2 in Coastal Kenya during the first two waves: a retrospective genomic study. *eLife* **11**, e71703, doi:10.7554/eLife.71703 (2022).
33 Gomaa, M. R. *et al.* Incidence, household transmission, and neutralizing antibody seroprevalence of Coronavirus Disease 2019 in Egypt: Results of a community-based cohort. *PLoS Pathog* **17**, e1009413, doi:10.1371/journal.ppat.1009413 (2021).

34 Cohen, C. *et al.* SARS-CoV-2 incidence, transmission, and reinfection in a rural and an urban setting: results of the PHIRST-C cohort study, South Africa, 2020-21. *Lancet Infect Dis*, doi:10.1016/s1473-3099(22)00069-x (2022).
35 Semakula, M. *et al.* The secondary transmission pattern of COVID-19 based on contact tracing in Rwanda. *BMJ Global Health* **6**, e004885, doi:10.1136/bmjgh-2020-004885 (2021).
36 Ratovoson, R. *et al.* Household transmission of COVID-19 among the earliest cases in Antananarivo, Madagascar. *Influenza Other Respir Viruses* **16**, 48-55, doi:10.1111/irv.12896 (2022).
37 Agoti, C. N. *et al.* Genomic analysis of respiratory syncytial virus infections in households and utility in inferring who infects the infant. *Sci Rep* **9**, 10076, doi:10.1038/s41598-019-46509-w (2019).
38 Gallego-García, P. *et al.* Limited genomic reconstruction of SARS-CoV-2 transmission history within local epidemiological clusters. *Virus Evolution* **8**, doi:10.1093/ve/veac008 (2022).