Rapid transmission and tight bottlenecks constrain the evolution of highly transmissible SARS-CoV-2 variants

Emily E. Bendall¹, Amy Callear², Amy Getz², Kendra Goforth², Drew Edwards², Arnold S. Monto²,
Emily T. Martin², Adam S. Lauring¹ ³ *

¹Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA
²Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA
³Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

* Corresponding Author:

Adam Lauring

MS2 4742C, SPC 1621
1137 Catherine Street, Ann Arbor, MI 48109
alauring@med.umich.edu

Keywords: SARS-CoV-2; transmission; bottleneck; genomic epidemiology

Running Title: SARS-CoV-2 diversity in households

Abstract Word Count: 240

Main Text Word Count: 3410
Abstract

Transmission bottlenecks limit the spread of novel mutations and reduce the efficiency of natural selection along a transmission chain. Many viruses exhibit tight bottlenecks, and studies of early SARS-CoV-2 lineages identified a bottleneck of 1-3 infectious virions. While increased force of infection, host receptor binding, or immune evasion may influence bottleneck size, the relationship between transmissibility and the transmission bottleneck is unclear. Here, we compare the transmission bottleneck of non-variant-of-concern (non-VOC) SARS-CoV-2 lineages to those of the Alpha, Delta, and Omicron variants. We sequenced viruses from 168 individuals in 65 multiply infected households in duplicate to high depth of coverage. In 110 specimens collected close to the time of transmission, within-host diversity was extremely low. At a 2% frequency threshold, 51% had no intrahost single nucleotide variants (iSNV), and 42% had 1-2 iSNV. In 64 possible transmission pairs with detectable iSNV, we identified a bottleneck of 1 infectious virion (95% CI 1-1) for Alpha, Delta, and Omicron lineages and 2 (95% CI 2-2) in non-VOC lineages. The latter was driven by a single iSNV shared in one non-VOC household. The tight transmission bottleneck in SARS-CoV-2 is due to low genetic diversity at the time of transmission, a relationship that may be more pronounced in rapidly transmissible variants. The tight bottlenecks identified here will limit the development of highly mutated VOC in typical transmission chains, adding to the evidence that selection over prolonged infections in immunocompromised patients may drive their evolution.
Introduction

Viral populations are often subject to multiple bottleneck events as they evolve within and between hosts. These bottlenecks drastically reduce the size and genetic diversity of the population, which will affect how new mutations spread through host populations (1, 2). In the setting of a tight transmission bottleneck, most mutations that arise within a host are not propagated between them. Bottlenecks also reduce the virus’s effective population size, which captures the number of virions that reproduce and genetically contribute to the next generation; selection is less effective in smaller populations. Therefore, tight bottlenecks constrain adaptive evolution by limiting the spread of newly arising mutations and reducing the efficiency of selection on these mutations along transmission chains. Many viruses, such as HIV (3, 4), influenza (5), and SARS-CoV-2 (6–10), have tight bottlenecks, with 1-3 distinct viral genomes transmitted.

The size of the transmission bottleneck may be impacted by viral dynamics, route of infection, or molecular interactions at the virus-host interface. For example, it has been suggested that transmissibility, or force of infection, may influence bottleneck size. Increased transmissibility may lead to wider bottlenecks in several ways. First, increasing the infectious dose, perhaps through increased shedding in the donor host or increased intensity of contact, can lead to wider bottlenecks as shown in experimental infections of influenza A virus (11, 12) and tobacco etch virus (13). Additionally, the number of virions that initially infect cells is directly related to bottleneck size (14). More transmissible viruses may have an increased ability to infect
individual cells, such as through increased receptor affinity or escape from intrinsic or innate immunity.

While early studies of SARS-CoV-2 transmission estimated a tight transmission bottleneck, the last 20 months of the pandemic have witnessed the emergence of highly transmissible variants of concern (VOC). In December 2020, B.1.1.7 (Alpha) was detected for the first time with a substantial increase in transmissibility over previous SARs-CoV-2 lineages (15). Since then, additional variants of concern characterized by an increase in transmissibility have arisen. The Alpha, Beta, Gamma, Delta, and Omicron VOC are 25-100% more transmissible than the original Wuhan strain (16). There are multiple and overlapping mechanisms for the increased transmissibility in SARS-CoV-2 that may influence bottleneck size, including increased binding to ACE2 (17–20), increased viral shedding (21, 22), innate immune evasion (23), rapid cellular penetration (18), and alternative entry pathways (24, 25).

Here we explore the relationship between viral transmissibility and transmission bottlenecks by comparing bottleneck size across multiple VOC and pre-VOC lineages. We sampled viral populations from two household cohorts in Michigan, obtaining high depth of coverage sequence from 168 individuals in 65 households. We found that bottleneck size did not vary significantly between transmission pairs infected with pre-VOC lineages and those infected with highly transmissible Alpha, Delta, or Omicron (BA.1) lineages. This tight bottleneck was largely the result of limited diversity in the donor host at the time of transmission.
Methods

Households and sample collection

Households were enrolled through two household cohorts in Southeast Michigan – MHome and the Household Influenza Vaccine Evaluation Study (HIVE). MHome is a case ascertained household cohort in which households are recruited following identification of an index case who meets a case definition for COVID-like illness and is positive for SARS-CoV-2 by clinical testing. Households in this study were enrolled between November 18, 2020 and January 19, 2022. HIVE is a prospective household cohort with year-round surveillance for symptomatic acute respiratory illness. We identified all HIVE households with ≥1 individuals positive for SARS-CoV-2 between June 1, 2021 and January 18, 2022. For both studies, written informed consent (paper or electronic) was obtained from adults (aged >18). Parents or legal guardians of minor children provided written informed consent on behalf of their children. Both study protocols were reviewed and approved by the University of Michigan Institutional Review Board (HIVE: HUM118900 & HUM198212, MHome: HUM180896).

In MHome, index enrollees meeting the case definition (at least one the following: cough, difficulty breathing, or shortness of breath; or at least two of the following: fever, chills, rigors, myalgia, headache, sore throat, new loss of smell or taste) with a positive clinical test result within the last 7 days are invited to enroll themselves and their household members. Nasal swabs were collected on days 0, 5, and 10 after enrollment for all participating household members. For HIVE, study participants were instructed to collect a nasal swab at the onset of illness, with weekly active confirmation of illness status by study staff. Eligible illness was
defined as two or more of cough, nasal congestion, sore throat, chills, fever/feverish, body aches, or headache (for participants 3 years & older) or two or more of cough, runny nose/nasal congestion, fever/feverish, fussiness/irritability, decreased appetite, trouble breathing, or fatigue (for participants under 3 years old). If a participant had symptoms of a respiratory illness, specimens were collected from all members of that household on days 0, 5, and 10 of the index illness. For both cohorts all samples were nasal swabs that were self-collected, or in the case of young children, parent-collected following an established protocol (26). In both cohorts, participants were questioned about the day of symptom onset and duration of symptoms. In MHome, the index case was defined as the individual with the earliest symptom onset date. If two or more individuals shared the earliest onset date, they were considered to be co-index cases.

**Viral sequencing**

All samples were tested by quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) with either the TaqPath COVID-19 Combo Kit from Thermofisher (MHome) or CDC Influenza SARS-CoV-2 Multiplex Assay (HIVE). We sequenced the first positive sample in each individual with a cycle threshold (Ct) value ≤30 from each individual. RNA was extracted using the MagMAX viral/pathogen nucleic acid purification kit (ThermoFisher) and a KingFisher Flex instrument. Sequencing libraries were prepared using the NEBNext ARTIC SARS-CoV-2 Library Prep Kit (NEB) and ARTIC V3 (MHome, through November 10, 2021) and V4 (MHome, after November 10, 2021; HIVE) primer sets. After barcoding, libraries were pooled in equal volume. The pooled libraries (up to 96 samples per pool) were size selected by gel extraction and
sequenced on an illumina MiSeq (2x250, v2 chemistry). We sequenced all samples in duplicate from the RNA extraction step onwards, randomizing sample position on the plate between replicates.

We aligned the sequencing reads to the MN908947.3 reference using BWA-mem v 0.7.15 (27). Primers were trimmed and consensus sequences were generated using iVar v1.2.1 (28). Intrahost single nucleotide variants (iSNV) were identified for each replicate separately using iVar (28) with the following criteria: average genome wide coverage >500x, frequency 0.02-0.98, p-value <1x10^{-5}, variant position coverage depth > 400x. We also masked ambiguous and homoplastic sites (29). Finally, to minimize the possibility of false variants being detected, the variants had to be present in both sequencing replicates. Indels were not evaluated.

Delineation of transmission chains and SARS-CoV-2 lineages

Alignments of consensus sequences within each household were manually inspected. We considered infections to be consistent with household transmission if the consensus sequences differed by ≤2 mutations (30). We excluded individuals whose consensus sequences were inconsistent with household transmission but retained the rest of the household if there was evidence of household transmission among the other members. Households were split and analyzed separately if the consensus sequences supported multiple independent transmission chains within the household. If necessary, we reassigned the index case, so that the index case was part of the transmission chain.
For households with genetically linked infections, we further analyzed all samples with high quality sequencing (>500x coverage) from households with ≥2 members. We used Nextclade to annotate clades and variants of concern (31). We used the WHO definition to classify variants of concern (i.e., Alpha, Beta, Gamma, Delta, and Omicron: BA1)(32). Variants of interest were included in the non-variants of concern group for all analyses.

**Infection dynamics**

Serial intervals were calculated as the time between symptom onset of the index and each household contact and compared across clades using an ANOVA. Additionally, the times between symptom onset and sample collection for index cases were calculated. Serial intervals and time to sampling across clades were compared using an ANOVA followed by a Tukey HSD. We also compared the Ct values from the nucleocapsid gene of sequenced samples and the other positive non-sequenced samples for index cases.

**Bottleneck estimation**

We defined the possible transmission pairs within each household as follows: the index was allowed to be the donor for household contacts, and the household contacts were allowed to be donors to each other. The only case in which the index case was allowed to be the recipient was when there were co-index cases. Co-index cases were allowed to be both donor and recipient with respect to the other co-index. After defining the transmission pairs, we applied the approximate beta-binomial approach (33). This method accounts for the variant calling frequency threshold and stochasticity in the recipient after transmission. We estimated the
bottleneck size for each transmission pair individually and also calculated an overall bottleneck size for each clade using a weighted sum of loglikelihoods (33). We re-calculated the above bottleneck estimates after merging replicate aligned fastq files to examine the impact of our variant calling strategy.

Data and materials availability

Raw sequencing reads are available on the NCBI short read archive under BioProject PRJNA889424. Data and scripts necessary to replicate the analyses are available on github (https://github.com/lauringlab/SARS-CoV-2_VOC_transmission_bottleneck).

Results

We used high depth of coverage sequencing to characterize SARS-CoV-2 populations collected from individuals enrolled in a prospective surveillance cohort (HIVE) and a case-ascertained household cohort (MHome). There were 65 multiply infected households (infections ≤14 days apart) with 168 cases. High quality, whole genome sequences (see Methods) were obtained with technical replicates from 131 cases. Depth of coverage was generally high and iSNV frequency was similar across both replicates (Figure S1). There were five households that had consensus sequences inconsistent with household transmission (Figure S2). Of these five, two households with two individuals each were excluded. In two households, there was a single individual whose consensus sequence differed from the others and was excluded. In the final household, the consensus sequences were consistent with two separate transmission pairs, and these were analyzed separately. All 5 households with multiple introductions were due to
either Delta or Omicron viruses, consistent with high community prevalence during these waves (34). The final transmission analysis dataset included 45 households, 110 individuals, and 134 possible transmission pairs (Table 1). Alpha (B.1.1.7), Gamma (P.1), Delta (AY.3, AY.4, AY.39, AY.44, AY.100), and Omicron (BA.1, BA.1.1) were represented in these households. Variants of interest included one household with Lambda (C.37).

There was rapid transmission of SARS-CoV-2 in the sampled households. The median serial interval ranged between 2 and 3.5 with no significant difference observed between clades (df =4, F =.879, p =0.483, Figure 1A, Figure S3). Households with Delta and Omicron had a greater range of serial intervals. Viral specimens were collected soon after symptom onset in both household studies, with a clade-specific medians ranging from 2-6.5 days. Omicron had a shorter time between index symptom onset and sample collection for sequencing than non-VOC (df=3, F=8.138, p <0.001) and Alpha (p =0.01) (Figure 1B, S3). This is likely due to the number of Omicron cases in HIVE households, which had a shorter time between index symptom onset and sample collection for sequencing than MHome households (df =1, F =15.363, p < 0.001).

We further examined the timing of index case sampling by trending RT-qPCR Ct values for all index case specimens. In nearly all cases, the index cases were sampled at peak viral shedding (Figure 1C). Therefore, our sequence data for the index cases should be reflective of the genetic diversity present in donor hosts when risk of household transmission was highest. Consistent with the short time between the infection onset and sample collection, we found low genetic
diversity in nearly all specimens. (Figure 2A). A majority (56/110, 51%) had no iSNV above the 2% frequency threshold; 42% (46/110) of samples had 1-2 iSNV; and 7% (8/110) had ≥ 3 iSNV. There were no specimens with more than 5 iSNV. Fifty-two percent of iSNV were present at <10% frequency within hosts, Figure 2B).

Bottleneck size is calculated based on shared diversity between members of a transmission pair. Within each household, possible transmission pairs included the index case as donor and each household contact as a recipient, and household contacts as donors for other household contact recipients. While the majority of sampled households had only two cases, 12 had three cases, and 4 had four cases (Figure 3A). The number of possible transmission pairs per household ranged from 1 to 12 (Table S1). When we compared the frequency of iSNV in the donors and recipients, we found only a single shared iSNV – C29708T (noncoding) – in 6 possible transmission pairs from a single household (Figure 3B). This iSNV was present in all three individuals in the household at a frequency of 0.56, 0.97, and 0.24 respectively. All other iSNV were either absent (frequency of 0) or completely fixed (frequency of 1) in the other individual of the transmission pair for all households. This pattern is highly suggestive of a narrow bottleneck.

We used the beta binomial model (33) to obtain a quantitative estimate of the transmission bottleneck. Because bottleneck size can only be calculated when there are iSNV in the transmission donor (see Figure 2A), we were able to use 64 potential pairs in this analysis (Table S1). All VOC clades had an overall bottleneck size of 1 (Alpha, Delta, Omicron: 95% CI 1:1,
Gamma: 95% CI 1:7). The Non-VOC clades had an overall bottleneck size of 2 (95% CI 2:2),
which was driven entirely by the single shared iSNV in one household. The 6 transmission pairs
in this household exhibited bottlenecks of 2, 4, and 6 (Table S2). All other transmission pairs
had a bottleneck size of 1 inclusive of all clades. Across all transmission pairs, the upper bound
of the 95% confidence interval varied greatly, from 1 to 200, the maximum bottleneck size we
evaluated (Table S2).

We were stringent in our variant calling criteria and required iSNV to be present in both
sequencing replicates, because false positive iSNV can artifactually inflate bottleneck estimates
(7, 35–37). To ensure that our stringency did not lead to an underestimate, we re-analyzed our
dataset after merging sequencing reads across the technical replicates. This had only a small
effect on the number of iSNV identified in each specimen (Figure S4). Thirty-nine out of 110
specimens still had no iSNV present, and all but 2 specimens had ≤8 iSNV. The remaining two
specimens had 25 and 57 iSNV. The newly detected iSNV in the merged dataset tended to be
present at very low frequency (<3%) and shifted the iSNV frequency distribution toward lower
values (Figure S4). In this lower stringency dataset, an additional 19 transmission pairs had iSNV
in the donor. However, the bottleneck sizes for all clades were identical to the previous
estimates (Table S3). This suggests that the tight bottlenecks we estimated were not due to
overly stringent variant calling.

Discussion
Here, we used in depth sequencing of two well-sampled household cohorts to define the relationship between transmissibility and transmission bottleneck size. We found that all clades exhibited short serial intervals in our households and low genetic diversity in specimens collected close to the time of transmission. This limited genetic diversity across all clades resulted in a tight estimated bottleneck. In line with bottleneck estimates for first-wave lineages of SARS-CoV-2 we found that VOC clades had a bottleneck of 1 and non-VOC had a bottleneck of 2. These very tight bottleneck estimates were robust to reductions in the stringency in variant-calling.

Consistent with prior studies of SARS-CoV-2 and other viruses, we found low genetic diversity within and between hosts. Allowing for slight differences due to analytic pipelines, previous studies have largely reported low within-host genetic diversity in SARS-CoV-2 (6, 9, 38–40).

Much of this diversity is not shared between hosts, as multiple studies in different settings have measured a tight transmission bottleneck for SARS-CoV-2 (6–10). Tight bottlenecks appear to be broadly applicable across routes of infection and viral family. Potato Y virus (0.5-3.2) and Cucumber mosaic virus (1-2), both transmitted by aphids (41, 42), along with Influenza (1-2), HIV (3, 4), Venezuelan equine encephalitis (43), and HCV (44) have tight bottlenecks.

Additionally, we demonstrate that increased transmissibility, whether through force of infection or immune escape, doesn’t change the bottleneck size for SARS-CoV-2. Genetic diversity constrains bottleneck sizes, and with sufficiently low genetic diversity the bottleneck cannot be greater than one. For both non-VOC and VOC, the short generation time of SARS-
CoV-2 does not allow for diversity to accumulate in the donor, much less transmit. These effects may be exaggerated in highly transmissible variants if time to transmission is shortened. While we did not find variant-specific differences in serial interval in our cohorts, multiple studies that explicitly modeled generation time during household transmission have shown shorter generation times as the pandemic has progressed. Even before variants of concern arose, the generation time of SARS-CoV-2 was decreasing (45), and this trend continued as variants of concern arose with Delta (3.2 days) exhibiting a shorter generation time than Alpha (4.5 days) (46). A shortening of generation could potentially have a larger impact on bottleneck size for other viruses, particularly those that generate more diversity than SARS-CoV-2 prior to transmission.

Our work highlights how transmission bottlenecks, as typically measured, are distinct from infectious dose. Within-host processes in the recipient influence bottleneck size, because not all virions that initiate an infection go on to establish a genetic lineage (1). After infection begins, stochastic loss (genetic drift) during exponential growth, superinfection exclusion, cell-to-cell heterogeneity, and host immune response cause some virions to be lost (47). These within-host processes combined with the starting genetic diversity cause bottleneck size to, in many cases, be smaller than the infectious dose. In experimental systems, genetic barcoding and more frequent sampling of donor and recipient hosts can be used to link bottlenecks to infectious dose and identify lineages that are lost (12, 48).
Our study is subject to at least three limitations. First, in all studies of natural transmission, there is always some ambiguity about who infected whom. In two-infection households, it is possible that both were exposed to a common donor outside the household, and in households with >2 cases, there are multiple possible transfection pairs. Because individuals who don’t transmit to each other are unlikely to share diversity, incorrect pairing will underestimate the bottleneck (5). However, we found that all transmission pairs had equal bottlenecks even when we tested mutually exclusive transmission pairs. Second, virus populations may be spatially segregated within hosts, and the transmitted population may not have been well sampled by our analysis of nasal swabs (49–53). However, given the low viral diversity identified in nearly all cases, even spatially segregated viral populations are likely to be genetically similar to each other. Third, rare diversity may have been under sampled in the donors and recipients due to the sensitivity of our sequencing approach. This possibility was addressed in our analysis of merged technical replicates. Given that more common variants (10-50% frequency) were not shared between hosts, it is unlikely that even perfect detection would find shared iSNV at lower frequencies.

Understanding how different viral properties promote or impede evolution is critical for predicting and effectively monitoring the course of the COVID pandemic. The tight bottlenecks we have estimated for SARS-CoV-2 VOC will both limit the spread of new mutations and reduce the effectiveness of natural selection. Weakened selection will inhibit the evolution of new lineages and may be especially important for new VOC. Whereas other lineages may evolve through non-selective mechanisms, such as genetic drift, the existing VOC have exhibited
strong signals of prior positive selection at the time of their emergence (16)(54–56). The tight bottlenecks identified here will limit the development of highly mutated VOC in typical transmission chains, adding to the evidence that selection over prolonged infections in immunocompromised patients may drive the evolution of SARS-CoV-2 variants of concern (6, 15, 57, 58).

Acknowledgements

We thank all individuals who participated in this study. This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. 75N93021C00015 and R01 AI148371 and from the Centers for Disease Control and Prevention, under U01IP001034.
References

1. Zwart MP, Elena SF. 2015. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. Annu Rev Virol 2:161–79.

2. McCrone JT, Lauring AS. 2018. Genetic bottlenecks in intraspecies virus transmission. Curr Opin Virol 28:20–25.

3. Edwards CT, Holmes EC, Wilson DJ, Viscidi RP, Abrams EJ, Phillips RE, Drummond AJ. 2006. Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. BMC Evol Biol 6:28.

4. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping L-H, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci 105:7552–7557.

5. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauling AS. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. eLife 7:e35962.

6. Braun K, Moreno G, Wagner C, Accola MA, Rehrauer WM, Baker D, Koelle K, O’Connor DH, Bedford T, Friedrich TC, Moncla LH. 2021. Limited within-host diversity and tight
transmission bottlenecks limit SARS-CoV-2 evolution in acutely infected individuals.

bioRxiv https://doi.org/10.1101/2021.04.30.440988.

7. Martin MA, Koelle K. 2021. Comment on “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2.” Sci Transl Med 13.

8. Nicholson MD, Endler L, Popa A, Genger J-W, Bock C, Michor F, Bergthaler A. 2021. Response to comment on “Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2.” Sci Transl Med 13:eabj3222.

9. Hannon WW, Roychoudhury P, Xie H, Shrestha L, Addetia A, Jerome KR, Greninger AL, Bloom JD. 2022. Narrow transmission bottlenecks and limited within-host viral diversity during a SARS-CoV-2 outbreak on a fishing boat. Virus Evol 8:1–9.

10. Li B, Deng A, Li K, Hu Y, Li Z, Shi Y, Xiong Q, Liu Z, Guo Q, Zou L, Zhang H, Zhang M, Ouyang F, Su J, Su W, Xu J, Lin H, Sun J, Peng J, Jiang H, Zhou P, Hu T, Luo M, Zhang Y, Zheng H, Xiao J, Liu T, Tan M, Che R, Zeng H, Zheng Z, Huang Y, Yu J, Yi L, Wu J, Chen J, Zhong H, Deng X, Kang M, Pybus OG, Hall M, Lythgoe KA, Li Y, Yuan J, He J, Lu J. 2022. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Nat Commun 13:460.

11. Tao H, Steel J, Lowen AC. 2014. Intrahost Dynamics of Influenza Virus Reassortment. J Virol 88:7485–7492.
12. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, Garcia-Sastre A, tenOever BR. 2014. Influenza A Virus Transmission Bottlenecks Are Defined by Infection Route and Recipient Host. Cell Host Microbe 16:691–700.

13. Zwart MP, Daròs J-A, Elena SF. 2011. One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. PLOS Pathog 7:e1002122.

14. Koelle K, Lin J, Zhu H, Antia R, Lowen AC, Weissman D. 2022. Masks Do No More Than Prevent Transmission: Theory and Data Undermine the Variolation Hypothesis. medRxiv https://doi.org/10.1101/2022.06.28.22277028.

15. Hill V, Du Plessis L, Peacock TP, Aggarwal D, Colquhoun R, Carabelli AM, Ellaby N, Gallagher E, Groves N, Jackson B, McCrone JT, O’Toole Á, Price A, Sanderson T, Scher E, Southgate J, Volz E, Barclay WS, Barrett JC, Chand M, Connor T, Goodfellow I, Gupta RK, Harrison EM, Loman N, Myers R, Robertson DL, Pybus OG, Rambaut A, The COVID-19 genomics UK (COG-UK) consortium. 2022. The Origins and Molecular Evolution of SARS-CoV-2 Lineage B.1.1.7 in the UK. Virus Evol veac080.

16. Telenti A, Hodcroft EB, Robertson DL. 2022. The Evolution and Biology of SARS-CoV-2 Variants. Cold Spring Harb Perspect Med 12:1–24.

17. Cai Y, Zhang J, Xiao T, Lavine CL, Rawson S, Peng H, Zhu H, Anand K, Tong P, Gautam A, Lu S, Sterling SM, Walsh Jr. RM, Rits-Volloch S, Lu J, Wesemann DR, Yang W, Seaman MS, Chen B. 2021. Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. Science 373:642–648.
18. Zhang J, Xiao T, Cai Y, Lavine CL, Peng H, Zhu H, Anand K, Tong P, Gautam A, Mayer ML, Walsh RM, Rits-Volloch S, Wesemann DR, Yang W, Seaman MS, Lu J, Chen B. 2021. Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. Science 374:1353–1360.

19. Araf Y, Akter F, Tang Y, Fatemi R, Parvez MdSA, Zheng C, Hossain MdG. 2022. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. J Med Virol 94:1825–1832.

20. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. 2022. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol 94:1641–1649.

21. Syed AM, Taha TY, Khalid MM, Tabata T, Chen IP, Sreekumar B, Chen P-Y, Hayashi JM, Soczek KM, Ott M, Doudna JA. 2021. Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles. Science 374:1626–1632.

22. Puhach O, Adea K, Hulo N, Sattonnet P, Genecand C, Iten A, Jacquérisoz F, Kaiser L, Vetter P, Eckerle I, Meyer B. 2022. Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron SARS-CoV-2. Nat Med 28:1491–1500.

23. Thorne LG, Bouhaddou M, Reuschl A-K, Zuliani-Alvarez L, Polacco B, Pelin A, Batra J, Whelan MVX, Hosmillo M, Fossati A, Ragazzini R, Jungreis I, Ummadi M, Rojc A, Turner J, Bischof ML, Obernier K, Braberg H, Soucheray M, Richards A, Chen K-H, Harjai B, Memon D, Hiatt J, Rosales R, McGovern BL, Jahun A, Fabius JM, White K,
Evolution of enhanced innate immune evasion by SARS-CoV-2. 7897. Nature 602:487–495.

Hui KPY, Ho JCW, Cheung M, Ng K, Ching RHH, Lai K, Kam TT, Gu H, Sit K-Y, Hsin MKY, Au TWK, Poon LLM, Peiris M, Nicholls JM, Chan MCW. 2022. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. 7902. Nature 603:715–720.

Meng B, Abdullahi A, Ferreira IATM, Goonawardane N, Saito A, Kimura I, Yamasoba D, Gerber PP, Fathi S, Rathore S, Zepeda SK, Papa G, Kemp SA, Ikeda T, Toyoda M, Tan TS, Kuramochi J, Mitsunaga S, Ueno T, Shirakawa K, Takaori-Kondo A, Brevini T, Mallery DL, Charles OJ, The CITIID-NIHR BioResource COVID-19 Collaboration, Baker S, Dougan G, Hess C, Kingston N, Lehner PJ, Lyons PA, Matheson NJ, Ouwehand WH, Saunders C, Summers C, Thaventhiran JED, Toshner M, Weekes MP, Maxwell P, Shaw A, Bucke A, Calder J, Canna L, Domingo J, Elmer A, Fuller S, Harris J, Hewitt S, Kennet J, Jose S, Kourampa J, Meadows A, O’Brien C, Price J, Publico C, Rastall R, Ribeiro C, Rowlands J, Ruffolo V, Tordesillas H, Bullman B, Dunmore BJ, Gräf S, Hodgson J, Huang C, Hunter K, Jones E, Legchenko E, Matara C, Martin J, Mescia F, O’Donnell C, Pointon L, Shih J, Sutcliffe R, Tilly T, Treacy C, Tong Z, Wood J, Wylot M, Betancourt A, Bower G, Cossetti C, De Sa A, Epping M, Fawke S, Gleadall N, Grenfell R, Hinch A, Jackson S, Jarvis I, Krishna B, Nice F, Omarjee O, Perera M, Potts M, Richoz N, Romashova V, Stefanucci L, Strezlecki M, Turner L, De Bie EMDD, Bunclark K, Josipovic M, Mackay M, Butcher H, Caputo D, Chandler M, Chinnery P, Clapham-Riley D, Dewhurst E,
Fernandez C, Furlong A, Graves B, Gray J, Hein S, Ivers T, Le Gresley E, Linger R, Kanicki M, King R, Kingston N, Meloy S, Moulton A, Muldoon F, Ovington N, Papadion S, Penkett CJ, Phelan I, Ranganath V, Paraschiv R, Sage A, Sambrook J, Scholtes I, Schon K, Stark H, Stirrups KE, Townsend P, Walker N, Webster J, The Genotype to Phenotype Japan (G2P-Japan) Consortium, Butlertanaka EP, Tanaka YL, Ito J, Uriu K, Kosugi Y, Suganami M, Oide A, Yokoyama M, Chiba M, Motozono C, Nasser H, Shimizu R, Kitazato K, Hasebe H, Irie T, Nakagawa S, Wu J, Takahashi M, Fukuhara T, Shimizu K, Tsushima K, Kubo H, Kazuma Y, Nomura R, Horisawa Y, Nagata K, Kawai Y, Yanagida Y, Tashiro Y, Tokunaga K, Ozono S, Kawabata R, Morizako N, Sadamasu K, Asakura H, Nagashima M, Yoshimura K, Ecuador-COVID19 Consortium, Cárdenas P, Muñoz E, Barragan V, Márquez S, Prado-Vivar B, Becerra-Wong M, Caravajal M, Trueba G, Rojas-Silva P, Grunauer M, Gutierrez B, Guadalupe JJ, Fernández-Cadena JC, Andrade-Molina D, Baldeon M, Pinos A, Bowen JE, Joshi A, Walls AC, Jackson L, Martin D, Smith KGC, Bradley J, Briggs JAG, Choi J, Masissoon E, Meyer KB, Mlcochova P, Ceron-Gutierrez L, Doffinger R, Teichmann SA, Fisher AJ, Pizzuto MS, de Marco A, Corti D, Hosmillo M, Lee JH, James LC, Thukral L, Veesler D, Sigal A, Sampaziotis F, Goodfellow IG, Matheson NJ, Sato K, Gupta RK. 2022. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. Nature 603:706–714.

26. Malosh RE, Petrie JG, Callear AP, Monto AS, Martin ET. 2021. Home collection of nasal swabs for detection of influenza in the Household Influenza Vaccine Evaluation Study. Influenza Other Respir Viruses 15:227–234.

27. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
28. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, Tan AL, Paul LM, Brackney DE, Grewal S, Gurfield N, Van Rompay KKA, Isern S, Michael SF, Coffey LL, Loman NJ, Andersen KG. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol 20:8.

29. De Maio N, Walker C, Borges R, Weilguny L, Slodkowicz G, Goldman Ni. 2020. Issues with SARS-CoV-2 sequencing data. Virological.

30. Bendall E, Paz-Bailey G, Santiago GA, Porucznik CA, Stanford JB, Stockwell MS, Duque J, Jedy Z, Veuilla V, Major C, Rivera-Amill V, Rolfes MA, Dawood FS, Lauring AS. 2022. SARS-CoV-2 genomic diversity in households highlights the challenges of sequence-based transmission inference. medRxiv https://doi.org/10.1101/2022.08.09.22278452.

31. Aksametov I, Roemer C, Hodcroft EB, Neher RA. 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw 6:3773.

32. 2022. Tracking SARS-CoV-2 variants. https://www.who.int/activities/tracking-SARS-CoV-2-variants. Retrieved 29 August 2022.

33. Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E, Koelle K. 2017. Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. J Virol 91:e00171-17.

34. Petrie JG, Eisenberg MC, Lauring AS, Gilbert J, Harrison SM, DeJonge PM, Martin ET. 2022. The variant-specific burden of SARS-CoV-2 in Michigan: March 2020 through November 2021. J Med Virol 94:5251–5259.
35. Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, Sebra R, Halpin RA, Guan Y, Twaddle A, DePasse JV, Stockwell TB, Wentworth DE, Holmes EC, Greenbaum B, Peiris JSM, Cowling BJ, Ghedin E. 2016. Quantifying influenza virus diversity and transmission in humans. 2. Nat Genet 48:195–200.

36. Xue KS, Bloom JD. 2019. Reconciling disparate estimates of viral genetic diversity during human influenza infections. 9. Nat Genet 51:1298–1301.

37. Popa A, Genger J-W, Nicholson MD, Penz T, Schmid D, Aberle SW, Agerer B, Lercher A, Endler L, Colaço H, Smyth M, Schuster M, Grau ML, Martínez-Jiménez F, Pich O, Borena W, Pawelka E, Keszei Z, Senekowitsch M, Laine J, Aberle JH, Redlberger-Fritz M, Karolyi M, Zoufaly A, Maritschnik S, Borkovec M, Hufnagl P, Nairz M, Weiss G, Wolfinger MT, von Laer D, Superti-Furga G, Lopez-Bigas N, Puchhammer-Stöckl E, Allerberger F, Michor F, Bock C, Bergthaler A. 2020. Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Sci Transl Med 12:eabe2555.

38. Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, Andersson M, Otecko N, Wise EL, Moore N, Lynch J, Kidd S, Cortes N, Mori M, Williams R, Vernet G, Justice A, Green A, Nicholls SM, Ansari MA, Abeler-Dörner L, Moore CE, Peto TEA, Eyre DW, Shaw R, Simmonds P, Buck D, Todd JA, on behalf of the Oxford Virus Sequencing Analysis Group (OVSG), Connor TR, Ashraf S, da Silva Filipe A, Shepherd J, Thomson EC, The COVID-19 Genomics UK (COG-UK) Consortium, Bonsall D, Fraser C, Golubchik T. 2021. SARS-CoV-2 within-host diversity and transmission. Science 372:eabg0821.
39. Tonkin-Hill G, Martincorena I, Amato R, Lawson AR, Gerstung M, Johnston I, Jackson DK, Park N, Lensing SV, Quail MA, Gonçalves S, Ariani C, Spencer Chapman M, Hamilton WL, Meredith LW, Hall G, Jahun AS, Chaudhry Y, Hosmillo M, Pinckert ML, Georgana I, Yakovleva A, Caller LG, Caddy SL, Feltwell T, Khokhar FA, Houldcroft CJ, Curran MD, Parmar S, The COVID-19 Genomics UK (COG-UK) Consortium, Alderton A, Nelson R, Harrison EM, Sillitoe J, Bentley SD, Barrett JC, Torok ME, Goodfellow IG, Langford C, Kwiatkowski D, Wellcome Sanger Institute COVID-19 Surveillance Team. 2021. Patterns of within-host genetic diversity in SARS-CoV-2. eLife 10:e66857.

40. Valesano AL, Rumfelt KE, Dimcheff DE, Blair CN, Fitzsimmons WJ, Petrie JG, Martin ET, Lauring AS. 2021. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. PLOS Pathog 17:e1009499.

41. Ali A, Li H, Schneider WL, Sherman DJ, Gray S, Smith D, Roossinck MJ. 2006. Analysis of Genetic Bottlenecks during Horizontal Transmission of Cucumber Mosaic Virus. J Virol 80:8345–8350.

42. Moury B, Fabre F, Senoussi R. 2007. Estimation of the number of virus particles transmitted by an insect vector. Proc Natl Acad Sci 104:17891–17896.

43. Smith DR, Adams AP, Kenney JL, Wang E, Weaver SC. 2008. Venezuelan equine encephalitis virus in the mosquito vector Aedes taeniorhynchus: infection initiated by a small number of susceptible epithelial cells and a population bottleneck. Virology 372:176–186.
44. Bull RA, Luciani F, McElroy K, Gaudieri S, Pham ST, Chopra A, Cameron B, Maher L,
Dore GJ, White PA, Lloyd AR. 2011. Sequential Bottlenecks Drive Viral Evolution in
Early Acute Hepatitis C Virus Infection. PLoS Pathog 7:e1002243.

45. Hart WS, Abbott S, Endo A, Hellewell J, Miller E, Andrews N, Maini PK, Funk S,
Thompson RN. 2022. Inference of the SARS-CoV-2 generation time using UK household
data. eLife 11:e70767.

46. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, Thompson RN. 2022.
Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis.
Lancet Infect Dis 22:603–610.

47. Gutiérrez S’n, Michalakis Y, Blanc Ś phane. 2012. Virus population bottlenecks during
within-host progression and host-to-host transmission. Curr Opin Virol 2.

48. Weger-Lucarelli J, Garcia SM, Rückert C, Byas A, O’Connor SL, Aliota MT, Friedrich TC,
O’Connor DH, Ebel GD. 2018. Using barcoded Zika virus to assess virus population
structure in vitro and in Aedes aegypti mosquitoes. Virology 521:138–148.

49. Hamada N, Hara K, Kashiwagi T, Imamura Y, Nakazono Y, Watanabe H, Imamura Y,
Chijiwa K. 2012. Intrahost emergent dynamics of oseltamivir-resistant virus of pandemic
influenza A (H1N1) 2009 in a fatally immunocompromised patient. J Infect Chemother
18:865–871.

50. Gallagher ME, Brooke CB, Ke R, Koelle K. 2018. Causes and Consequences of Spatial
Within-Host Viral Spread. 11. Viruses 10:627.
Desai N, Neyaz A, Szabolcs A, Shih AR, Chen JH, Thapar V, Nieman LT, Solovyov A, Mehta A, Lieb DJ, Kulkarni AS, Jaicks C, Xu KH, Raabe MJ, Pinto CJ, Juric D, Chebib I, Colvin RB, Kim AY, Monroe R, Warren SE, Danaher P, Reeves JW, Gong J, Rueckert EH, Greenbaum BD, Hacohen N, Lagana SM, Rivera MN, Sholl LM, Stone JR, Ting DT, Deshpande V. 2020. Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection. 1. Nat Commun 11:6319.

Ganti K, Bagga A, Ferreri LM, Geiger G, Carnaccini S, Caceres CJ, Seibert B, Li Y, Wang L, Kwon T, Li Y, Morozov I, Ma W, Richt JA, Perez DR, Koelle K, Lowen AC. 2022. Influenza A virus reassortment in mammals gives rise to genetically distinct within-host sub-populations. bioRxiv https://doi.org/10.1101/2022.02.08.479600.

Farjo M, Koelle K, Martin MA, Gibson LL, Walden KKO, Rendon G, Fields CJ, Alnaji FG, Gallagher N, Luo CH, Mostafa HH, Manabe YC, Pekosz A, Smith RL, McManus DD, Brooke CB. 2022. Within-host evolutionary dynamics and tissue compartmentalization during acute SARS-CoV-2 infection. bioRxiv https://doi.org/10.1101/2022.06.21.497047.

van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. 2020. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. 1. Nat Commun 11:5986.

MacLean OA, Orton R, Singer JB, Robertson DL. 2020. Response to “On the origin and continuing evolution of SARS-CoV-2.”

Molina-Mora JA, Reales-González J, Camacho E, Duarte-Martínez F, Tsukayama P, Soto-Garita C, Brenes H, Cordero-Laurent E, Santos AR dos, Salgado CG, Silva CS, Souza JS
Overview of the SARS-CoV-2 genotypes circulating in Latin America during 2021. bioRxiv https://doi.org/10.1101/2022.08.19.504579.

Recurrent SARS-CoV-2 mutations in immunodeficient patients. Virus Evol 8:veac050.

Investigating the evolutionary origins of the first three SARS-CoV-2 variants of concern. Front Virol 2.
Figure Legends

Figure 1. Serial interval and timing of sample collection. (A) Days between index symptom onset and household contact symptom onset for the indicated clades. “Non-VOC” includes all lineages not designated as a WHO variant of concern. No Beta variant transmission pairs were analyzed. (B) Days between symptom onset and collection of the sequenced specimen for the index case. Index cases from MHome are indicated in teal, and index cases from HIVE are indicated in red. Omicron had a shorter time between index symptom onset and sample collection for sequencing than non-VOC (df=3, F=8.138, p < 0.001) and Alpha. HIVE households had a shorter time than MHome households (df =1, F =15.363, p < 0.001). (C) RT-qPCR cycle threshold values (inverted y-axis) for all specimens collected from index cases. Sequenced specimens are indicated with filled circles.

Figure 2. Genetic diversity in sequenced specimens. (A) Histogram of the number of iSNV per specimen. (B) iSNV frequency histogram.

Figure 3. Diversity across transmission pairs. (A) The number of individuals per household with sequenced specimens. Colors represent the different clades. (B) Shared genetic diversity between transmission pairs. Each point is an iSNV within a transmission pair. Red points indicate mutation C29708T, which was shared in a single household (see text).
Figure S1. Sequencing coverage and consistency. (A) Boxplot of median (+/- IQR) coverage across the genome in 400bp non-overlapping sliding windows. (B) Frequency of iSNV in each replicate for iSNV that were identified in both replicates.

Figure S2. Inclusion and exclusion of transmission pairs. (A) Examples of possible transmission pairs in households. In each panel, the index cases are in blue, and the household contacts are in black. The grey arrows indicate transmission pairs, and they point from the donor to the recipient. (B) Consensus genome alignments inconsistent with household transmission. The genomes were visualized using Nextclade. Both Nextclade and Pango lineages are reported. Colored bars are mutations with reference to the Wuhan-Hu-1/2019 (MN908947) strain. Gray partial bars indicate missing data. Asterisks next to household names indicate households that were removed from further analyses. Asterisks next to sample names indicate samples that were removed from further analyses, while the rest of the household was retained. The black cross (†) indicates a household with two separate transmission pairs.

Figure S3. Timing of symptom onset and specimen collection by household. Each panel shows a household, grouped by the indicated clades. Within each household, blue symbols indicate index case(s) and black symbols indicate household contact(s). Open triangles indicate time of symptom onset and filled triangles indicate specimens that were sequenced. If there is no symptom onset, the case was considered to be asymptomatic.
Figure S4. Timing of symptom onset and specimen collection by household. Each panel shows a household, grouped by the indicated clades. Within each household, blue symbols indicate index case(s) and black symbols indicate household contact(s). Open triangles indicate time of symptom onset and filled triangles indicate specimens that were sequenced. If there is no symptom onset, the case was considered to be asymptomatic.

Figure S5. iSNV detected when sequencing replicates are merged. (A) The number of iSNV per specimen. Nearly still had a low number of iSNV. However, merging the reads greatly increased the number of iSNV in two individuals. These iSNV were near the 2% threshold. (B) Most iSNV were found at low frequencies. The frequency distribution shifted toward lower frequencies compared to when iSNV had to be detected in both replicates.
### Table 1

|                                | Non-VOC | Alpha | Gamma | Delta | Omicron | Total |
|--------------------------------|---------|-------|-------|-------|---------|-------|
| **Individuals with successful** |         |       |       |       |         |       |
| sequencing                     | 22      | 21    | 3     | 25    | 40      | 111   |
| **Households with successful** |         |       |       |       |         |       |
| sequencing*                    | 11      | 7     | 1     | 12    | 15      | 46    |
| **Possible transmission pairs**|         |       |       |       |         |       |
|                                | 26      | 34    | 2     | 19    | 55      | 134   |
| **Transmission pairs included**|         |       |       |       |         |       |
| in bottleneck analysis**       | 15      | 19    | 1     | 12    | 17      | 64    |

* Households that have 2 or more individuals with successful sequencing

** Only includes transmission pairs where there are iSNV in the donor
Figure 1. Serial interval and timing of sample collection. (A) Days between index symptom onset and household contact symptom onset for the indicated clades. “Non-VOC” includes all lineages not designated as a WHO variant of concern. No Beta variant transmission pairs were analyzed. (B) Days between symptom onset and collection of the sequenced specimen for the index case. Index cases from MHome are indicated in teal, and index cases from HIVE are indicated in red. Omicron had a shorter time between index symptom onset and sample collection for sequencing than non-VOC (df=3, F=8.138, p < 0.001) and HIVE households had a shorter time than MHome households (df =1, F =15.363, p < 0.001). (C) RT-qPCR cycle threshold values (inverted y-axis) for all specimens collected from index cases. Sequenced specimens are indicated with filled circles.
Figure 2. Genetic diversity in sequenced specimens. (A) Histogram of the number of iSNV per specimen. (B) iSNV frequency histogram.
Figure 3. Diversity across transmission pairs. (A) The number of individuals per household with sequenced specimens. Colors represent the different clades. (B) Shared genetic diversity between transmission pairs. Each point is an iSNV within a transmission pair. Red points indicate mutation C29708T, which was shared in a single household.
Figure S1. Sequencing coverage and consistency. (A) Boxplot of median (+/− IQR) coverage across the genome in 400bp non-overlapping sliding windows. (B) Frequency of iSNV in each replicate for iSNV that were identified in both replicates.
Figure S2. Inclusion and exclusion of transmission pairs. (A) Examples of possible transmission pairs in households. In each panel, the index cases are in blue, and the household contacts are in black. The grey arrows indicate transmission pairs, and they point from the donor to the recipient. (B) Consensus genome alignments inconsistent with household transmission. The genomes were visualized using Nextclade. Both Nextclade and Pango lineages are reported. Colored bars are mutations with reference to the Wuhan-Hu-1/2019 (MNR908947) strain. Gray partial bars indicate missing data. Asterisks next to household names indicate households that were removed from further analyses. Asterisks next to sample names indicate samples that were removed from further analyses, while the rest of the household was retained. The black cross (†) indicates a household with two separate transmission pairs.
### Non-VOC

| HH1 | HH10 | HH18 | HH2 |
|-----|------|------|-----|
| ![Graph](image1.png) | ![Graph](image2.png) | ![Graph](image3.png) | ![Graph](image4.png) |

| HH3 | HH4 | HH5 | HH6 |
|-----|-----|-----|-----|
| ![Graph](image5.png) | ![Graph](image6.png) | ![Graph](image7.png) | ![Graph](image8.png) |

| HH7 | HH8 | HH9 |
|-----|-----|-----|
| ![Graph](image9.png) | ![Graph](image10.png) | ![Graph](image11.png) |

### Delta

| HH20 | HH21 | HH22 | HH24 |
|------|------|------|------|
| ![Graph](image12.png) | ![Graph](image13.png) | ![Graph](image14.png) | ![Graph](image15.png) |

| HH25.1 | HH25.2 | HH26 | HH32.1 |
|--------|--------|------|--------|
| ![Graph](image16.png) | ![Graph](image17.png) | ![Graph](image18.png) | ![Graph](image19.png) |

| HH37 | HH40 | HH42 | HH46 |
|------|------|------|------|
| ![Graph](image20.png) | ![Graph](image21.png) | ![Graph](image22.png) | ![Graph](image23.png) |

### Alpha

| HH11 | HH12 | HH13 | HH15 |
|------|------|------|------|
| ![Graph](image24.png) | ![Graph](image25.png) | ![Graph](image26.png) | ![Graph](image27.png) |

| HH16 | HH17 | HH19 |
|------|------|------|
| ![Graph](image28.png) | ![Graph](image29.png) | ![Graph](image30.png) |

Figure S3. Timing of symptom onset and specimen collection by household. Each panel shows a household, grouped by the indicated clades. Within each household, blue symbols indicate index case(s) and black symbols indicate household contact(s). Open triangles indicate time of symptom onset and filled triangles indicate specimens that were sequenced. If there is no symptom onset, the case was considered to be asymptomatic.
Figure S4. Timing of symptom onset and specimen collection by household. Each panel shows a household, grouped by the indicated clades. Within each household, blue symbols indicate index case(s) and black symbols indicate household contact(s). Open triangles indicate time of symptom onset and filled triangles indicate specimens that were sequenced. If there is no symptom onset, the case was considered to be asymptomatic.
Figure S5. iSNV detected when sequencing replicates are merged. (A) The number of iSNV per specimen. Nearly still had a low number of iSNV. However, merging the reads greatly increased the number of iSNV in two individuals. These iSNV were near the 2% threshold. (B) Most iSNV were found at low frequencies. The frequency distribution shifted toward lower frequencies compared to when iSNV had to be detected in both replicates.