

RESEARCH ARTICLE

In depth sequencing of a serially sampled household cohort reveals the within-host dynamics of Omicron SARS-CoV-2 and rare selection of novel spike variants

Emily E. Bendall¹, Derek Dimcheff², Leigh Papalambros², William J. Fitzsimmons², Yuwei Zhu³, Jonathan Schmitz⁴, Natasha Halasa⁵, James Chappell⁵, Emily T. Martin⁶, Jessica E. Biddle⁷, Sarah E. Smith-Jeffcoat⁷, Melissa A. Rolfes⁷, Alexandra Mellis⁷, H. Keipp Talbot^{8,9}, Carlos Grijalva⁹, Adam S. Lauring^{1,2*}

1 Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, United States of America, **2** Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **4** Department of Pathology, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **5** Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **6** Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, United States of America, **7** Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, **8** Department of Health Policy, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **9** Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America

* alauring@med.umich.edu



OPEN ACCESS

Citation: Bendall EE, Dimcheff D, Papalambros L, Fitzsimmons WJ, Zhu Y, Schmitz J, et al. (2025) In depth sequencing of a serially sampled household cohort reveals the within-host dynamics of Omicron SARS-CoV-2 and rare selection of novel spike variants. PLoS Pathog 21(4): e1013134. <https://doi.org/10.1371/journal.ppat.1013134>

Editor: Anice C. Lowen, Emory University School of Medicine, UNITED STATES OF AMERICA

Received: January 10, 2025

Accepted: April 16, 2025

Published: April 28, 2025

Copyright: © This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data availability statement: Raw sequence reads are available at the NCBI Sequence Read Archive, Bioproject PRJNA1159790.

Abstract

SARS-CoV-2 has undergone repeated and rapid evolution to circumvent host immunity. However, outside of prolonged infections in immunocompromised hosts, within-host positive selection has rarely been detected. Here we combine daily longitudinal sampling of individuals with replicate sequencing to increase the accuracy of and lower the threshold for variant calling. We sequenced 577 specimens from 105 individuals in a household cohort during the BA.1/BA.2 variant period. Individuals exhibited extremely low viral diversity, and we estimated a low within-host evolutionary rate. Within-host dynamics were dominated by genetic drift and purifying selection. Positive selection was rare but highly concentrated in spike. A Wright Fisher Approximate Bayesian Computational model identified positive selection at 14 loci with 7 in spike, including S:448 and S:339. This detectable immune-mediated selection is unusual in acute respiratory infections and may be caused by the relatively narrow antibody repertoire in individuals during the early Omicron phase of the SARS-CoV-2 pandemic.

Metadata and analysis code are available at https://github.com/lauringlab/SARS-CoV-2_Within-host_RVTN

Funding: Primary funding for the RVTN study was provided by the US Centers for Disease Control and Prevention (CDC 75D30121C11656). CGG was partially supported by NIH K24 AI148459. Scientists from the US CDC participated in all aspects of this study, including its design, analysis, interpretation of data, writing the report, and the decision to submit the article for publication. Sequencing and associated analysis was supported by a NIH R01 AI148371 (to ASL and ETM), the Penn Center for Excellence in Influenza Research and Response, Penn-CEIRR, NIH 75N93021C00015 (to ASL and ETM), and the Michigan Infectious Disease Genomics Center, NIH U19 AI181767 (to ASL).

Competing interests: The authors have read the journal's policy and the authors of this manuscript declare the following competing interests: Carlos Grijalva reports grants from NIH, CDC, AHRQ, FDA, Campbell Alliance/Syneos Health, consulting fees and participating on an advisory board for Merck, outside the submitted work. Natasha Halasa reports grants from Sanofi, Quidel, and Merck, outside the submitted work. James Chappell reports research support from Merck outside the submitted work. Adam Luring reports receiving grants from CDC, NIAID, Burroughs Wellcome Fund, Flu Lab, and consulting fees from Roche, outside the submitted work. Emily Martin reports receiving a grant from Merck, outside the submitted work.

Author summary

Throughout the SARS-CoV-2 pandemic new variants have continually arisen due to population level immunity from vaccinations and previous infections. A similar process within-hosts may be occurring, where new variants evolve during the course of an infection to escape partial immunity from previous exposures. However, within-host positive selection has rarely been detected in acute SARS-CoV-2 infections. Here we studied individuals from a case-ascertained household cohort during the BA.1/BA.2 wave. Nasal swab specimens were collected daily for 10 days and sequenced in replicate for high accuracy. Individuals exhibited extremely low viral diversity with stochastic processes dominating. Positive selection was rare but highly concentrated in spike. The majority of spike mutations we identified as positively selected have been implicated in immune evasion previously. This infrequent but detectable positive selection may be due to the timing of these infections relative to the emergence of SARS-CoV-2

Introduction

As SARS-CoV-2 continues to circulate, population immunity from infections and vaccinations has resulted in the evolution of new variants that quickly become the dominant circulating strain [1,2]. This has contributed to decreased vaccine effectiveness, and in response, multiple reformulations of the SARS-CoV-2 vaccines [3–5]. The continual evolution of SARS-CoV-2 as a result of selection from the host adaptive immune system is likely to continue. Similar to this global antigenic drift, partial immunity from previous exposure may lead to the selection of new antigenic variants within hosts [6,7]. Because all variation originates from intrahost processes, understanding within-host dynamics is crucial to understanding the evolutionary trajectory of SARS-CoV-2.

To date, there has been limited evidence of positive selection of immune escape variants within individuals with acute, self-limited SARS-CoV-2 infections. We and others have found that SARS-CoV-2 infections exhibit low genetic diversity and few *de novo* mutations that reach significant frequencies [8–11]. Select studies have identified spike variants in sites known to confer antibody resistance [8,11]. Additionally, Farjo *et al.* found nonsynonymous intrahost single nucleotide variants (iSNVs) to be enriched in individuals who had been vaccinated or previously infected [11]. Regions of within-host positive selection in non-spike regions have also been detected when comparing intrahost diversity of synonymous and nonsynonymous variants (p_N/p_S) [12]. However, genetic hitchhiking (i.e., changes in a mutation's frequency as a result of selection on a linked site on the same genome/chromosome) and genetic drift make it difficult to accurately detect positive selection with viruses from only a single timepoint [13].

Most studies of serially sampled individuals come from prolonged infections in immunocompromised patients, where immune escape variants have repeatedly been found [14–18]. Prolonged infections release the virus from the frequent population

bottlenecks characteristic of acute infections, increasing the amount of genetic variation and allowing time for selection to occur [19]. The selection pressures in immunocompromised individuals may differ from those in immunocompetent individuals with acute infections, with selection for increased cell-cell transmission and viral packaging [17]. Additionally, monoclonal antibodies commonly used to treat immunocompromised individuals may exert more targeted selection than a polyclonal response from prior exposure in immunocompetent individuals [20].

To more thoroughly examine the role of positive selection within hosts during acute SARS-CoV-2 infections, we studied individuals from a case-ascertained household cohort, in which nasal swab specimens were collected daily for 10 days after enrollment. All specimens were sequenced in duplicate, allowing for robust variant calling at a very low frequency threshold (0.5%). With serial sampling and low frequency variant calling, we were able to define the within-host divergence of SARS-CoV-2 populations, detect genetic hitchhiking, and identify rare, but potentially significant, instances of positive selection in spike.

Methods

Cohort and specimens

Households were enrolled through the CDC-sponsored Respiratory Virus Transmission Network – Sentinel (RVTN-S), a case ascertained household transmission study coordinated at Vanderbilt University Medical Center [21]. All individuals provided written, informed consent and parents/guardians provided written, informed consent for minors. Individuals included in the current study were enrolled in Nashville, TN from September 2021 to February 2022. The study was reviewed and approved by the Vanderbilt University Medical Center Institutional Review Board (see 45 C.F.R. part 46.114; 21 C.F.R. part 56.114). Index cases (i.e., the first household members with laboratory-confirmed SARS-CoV-2 infection) were identified and recruited from ambulatory clinics, emergency departments, or other settings that performed SARS-CoV-2 testing. Index cases and their households were screened and enrolled within 6 days of the earliest symptom onset date within the household. Vaccination status was determined by plausible self-report (report of a manufacturer and either a date or location) or vaccine verification through vaccination cards, state registries, and medical records. Only vaccines received more than 14 days before the date of the earliest symptom onset in the household were considered.

Nasal swabs specimens were self- or parent-collected daily from all enrolled household members during follow-up for 10 days and tested for SARS-CoV-2. Nasal swabs were tested by transcription mediated amplification using the Panther Hologic system. All available specimens were processed for sequencing as described below.

Sequencing and variant calling

SARS-CoV-2 positive specimens with a cycle threshold (Ct) value ≤ 32 were sequenced in duplicate after the RNA extraction step. 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 V5.3.2 primer sets. After barcoding, libraries were pooled in equal volume. The pooled libraries (up to 96 specimens per pool) were size selected by gel extraction and sequenced on an Illumina NextSeq (2x300, P1 chemistry).

For the first specimen per individual with adequate sequencing, we aligned the sequencing reads to the MN908947.3 reference using BWA-mem v0.7.15 [22]. Primers were trimmed using iVar v1.2.1 [23]. Reads from both replicates were combined and used to make a within host consensus sequence using a script from Xue et al [24]. Specimens were considered successfully sequenced if both replicates had an average genome wide coverage $> 1000\times$. All specimens were aligned to their respective within-host consensus sequences. Primers were trimmed using iVar and reads from amplicons with mismatched primers were masked. Intrahost single nucleotide variants (iSNV) were identified for each replicate separately using iVar [23] with the following criteria: frequency 0.005–0.995, p-value $< 1 \times 10^{-5}$, variant position coverage depth $> 400\times$. We also masked ambiguous and homoplastic sites that have previously been designated as probably

erroneous [25]. Specific to this study, T11075C was found at low frequencies in 48 individuals. This is indicative of a sequencing artifact, and T11075C was also masked. Finally, to minimize the possibility of false variants being detected, the variants had to be present in both sequencing replicates. The variant frequencies were averaged for all analyses. Indels were not evaluated. Lineages were determined with Nextclade [26] and Pango [27,28], based on the within-host consensus sequence.

iSNV Dynamics and divergence rates

We calculated the divergence rate as in Xue *et al* [24]. Briefly, we calculated the rate of evolution by summing the frequencies of within-host mutations (non-consensus allele in first specimen) and dividing by the number of available sites and the time since the infection began. We calculated the rates separately for nonsynonymous and synonymous mutations. We used 0.77 for the proportion of available sites for nonsynonymous mutations and 0.23 for synonymous. To determine the number of available sites, we multiplied the proportion of sites available by the length of the coding sequence of the MN908947.3 reference. Because symptoms typically start 2–3 days post infection and nasal swab collection occurred after symptom onset among most individuals, we added 2 days to the time since symptom onset to obtain the time elapsed between infection and sampling [29–31]. We excluded individuals who were asymptomatic from the divergence rate analysis, as we are not able to date their infection by symptom onset (e.g., 2–3 days prior as above). Because the calculated rate of divergence varied over the course of the infection, we also calculated the rate using the specimen with the highest viral load for each individual to control for timing within the course of the infection (S1 Fig). In addition, we used linear regression to estimate the divergence rates in individuals with multiple specimens. We calculated per-site viral divergence for each specimen. For each person, a linear regression was performed with the per specimen divergences and the days post infection with viral load as a covariate. A person's divergence rate was the slope of this regression line. The rate was calculated for the whole genome and for each gene separately.

Mann-Whitney U tests were used to determine if the number iSNV per specimen and iSNV frequencies differed by mutation type, vaccination, and age group. Kruskal-Wallis tests were performed to determine if the number iSNV per specimen and iSNV frequencies differed by clade and days post symptom onset. Generalized linear models were used to determine if viral load impacted iSNV frequency and iSNV count (Poisson distributions). Mann-Whitney U tests were used to determine if the divergence rate differed by vaccination and age group. Kruskal-Wallis tests were performed to determine if divergence rate differed by clade, gene and days post infection. A generalized linear model was used to determine if viral load impacted divergence rates. All analyses were conducted using R version 4.3.1.

Analysis of selection

The study period included the Delta, BA.1, and BA.2 variant periods of the SARS-CoV-2 pandemic. For each of these clades, we looked at the lineage-defining mutations in spike of the subsequent wave (i.e., BA.1, BA.2, and BA.4/BA.5). We compared the iSNV within our specimens to these lineage defining mutations.

We also used Wright Fisher Approximate Bayesian Computation (WFABC) to estimate the effective populations size (N_e) and per locus selection coefficient (s) based on allele trajectories [32]. Generation times of 8 hours and 12 hours were used [33–35]. To maximize the number of loci used in the calculation of N_e and to avoid violating the assumption that most loci are neutral, we estimated a single N_e using all loci from individuals in which the first two specimens sequenced were collected one day apart. 10,000 bootstrap replicates were performed to obtain a posterior distribution. A fixed N_e was used for the per locus selection coefficient simulations, with the analysis repeated for the mean N_e , and the ± 1 standard deviation N_e estimated from the previous step. A uniform prior between s of -0.5 and 0.5 was used with 100,000 simulations and an acceptance rate of 0.01. We estimated the 95% highest posterior density intervals using the *boa* package [36] in R. We considered a site to be positively selected if the 95% highest posterior density did not include 0 for all three effective population sizes.

To understand how within-host selection relates to between host selection, we used the SARS-CoV-2 Nextstrain build [37] (nextstrain/ncov, the Nextstrain team) to examine the global frequencies of iSNV that were under positive within-host selection in our study. We also compared the selection coefficients we estimated to the selection coefficients that Bloom and Neher [38] estimated from the global phylogeny.

Results

There were 212 SARS-CoV-2 infected individuals enrolled from September 2021 to February 2022 in this case-ascertained household cohort. None of the individuals enrolled received monoclonal antibodies or antivirals. Of these, we successfully sequenced 577/825 (70%) specimens from 105 individuals. Ninety-nine out of 105 (94%) individuals had multiple specimens successfully sequenced (Fig 1A, S1 Table). Consistent with the viruses circulating in the United States during this timeframe, the individuals in the study were infected with Delta, BA.1, and BA.2. Depth of coverage was generally high (S2 Fig) and iSNV frequency was similar between replicates (Fig 1B). The number of iSNV detected was weakly related to sequencing depth with an adjusted R^2 of 0.04 (S2 Fig, $t=5.201$, $p=2.77 \times 10^{-7}$).

iSNV dynamics

The allele frequencies of identified iSNV were generally very low, with the majority of iSNV present at $\leq 2\%$ frequency (Fig 2A). In our cohort, the frequencies of iSNV in vaccinated individuals (median = 0.0151) were higher than in unvaccinated individuals (median = 0.0127, $p=0.022$, S2 Table), but this difference was extremely small and unlikely to be biologically significant (S3 Fig). Frequencies of iSNV also varied by the day of sampling ($p=0.002$, S3 Fig, S2 Table) but did not differ based on host age, SARS-CoV-2 clade, mutation type (i.e., nonsynonymous vs. synonymous; S3 Fig), or viral load (S4 Fig and S3 Table).

All specimens had between 0–12 iSNV identified at an allele frequency $\geq 0.5\%$ (Fig 2B). Unvaccinated individuals (median = 2.61, $p<0.001$) and children (median = 2, $p=0.011$) had greater numbers of iSNV per specimen than vaccinated individuals (median = 1.82) and adults (median = 1, Fig 2C and 2D, S2 Table). BA.1 had fewer iSNV per specimen

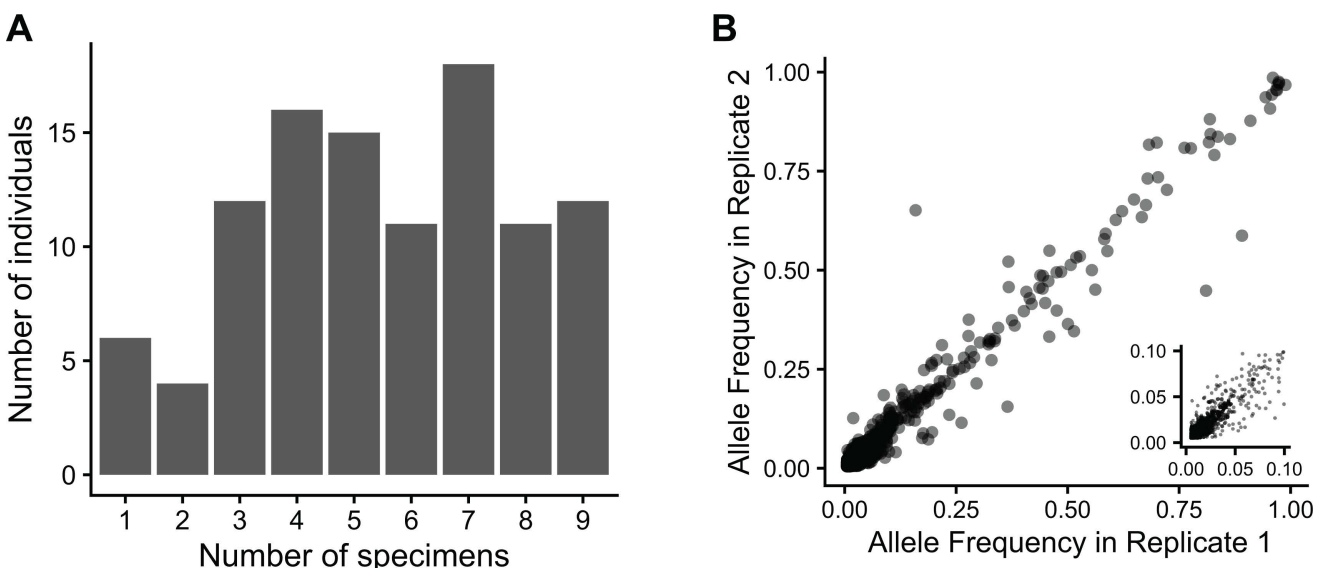


Fig 1. Sequencing summary. (A) The number of specimens per person successfully sequenced. (B) intra-host single nucleotide variants (iSNV) frequency is consistent across replicates for iSNV that pass quality control filtering. The insert shows iSNV frequency up to 0.1.

<https://doi.org/10.1371/journal.ppat.1013134.g001>

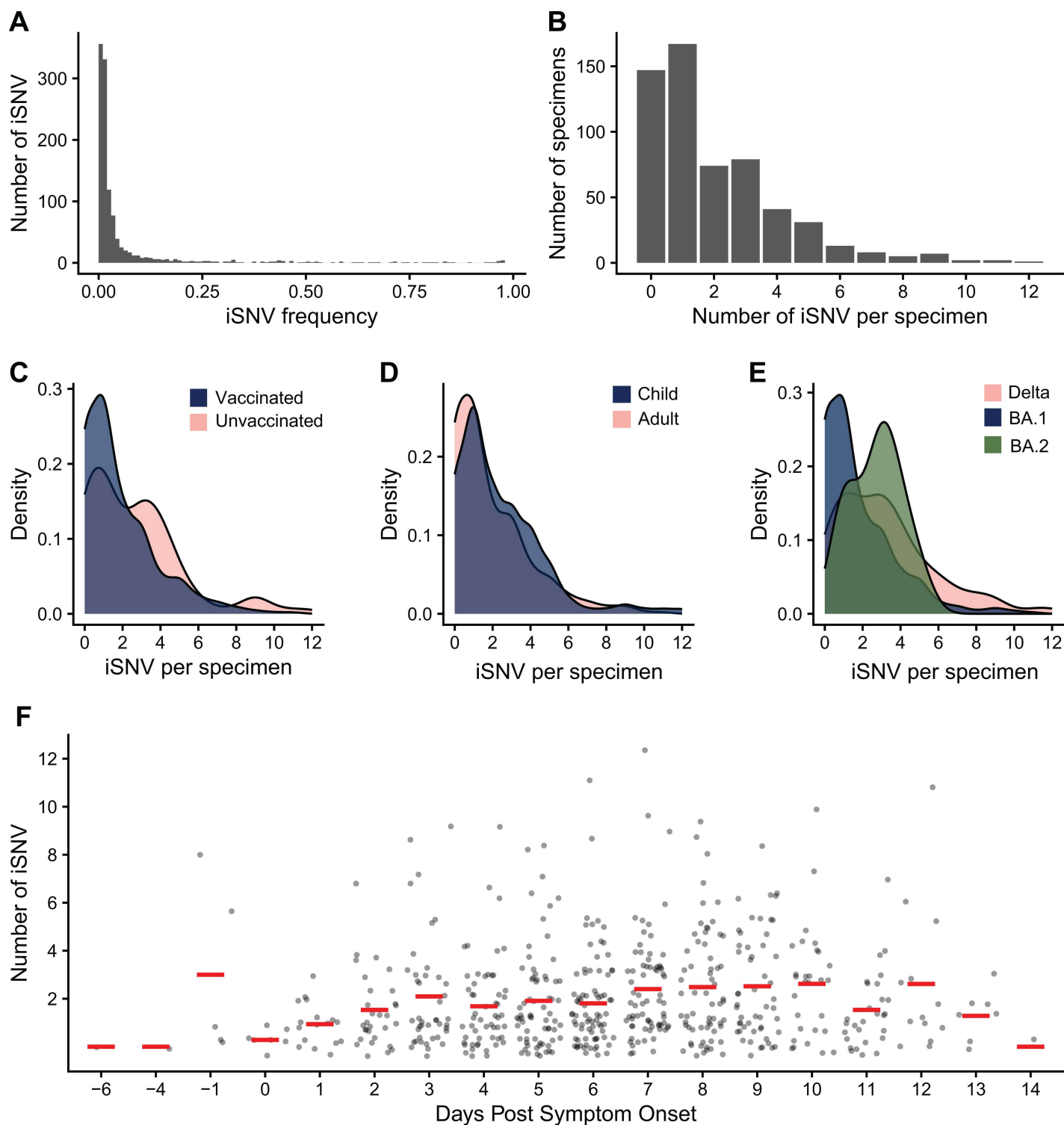


Fig 2. Intrahost single nucleotide variant dynamics. (A) iSNV frequency. (B) The number of iSNV per specimen. The number of iSNV per specimen by (C) vaccination status, (D) age with child <18 and adult 18+, (E) clade, (F) and days post symptom onset. The red lines are the mean. Vaccinated individuals and BA.1 infections had fewer iSNV per specimen than unvaccinated individuals and Delta or BA.2 infections.

<https://doi.org/10.1371/journal.ppat.1013134.g002>

(median = 1, $p < 0.001$) than BA.2 (median = 3, $p = 0.033$) or Delta (median = 3, $p < 0.001$) infections ([Fig 2E](#), [S2 Table](#)). The number of iSNV per specimen increased as the infection progressed, and after 8–10 days post symptom onset, the number of iSNV decreased ($p = 0.005$, [Fig 2F](#), [S2 Table](#)). The time of sampling (days post symptom onset) did not noticeably differ by vaccination status, age, or clade ([S5 Fig](#)). The number of iSNV did not differ based on viral load ([S4 Fig](#) and [S3 Table](#)).

Within-host divergence rates

We estimated within-host evolutionary rates as nucleotide divergence per site per day on a per-specimen basis and by linear regression in individuals for whom we had multiple sequenced specimens. The genome-wide mean divergence rate was 5.03×10^{-7} nucleotide substitutions/site/day for nonsynonymous mutations and 1.08×10^{-6} for synonymous mutations. Although not statistically significant, the estimated divergence rate varied according to the day of sampling when using a per specimen estimate ([Fig 3](#)). The divergence rate increased from the onset of the infection until approximately day 5 for nonsynonymous sites and day 8 for synonymous sites and then decreased. The divergence rate wasn't related to viral load ([S4 Fig](#) and [S3 Table](#)).

For the rest of the comparisons using a per specimen estimate, the divergence rate from the specimen with the highest viral load was used. Children had higher rates for nonsynonymous mutations (mean = 1.11×10^{-6} vs 3.01×10^{-7}), but not synonymous mutations ($p = 0.019$, [Fig 3C](#), [S4 Table](#)), while rates for synonymous mutations were not associated with age. The divergence rate did not differ by vaccination status or clade ([Fig 3](#), [S4 Table](#)). There were significant differences in divergence rate based on gene ($p < 0.001$); notably, spike (mean = 9.96×10^{-7}) had a higher divergence rate compared to ORF1a (mean = 4.81×10^{-7}) for nonsynonymous mutations, but did not differ from any of the other genes ([Fig 3E](#) and [3F](#), [S5 Table](#)). Results obtained by linear regression were slightly different. The divergence rate did not differ by vaccination, age, clade, or gene ([S6 Fig](#), [S4 Table](#)). Divergence rate varied by gene for synonymous mutations ($p < 0.001$, [S6 Fig](#), [S4](#) and [S6 Tables](#)). ORF1b (mean = 2.27×10^{-6}) had a higher divergence rate than N (mean = -2.62×10^{-6}) and ORF8 (mean = -3.55×10^{-7}). The negative estimates were due to times when specimens were collected predominantly after the peak of viral load, in a contracting population.

Analysis of selection

We analyzed selection by first looking for iSNV that anticipated mutations that defined subsequent variants. Many lineage defining mutations are immune escape variants, and within-host selection for new antigenic variants may precede immune selection detectable at the population level as lineage defining mutations of subsequent variants. Two individuals with BA.1 had an iSNV that causes S:371F, a BA.2 lineage defining mutation ([Table 1](#)). These iSNV were at low frequencies, with a maximum observed frequency of 0.8% and 1.8%. There were 3 additional iSNV in the codon for a lineage defining mutation but resulted in a different amino acid substitution. This included a third iSNV at position 371.

Using a WFABC model, we estimated a within-host effective population size of 78 resulting in strong genetic drift. Fourteen iSNV from 11 individuals were under positive selection: 7 in spike, 6 in other coding regions and 1 in a non-coding region ([Fig 4A](#) and [4B](#), [Table 2](#)). The results were the same for 8hr and 12hr generation times. Of the iSNV found in coding regions, 10 were nonsynonymous, including 6 of the iSNV in spike. Two of the selected synonymous iSNV were in individuals that had nonsynonymous iSNV under positive selection, suggestive of linkage as the allele trajectories of the two iSNV were closely matched ([Fig 4C](#) and [4D](#)). Outside of spike ORF7a:T14I and ORF1a:S318L were found in 1 additional individual each. Both individuals only had a single sample and therefore no selection coefficient.

Three of the selected spike amino acid substitutions were in the RBD (Receptor Binding Domain). Outside of the RBD, two individuals shared the positively selected substitution, S:D574N. A third individual had S:D574N in 4 specimens, yet without a positive selection coefficient. None of the iSNV in future lineage defining codons had a positive selection

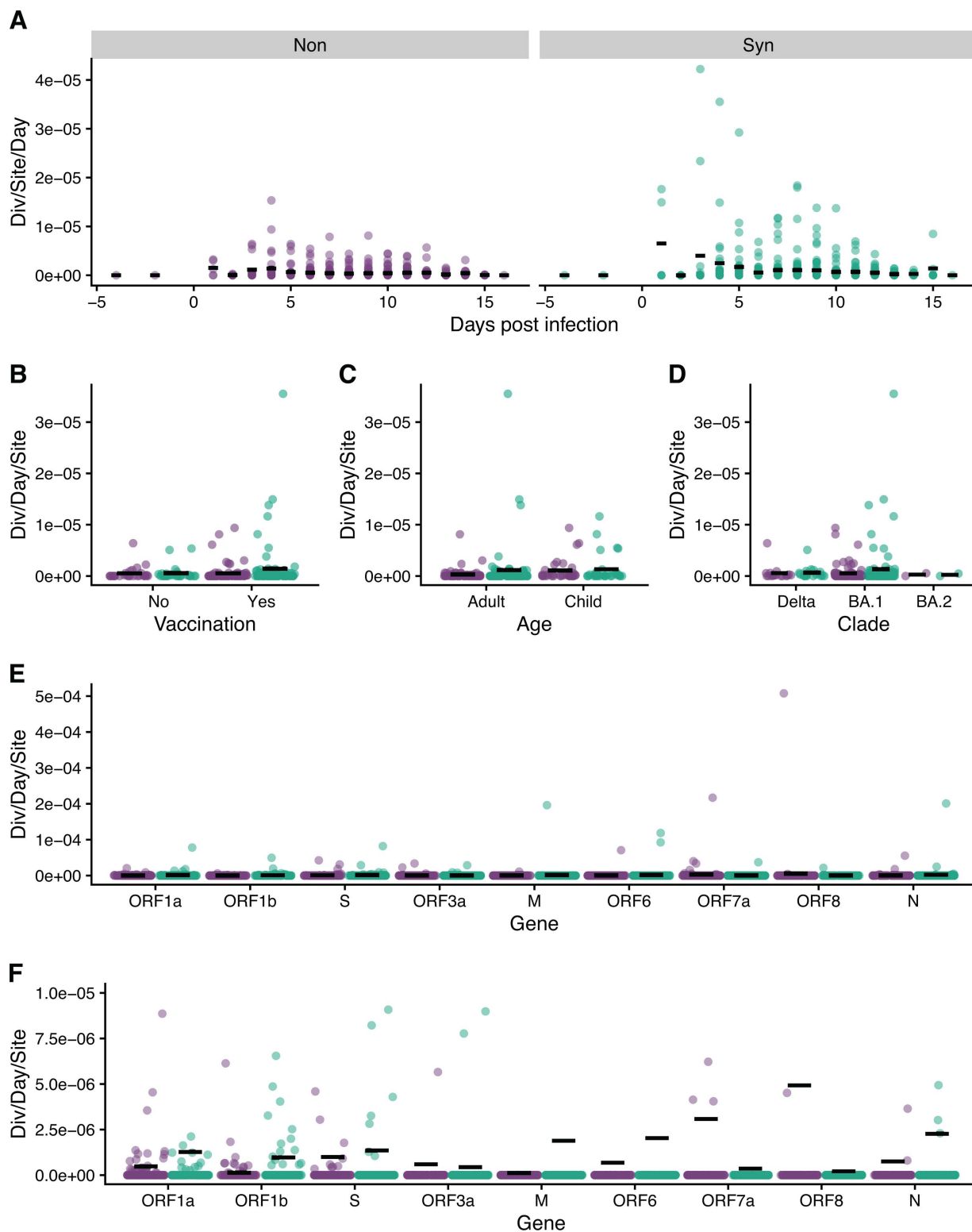


Fig 3. Divergence rates using a per specimen estimate. (A) Divergence rate (divergence/site/day) for all specimens by days post infection. Divergence rate (divergence/site/day) using the specimen with the highest viral titer by (B) vaccination status, (C) age, (D) clade, and (E) gene. (F) is a zoomed in version of (E), note y-axis. For nonsynonymous mutations children had higher rates than adults, and spike had a higher rate than ORF1a. Black lines are the mean divergence rate. Green is synonymous, and purple is nonsynonymous.

<https://doi.org/10.1371/journal.ppat.1013134.g003>

Table 1. iSNV that are in the same position as lineage defining mutations in spike for the subsequent variant wave.

Clade	Next Wave	Lineage Defining Mut.	Observed Mutation	Individual	Max Obs. Frequency
Delta	BA.1	N440K	N440Y	102101	0.049
Delta	BA.1	G446S	G446V	101201	0.018
Delta	BA.1	T547K	T547I	101703	0.008
BA.1	BA.2	S371F	L371I	102601	0.008
BA.1	BA.2	S371F	L371F	105701	0.018
BA.1	BA.2	S371F	L371F	107303	0.008

<https://doi.org/10.1371/journal.ppat.1013134.t001>

coefficient. However, one individual had both an iSNV in a lineage defining codon (S:547) and an iSNV with a positive selection coefficient in the viral replicase (ORF8:S54L). All of the nonsynonymous spike iSNV were in vaccinated individuals.

We used the SARS-CoV-2 nextstrain build to determine whether any of the iSNV with positive selection coefficients were also identified as increasing in global frequency [37]. None of the iSNV or resulting amino acid changes reached more than 5% globally. The selection coefficients we estimated were only weakly related to the between host selection coefficients estimated by Bloom and Neher [38] (S7 Fig).

Discussion

In this intensive evaluation of serially sampled individuals in a longitudinal household transmission study, we found that within-host SARS-CoV-2 populations are dominated by purifying selection and genetic drift. This results in low levels of diversity and low rates of divergence, consistent with previous studies [8–10,39,40]. There were differences in divergence rate based on age and in the frequency of iSNV based on vaccination, but these are unlikely to be biologically significant and are not necessarily causative. The lack of biologically meaningful differences between vaccinated and unvaccinated individuals may partially be the result of prior infection in the unvaccinated group. Multiple factors influenced the number of iSNV per specimen, notably day of sampling. Positive selection was rare, but when present, it tended to be enriched in key areas of spike and the RBD.

The low level of diversity is similar to what we and others have reported for SARS-CoV-2 [8–10,39,40]. Some study-specific differences in diversity are noteworthy. For example, Farjo *et al.* (with specimens from 40 individuals) observed higher numbers of iSNV in vaccinated individuals, while we found higher numbers of iSNV in unvaccinated individuals [11]. However, their quality metrics differed between vaccinated and unvaccinated individuals, and their sample size was smaller than the present study. Additionally Gu *et al.* found that the number of iSNV per specimen was higher in VOC compared to non-VOC clades, but did not find any differences between VOC clades [12]. In contrast, we found Delta and BA.2 had more iSNV than BA.1. Our frequency threshold for variant calling was lower, and potentially more sensitive to differences in iSNV number. Variation between cohorts likely contributes to differences between studies, but different study designs and methods also account for dissimilarities.

SARS-CoV-2 has comparable within-host dynamics to influenza A virus. The distribution of allele frequencies is very similar in influenza A and SARS-CoV-2, with most iSNV found at very low frequencies [24,41,42]. However, compared to studies of influenza with the same iSNV threshold, SARS-CoV-2 had fewer iSNV per specimen despite the genome being twice the size. SARS-CoV-2 also had lower divergence rates of 10^{-6} div/site/day for synonymous sites and 10^{-7} for nonsynonymous sites, compared to 10^{-5} and 10^{-6} for influenza A in synonymous and nonsynonymous sites respectively [24,41]. The lower within-host diversity of SARS-CoV-2 is largely attributable to the difference in mutation rates. With its proofreading capabilities, SARS-CoV-2 has a mutation rate of 9×10^{-7} mutations per nucleotide per replication cycle [43] compared to 2×10^{-6} in influenza A (using analogous assays) [44]. The strength of genetic drift may also contribute to the observed

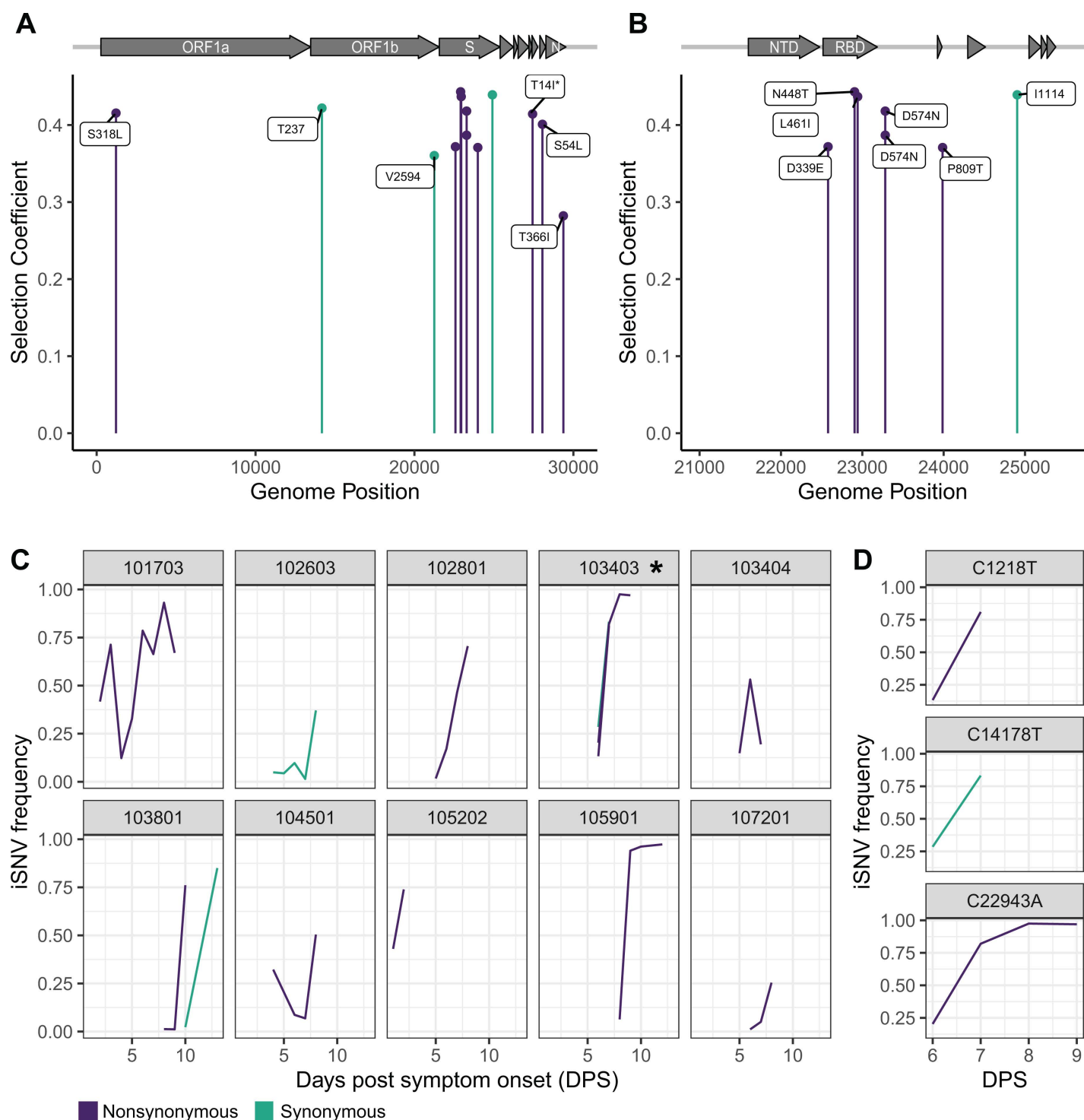


Fig 4. iSNV under positive selection and their corresponding allele trajectories. (A) WFABC selection coefficients for iSNV under positive selection for the whole genome and (B) for spike. (C) The allele trajectories of the iSNV with positive selection coefficients by individual. (D) The allele trajectories separated by locus for iSNV in individual 103403, denoted with an asterisk in (C). Green is synonymous, and purple is nonsynonymous. WFABC = Wright Fisher Approximate Bayesian Computation; iSNV = intra-host single nucleotide variants.

<https://doi.org/10.1371/journal.ppat.1013134.g004>

Table 2. iSNV with a positive selection coefficient. Selection coefficient values are from the 8hr generation time results.

Gene	iSNV	AA Mutation	Mutation Type	Selection Coefficient	Individual	Age	Clade	Vaccinated
ORF1a	C1218T	S318L	Non	0.4155	103403	42	BA.1	Yes
ORF1b	C14178T	237T	Syn	0.4220	103403	42	BA.1	Yes
ORF1b	G21249T	2594V	Syn	0.3603	102603	46	BA.1	Yes
S	T22579A	D339E	Non	0.3718	105202	54	BA.1	Yes
S	A22905C	L461I	Non	0.4369	103403	42	BA.1	Yes
S	C22943A	N448T	Non	0.4431	105901	49	BA.1	Yes
S	G23282A	D574N	Non	0.3868	103404	36	BA.1	Yes
S	G23282A	D574N	Non	0.4181	102801	39	BA.1	Yes
S	C23987A	P809T	Non	0.3708	107201	50	BA.1	Yes
S	C24904T	I114I	Syn	0.4392	103801	47	BA.1	No
ORF7a	C27434T	T14I	Non	0.4143	103801	47	BA.1	No
ORF8	C28054T	S54L	Non	0.4009	101703	10	Delta	No
N	C29370T	T366I	Non	0.2828	104501	10	BA.1	Yes
NA	G29742A	NA	NA	0.3544	106103	36	BA.1	Yes

<https://doi.org/10.1371/journal.ppat.1013134.t002>

differences. While both influenza A virus ($N_e \sim 150\text{--}300$) [41,42] and SARS-CoV-2 have small effective population sizes, the smaller effective population size in SARS-CoV-2 will result in more genetic drift. More of the variation will be lost from the population or not repeatedly sampled due to changes in population structure. These within-host dynamics are largely consistent with the neutral theory of evolution [45]. Strongly deleterious mutations are removed quickly from the population and the remaining variation is largely neutral.

Despite overall similar patterns of within-host dynamics between SARS-CoV-2 and influenza A virus, there are differences in the nature of selected sites. In influenza A virus, we have not found an overrepresentation of selected sites in hemagglutinin (HA), including antigenic sites, or in neuraminidase (NA) [41]. In contrast, in SARS-CoV-2 we found a greater number of positively selected sites in spike (7/13) and in the RBD (3) than expected by chance. This is consistent with selection for immune escape. Within the RBD, S:D339E was under positive selection. Although this exact amino acid substitution has not previously been known to be under selection, S:339 is the most variable amino acid in spike [46]. Additionally, G339D is a lineage defining mutation in BA.1, BA.2, BA.4, and BA.5 [47], and D339H is a lineage defining mutation for BA.2.75, XBB, and BA.2.86 [48,49]. Both of these amino acid substitutions have been shown to escape neutralizing antibodies [50,51].

In the RBD, S:448 is an epitope targeted by multiple monoclonal antibodies, including bebtelovimab, imdevimab, and cilgavimab [47]. These monoclonal antibodies have high similarity to germline encoded antibodies [52–54], making S:448 an epitope that is likely to be commonly targeted across individuals. Outside of the RBD, two individuals in different households had D574N under positive selection. This substitution has been observed in a long-term infection of an immunocompromised patient [55] and also detected in a small proportion of BA.5 lineages. Mutations at sites 371, 339, and 574 have all been shown to affect the propensity of the RBD to adopt a down versus up conformation, which can reduce neutralization by polyclonal serum antibodies by reducing antibody binding to RBD epitopes that are only accessible in the up-RBD conformation [56–58].

This infrequent but detectable positive selection may be due to the timing of these infections relative to viral emergence. This study enrolled individuals within approximately the first 18–24 months of the pandemic. At this time, only the Wuhan strain spike was used for vaccination, leading to a relatively narrow antibody repertoire. A narrow antibody repertoire may cause uniform selection pressure, with one or a few mutations being sufficient for SARS-CoV-2 to be resistant to

a majority of the host antibodies, similar to treatment with monoclonal antibodies [52,59]. In our study, six of the selected sites in spike, all of the nonsynonymous sites, and all of the selected sites in the RBD occurred in vaccinated individuals. Over time as the number of exposures and lineages individuals are exposed to increases, their antibody repertoires also increase [60,61]. As the antibody repertoire diversifies, individual mutations may make SARS-CoV-2 resistant to only a small proportion of antibodies, leading to weaker selection [61]. Earlier in the pandemic there may have been low levels of selection due to lack of even partial immunity, coinciding with a period of global evolutionary stasis [43].

Despite finding immunologically relevant iSNV, our results had low predictive power for trends in SARS-CoV-2 evolution globally. None of the iSNV under positive selection or the corresponding amino acid substitutions reached >5% frequency globally at any time. Two individuals with BA.1 infections had a lineage defining mutation, S:371F, for BA.2. However, the mutation remained at very low frequencies within these two individuals. In the first individual, the selection coefficient was not statistically significantly different than 0 ($s = -0.07$), and a selection coefficient was unable to be calculated for the second individual due to the number of specimens. With low effective population sizes and stochastic dynamics, our estimates of positive selection are conservative. However, combining within-host variant data with other sources (e.g., deep mutational scanning or inferred between-host selection coefficients) may be fruitful for understanding the evolutionary trajectory of SARS-CoV-2.

A major strength of this study is daily sampling, with up to 9 successfully sequenced specimens per individual, allowing us to examine allele trajectories. Summary statistics meant to detect selection can be misleading due to genetic linkage and hitchhiking [13]. These effects are especially prominent in cases where there are strong bottlenecks and low levels of recombination. With serial sampling, we were able to calculate selection coefficients and detect hitchhiking of synonymous mutations with a physically linked nonsynonymous mutation. To illustrate, in one individual, there were three iSNV with nearly identical allele trajectories: 2 nonsynonymous and 1 synonymous. Most likely, the nonsynonymous iSNV in ORF1a and the synonymous iSNV in ORF1b were swept along with the nonsynonymous iSNV in spike (L461I).

Our study has several limitations. First, our results may not generalize to other phases of the SARS-CoV-2 pandemic. The study took place over 6 months in the second year of the pandemic after the availability of a single vaccine formulation. Results may differ as vaccine and exposure history become more variable across the population and as SARS-CoV-2 has had a longer evolutionary history with human hosts. Indeed, we speculate that SARS-CoV-2 evolution during acute infections could become more similar to the dynamics of influenza A virus within hosts [41,42]. Second, there is always the possibility of inaccurate variant calls. However, this possibility was mitigated by sequencing all specimens in replicate and sequencing multiple specimens per person reduces this possibility. Third, SARS-CoV-2 has significant compartmentalization [62,63], and we are only sampling one location in the body; but when compared, nasal and saliva specimens have similar within-host dynamics dominated by stochastic processes [11]. Fourth, the bias towards symptomatic index cases may reduce the generalizability. Although contact cases had both symptomatic and asymptomatic infections included.

Across studies, acute respiratory viruses have similar within-host dynamics: tight bottlenecks, low genetic diversity, and populations dominated by purifying selection and genetic drift [8–12,19,39–42,64,65]. Overall, our findings are consistent with this pattern. However, nuanced differences exist between viruses, cohorts, and demographic features. In our cohort, within-host positive selection was rare, but appeared to frequently be immune mediated when present. As viruses adapt to human hosts and the population develops immunity, it will be important to follow the shifting impacts on within-host dynamics and selective pressure.

Supporting information

S1 Table. Demographic information and infection details for individuals in this study (n=105).

(PDF)

S2 Table. Comparisons of the number of iSNV per specimen and of iSNV frequency. For statistically significant differences the p values are bolded. χ^2 test statistics are from Kruskal-Wallis rank sum tests and W test statistics are from Mann-Whitney U tests.

(PDF)

S3 Table. Effects of viral load. Statistically significant differences are bolded. Z test statistics are from comparisons of iSNV number and iSNV frequency. T test statistics are from comparisons with divergence rate.

(PDF)

S4 Table. Comparisons of divergence rates. Statistically significant differences are bolded. χ^2 test statistics are from Kruskal-Wallis rank sum tests, and W test statistics are from Mann-Whitney U tests.

(PDF)

S5 Table. Post hoc (Dunn) tests for divergence rate between genes using a per specimen estimate. Statistically significant differences are bolded for the adjusted P values.

(PDF)

S6 Table. Post hoc (Dunn) tests for divergence rate of synonymous mutations between genes using linear regressions. Statistically significant differences are bolded for the adjusted P values.

(PDF)

S1 Fig. Trajectories of viral load by individual over the course of their infection.

(PDF)

S2 Fig. Sequencing coverage (A) Boxplots of coverage across the genome in non-overlapping windows of 400 bp for specimens with high quality sequencing. The box shows the first quartile, median, and third quartile. The whiskers are 1.5x interquartile range, and the dots are the outliers. **(B)** The number of iSNV per sample by average sequencing depth. iSNV = intra-host single nucleotide variants.

(PDF)

S3 Fig. iSNV frequency. (A) mutation type, **(B)** vaccination status, **(C)** age with child <18 and adult 18+, **(D)** clade, and **(E)** days post symptom onset. The red lines are the mean. iSNV = intra-host single nucleotide variants.

(PDF)

S4 Fig. Viral load and iSNV dynamics. Effects of viral load on **(A)** iSNV frequency, **(B)** iSNV number per specimen, and **(C)** divergence rates. Green is synonymous, and purple is nonsynonymous. iSNV = intra-host single nucleotide variants.

(PDF)

S5 Fig. Number of specimens collected per day post symptom onset. (A) vaccination status, **(B)** age with child <18 and adult 18+, and **(C)** clade.

(PDF)

S6 Fig. Divergence rate (divergence/site/day) using linear regressions. (A) vaccination status, **(B)** age with child <18 and adult 18+, **(C)** clade, and **(D)** gene (green synonymous, purple nonsynonymous).

(PDF)

S7 Fig. Comparison of the within-host selection coefficient and the population level selection coefficient for Bloom & Neher 2023. Green is synonymous and purple is nonsynonymous. Triangles are mutations in spike, and circles are in non-spike genes.

(PDF)

Acknowledgments

We thank all participants in the study for their time and effort and Jesse Bloom for critical comments on the preprint version of the manuscript. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC).

Author contributions

Conceptualization: H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Data curation: Emily E. Bendall, Derek Dimcheff, Leigh Papalambros, William J. Fitzsimmons, Yuwei Zhu, Jonathan Schmitz, Natasha Halasa, James Chappell, Emily T. Martin, Jessica E. Biddle, Sarah E. Smith-Jeffcoat, Melissa A. Rolfes, Alexandra Mellis, H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Formal analysis: Emily E. Bendall, Yuwei Zhu, Adam S. Luring.

Funding acquisition: Emily T. Martin, H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Investigation: Emily E. Bendall, Derek Dimcheff, Leigh Papalambros, William J. Fitzsimmons, Yuwei Zhu, Jonathan Schmitz, Natasha Halasa, James Chappell, Emily T. Martin, Jessica E. Biddle, Sarah E. Smith-Jeffcoat, Melissa A. Rolfes, Alexandra Mellis, H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Methodology: Emily E. Bendall, Derek Dimcheff, Leigh Papalambros, William J. Fitzsimmons, Yuwei Zhu, Jonathan Schmitz, Natasha Halasa, James Chappell, Emily T. Martin, Jessica E. Biddle, Sarah E. Smith-Jeffcoat, Melissa A. Rolfes, Alexandra Mellis, H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Project administration: H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Resources: Emily E. Bendall, Leigh Papalambros, William J. Fitzsimmons, Yuwei Zhu, Jonathan Schmitz, Natasha Halasa, James Chappell, Emily T. Martin, Jessica E. Biddle, Sarah E. Smith-Jeffcoat, Melissa A. Rolfes, Alexandra Mellis, H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

Writing – original draft: Emily E. Bendall, Adam S. Luring.

Writing – review & editing: Emily E. Bendall, Derek Dimcheff, Leigh Papalambros, William J. Fitzsimmons, Yuwei Zhu, Jonathan Schmitz, Natasha Halasa, James Chappell, Emily T. Martin, Jessica E. Biddle, Sarah E. Smith-Jeffcoat, Melissa A. Rolfes, Alexandra Mellis, H. Keipp Talbot, Carlos Grijalva, Adam S. Luring.

References

1. Carabelli AM, Peacock TP, Thorne LG, Harvey WT, Hughes J, COVID-19 Genomics UK Consortium, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol*. 2023;21(3):162–77. <https://doi.org/10.1038/s41579-022-00841-7> PMID: 36653446
2. Chaudhari AM, Joshi M, Kumar D, Patel A, Lokhande KB, Krishnan A, et al. Evaluation of immune evasion in SARS-CoV-2 Delta and Omicron variants. *Comput Struct Biotechnol J*. 2022;20:4501–16. <https://doi.org/10.1016/j.csbj.2022.08.010> PMID: 35965661
3. Pouwels KB, Pritchard E, Matthews PC, Stoesser N, Eyre DW, Vihta K-D, et al. Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nat Med*. 2021;27(12):2127–35. <https://doi.org/10.1038/s41591-021-01548-7> PMID: 34650248
4. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *N Engl J Med*. 2022;386(16):1532–46. <https://doi.org/10.1056/NEJMoa2119451> PMID: 35249272
5. Marks P. Fall 2022 COVID-19 Vaccine Strain Composition Selection Recommendation. Federal Drug Administration; 2022. Available: <https://www.fda.gov/media/159597/download?attachment>
6. Luo S, Reed M, Mattingly JC, Koelle K. The impact of host immune status on the within-host and population dynamics of antigenic immune escape. *J R Soc Interface*. 2012;9(75):2603–13. <https://doi.org/10.1098/rsif.2012.0180> PMID: 22572027
7. Volkov I, Pepin KM, Lloyd-Smith JO, Banavar JR, Grenfell BT. Synthesizing within-host and population-level selective pressures on viral populations: the impact of adaptive immunity on viral immune escape. *J R Soc Interface*. 2010;7(50):1311–8. <https://doi.org/10.1098/rsif.2009.0560> PMID: 20335194

8. Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, et al. SARS-CoV-2 within-host diversity and transmission. *Science*. 2021;372(6539). <https://doi.org/10.1126/science.abg0821>
9. Valesano AL, Rumpf KE, Dimcheff DE, Blair CN, Fitzsimmons WJ, Petrie JG, et al. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *PLoS Pathog*. 2021;17(4):e1009499. <https://doi.org/10.1371/journal.ppat.1009499> PMID: 33826681
10. Tonkin-Hill G, Martincorena I, Amato R, Lawson ARJ, Gerstung M, Johnston I, et al. Patterns of within-host genetic diversity in SARS-CoV-2. *Elife*. 2021;10:e66857. <https://doi.org/10.7554/eLife.66857> PMID: 34387545
11. Farjo M, Koelle K, Martin MA, Gibson LL, Walden KKO, Rendon G, et al. Within-host evolutionary dynamics and tissue compartmentalization during acute SARS-CoV-2 infection. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.06.21.497047>
12. Gu H, Quadeer AA, Krishnan P, Ng DYM, Chang LDJ, Liu GYZ, et al. Within-host genetic diversity of SARS-CoV-2 lineages in unvaccinated and vaccinated individuals. *Nat Commun*. 2023;14(1):1793. <https://doi.org/10.1038/s41467-023-37468-y> PMID: 37002233
13. Soni V, Terbot JW 2nd, Jensen JD. Population genetic considerations regarding the interpretation of within-patient SARS-CoV-2 polymorphism data. *Nat Commun*. 2024;15(1):3240. <https://doi.org/10.1038/s41467-024-46261-4> PMID: 38627371
14. Riddell AC, Kele B, Harris K, Bible J, Murphy M, Dakshina S, et al. Generation of Novel Severe Acute Respiratory Syndrome Coronavirus 2 Variants on the B.1.1.7 Lineage in 3 Patients With Advanced Human Immunodeficiency Virus-1 Disease. *Clin Infect Dis*. 2022;75(11):2016–8. <https://doi.org/10.1093/cid/ciac409> PMID: 35616095
15. Scherer EM, Babiker A, Adelman MW, Allman B, Key A, Kleinhenz JM, et al. SARS-CoV-2 Evolution and Immune Escape in Immunocompromised Patients. *N Engl J Med*. 2022;386(25):2436–8. <https://doi.org/10.1056/NEJMc2202861> PMID: 35675197
16. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021;592(7853):277–82. <https://doi.org/10.1038/s41586-021-03291-y> PMID: 33545711
17. Wilkinson SAJ, Richter A, Casey A, Osman H, Mirza JD, Stockton J, et al. Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol*. 2022;8(2):veac050. <https://doi.org/10.1093/ve/veac050> PMID: 35996593
18. Weigang S, Fuchs J, Zimmer G, Schnepf D, Kern L, Beer J, et al. Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants. *Nat Commun*. 2021;12(1):6405. <https://doi.org/10.1038/s41467-021-26602-3> PMID: 34737266
19. McCrone JT, Luring AS. Genetic bottlenecks in intraspecies virus transmission. *Curr Opin Virol*. 2018;28:20–5. <https://doi.org/10.1016/j.coviro.2017.10.008> PMID: 29107838
20. Focosi D, Maggi F, Franchini M, McConnell S, Casadevall A. Analysis of Immune Escape Variants from Antibody-Based Therapeutics against COVID-19: A Systematic Review. *Int J Mol Sci*. 2021;23(1):29. <https://doi.org/10.3390/ijms23010029> PMID: 35008446
21. RVTN-Sentinel-Protocol-20220110. 2021. Available: <https://archive.cdc.gov/#/details?url=https://www.cdc.gov/vaccines/covid-19/downloads/RVTN-Sentinel-Protocol-20220110.pdf>
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
23. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. 2019;20(1):8. <https://doi.org/10.1186/s13059-018-1618-7> PMID: 30621750
24. Xue KS, Bloom JD. Linking influenza virus evolution within and between human hosts. *Virus Evol*. 2020;6(1):veaa010. <https://doi.org/10.1093/ve/veaa010> PMID: 32082616
25. De Maio N, Walker C, Borges R, Weilguny L, Slodkiewicz G, Goldman Ni. Issues with SARS-CoV-2 sequencing data. *Virological*. 2020. [cited 6 Oct 2022]. Available.
26. Aksamentov I, Roemer C, Hodcroft E, Neher R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *JOSS*. 2021;6(67):3773. <https://doi.org/10.21105/joss.03773>
27. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;5(11):1403–7. <https://doi.org/10.1038/s41564-020-0770-5> PMID: 32669681
28. 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 pangolin tool. *Virus Evol*. 2021;7(2):veab064. <https://doi.org/10.1093/ve/veab064> PMID: 34527285
29. Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. Kinetics of Influenza A Virus Infection in Humans. *J Virol*. 2006;80(15):7590–9. <https://doi.org/10.1128/jvi.01623-05>
30. Beauchemin CAA, Handel A. A review of mathematical models of influenza A infections within a host or cell culture: lessons learned and challenges ahead. *BMC Public Health*. 2011;11 Suppl 1(Suppl 1):S7. <https://doi.org/10.1186/1471-2458-11-S1-S7> PMID: 21356136
31. Carrat F, Vergu E, Ferguson NM, Lemaître M, Cauchemez S, Leach S, et al. Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies. *American J Epidemiology*. 2008;167(7):775–85. <https://doi.org/10.1093/aje/kwm375>
32. Foll M, Shim H, Jensen JD. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol Ecol Resour*. 2015;15(1):87–98. <https://doi.org/10.1111/1755-0998.12280> PMID: 24834845
33. Schneider M, Ackermann K, Stuart M, Wex C, Protzer U, Schätzl HM, et al. Severe acute respiratory syndrome coronavirus replication is severely impaired by MG132 due to proteasome-independent inhibition of M-calpain. *J Virol*. 2012;86(18):10112–22. <https://doi.org/10.1128/JVI.01001-12> PMID: 22787216

34. Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. *Elife*. 2020;9:e57309. <https://doi.org/10.7554/eLife.57309> PMID: [32228860](https://pubmed.ncbi.nlm.nih.gov/32228860/)
35. Harcourt J, Tamin A, Lu X, Kamili S, Sakthivel SK, Murray J, et al. Isolation and characterization of SARS-CoV-2 from the first US COVID-19 patient. *bioRxiv*. 2020:2020.03.02.972935. <https://doi.org/10.1101/2020.03.02.972935> PMID: [32511316](https://pubmed.ncbi.nlm.nih.gov/32511316/)
36. Smith BJ. boa: AnRPackage for MCMC Output Convergence Assessment and Posterior Inference. *J Stat Soft*. 2007;21(11). <https://doi.org/10.18637/jss.v021.i11>
37. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121–3. <https://doi.org/10.1093/bioinformatics/bty407> PMID: [29790939](https://pubmed.ncbi.nlm.nih.gov/29790939/)
38. Bloom JD, Neher RA. Fitness effects of mutations to SARS-CoV-2 proteins. *Virus Evol*. 2023;9: vead055. <https://doi.org/10.1093/ve/vead055>
39. Bendall EE, Callear AP, Getz A, Goforth K, Edwards D, Monto AS, et al. Rapid transmission and tight bottlenecks constrain the evolution of highly transmissible SARS-CoV-2 variants. *Nat Commun*. 2023;14(1):272. <https://doi.org/10.1038/s41467-023-36001-5> PMID: [36650162](https://pubmed.ncbi.nlm.nih.gov/36650162/)
40. Braun K, Moreno G, Wagner C, Accola MA, Rehrauer WM, Baker D, et al. Limited within-host diversity and tight transmission bottlenecks limit SARS-CoV-2 evolution in acutely infected individuals. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.04.30.440988>
41. Bendall EE, Zhu Y, Fitzsimmons WJ, Rolfes M, Mellis A, Halasa N, et al. Influenza A virus within-host evolution and positive selection in a densely sampled household cohort over three seasons. *bioRxiv*. 2024:2024.08.15.608152. <https://doi.org/10.1101/2024.08.15.608152> PMID: [39229225](https://pubmed.ncbi.nlm.nih.gov/39229225/)
42. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Luring AS. Stochastic processes constrain the within and between host evolution of influenza virus. *Elife*. 2018;7:e35962. <https://doi.org/10.7554/eLife.35962> PMID: [29683424](https://pubmed.ncbi.nlm.nih.gov/29683424/)
43. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. *Nat Rev Microbiol*. 2023;21(6):361–79. <https://doi.org/10.1038/s41579-023-00878-2> PMID: [37020110](https://pubmed.ncbi.nlm.nih.gov/37020110/)
44. Nobusawa E, Sato K. Comparison of the mutation rates of human influenza A and B viruses. *J Virol*. 2006;80(7):3675–8. <https://doi.org/10.1128/JVI.80.7.3675-3678.2006> PMID: [16537638](https://pubmed.ncbi.nlm.nih.gov/16537638/)
45. Kimura M. The neutral theory of molecular evolution. *Sci Am*. 1979;241(5):98–100, 102, 108 passim. <https://doi.org/10.1038/scientificameri-can1179-98> PMID: [504979](https://pubmed.ncbi.nlm.nih.gov/504979/)
46. Guruprasad L, Naresh GK, Boggarapu G. Taking stock of the mutations in human SARS-CoV-2 spike proteins: From early days to nearly the end of COVID-19 pandemic. *Curr Res Struct Biol*. 2023;6:100107. <https://doi.org/10.1016/j.crstbi.2023.100107> PMID: [37841365](https://pubmed.ncbi.nlm.nih.gov/37841365/)
47. Cox M, Peacock TP, Harvey WT, Hughes J, Wright DW, COVID-19 Genomics UK (COG-UK) Consortium, et al. SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro studies. *Nat Rev Microbiol*. 2023;21(2):112–24. <https://doi.org/10.1038/s41579-022-00809-7> PMID: [36307535](https://pubmed.ncbi.nlm.nih.gov/36307535/)
48. Escalera-Zamudio M, Tan CCS, van Dorp L, Balloux F. Early evolution of BA.2.86 sheds light on the origins of highly divergent SARS-CoV-2 lineages. 2024. <https://doi.org/10.1101/2024.07.18.604213>
49. Tamura T, Ito J, Uriu K, Zahradnik J, Kida I, Anraku Y, et al. Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. *Nat Commun*. 2023;14(1). <https://doi.org/10.1038/s41467-023-38435-3>
50. Qu P, Evans JP, Zheng Y-M, Carlin C, Saif LJ, Oltz EM, et al. Evasion of neutralizing antibody responses by the SARS-CoV-2 BA.2.75 variant. *Cell Host Microbe*. 2022;30(11):1518–1526.e4. <https://doi.org/10.1016/j.chom.2022.09.015> PMID: [36240764](https://pubmed.ncbi.nlm.nih.gov/36240764/)
51. Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*. 2022;602(7898):657–63. <https://doi.org/10.1038/s41586-021-04385-3> PMID: [35016194](https://pubmed.ncbi.nlm.nih.gov/35016194/)
52. Halfmann PJ, Minor NR, Haddock Iii LA, Maddox R, Moreno GK, Braun KM, et al. Evolution of a globally unique SARS-CoV-2 Spike E484T monoclonal antibody escape mutation in a persistently infected, immunocompromised individual. *Virus Evol*. 2022;9(2):veac104. <https://doi.org/10.1093/ve/veac104> PMID: [37692895](https://pubmed.ncbi.nlm.nih.gov/37692895/)
53. Dong J, Zost SJ, Greaney AJ, Starr TN, Dingens AS, Chen EC, et al. Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. *Nat Microbiol*. 2021;6(10):1233–44. <https://doi.org/10.1038/s41564-021-00972-2> PMID: [34548634](https://pubmed.ncbi.nlm.nih.gov/34548634/)
54. Jones BE, Brown-Augsburger PL, Corbett KS, Westendorf K, Davies J, Cujec TP, et al. The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates. *Sci Transl Med*. 2021;13(593):eabf1906. <https://doi.org/10.1126/scitranslmed.abf1906> PMID: [33820835](https://pubmed.ncbi.nlm.nih.gov/33820835/)
55. Futatsusako H, Hashimoto R, Yamamoto M, Ito J, Matsumura Y, Yoshifuji H, et al. Longitudinal analysis of genomic mutations in SARS-CoV-2 isolates from persistent COVID-19 patient. *iScience*. 2024;27(5):109597. <https://doi.org/10.1016/j.isci.2024.109597> PMID: [38638575](https://pubmed.ncbi.nlm.nih.gov/38638575/)
56. Zhang QE, Lindenberger J, Parsons RJ, Thakur B, Parks R, Park CS, et al. SARS-CoV-2 Omicron XBB lineage spike structures, conformations, antigenicity, and receptor recognition. *Mol Cell*. 2024;84(14):2747–2764.e7. <https://doi.org/10.1016/j.molcel.2024.06.028> PMID: [39059371](https://pubmed.ncbi.nlm.nih.gov/39059371/)
57. Dadonaite B, Brown J, McMahon TE, Farrell AG, Figgins MD, Asarnow D, et al. Spike deep mutational scanning helps predict success of SARS-CoV-2 clades. *Nature*. 2024;631(8021):617–26. <https://doi.org/10.1038/s41586-024-07636-1> PMID: [38961298](https://pubmed.ncbi.nlm.nih.gov/38961298/)
58. Liu L, Iketani S, Guo Y, Chan JF-W, Wang M, Liu L, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature*. 2022;602(7898):676–81. <https://doi.org/10.1038/s41586-021-04388-0> PMID: [35016198](https://pubmed.ncbi.nlm.nih.gov/35016198/)
59. Bronstein Y, Adler A, Katash H, Halutz O, Herishanu Y, Levytskyi K. Evolution of spike mutations following antibody treatment in two immunocompromised patients with persistent COVID-19 infection. *J Med Virol*. 2022;94(3):1241–5. <https://doi.org/10.1002/jmv.27445> PMID: [34755363](https://pubmed.ncbi.nlm.nih.gov/34755363/)

60. Scheaffer SM, Lee D, Whitener B, Ying B, Wu K, Liang C-Y, et al. Bivalent SARS-CoV-2 mRNA vaccines increase breadth of neutralization and protect against the BA.5 Omicron variant in mice. *Nat Med*. 2023;29(1):247–57. <https://doi.org/10.1038/s41591-022-02092-8> PMID: [36265510](https://pubmed.ncbi.nlm.nih.gov/36265510/)
61. Schmidt F, Weisblum Y, Rutkowska M, Poston D, DaSilva J, Zhang F, et al. High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature*. 2021;600(7889):512–6. <https://doi.org/10.1038/s41586-021-04005-0> PMID: [34544114](https://pubmed.ncbi.nlm.nih.gov/34544114/)
62. Ke R, Martinez PP, Smith RL, Gibson LL, Mirza A, Conte M, et al. Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *Nat Microbiol*. 2022;7(5):640–52. <https://doi.org/10.1038/s41564-022-01105-z> PMID: [35484231](https://pubmed.ncbi.nlm.nih.gov/35484231/)
63. Ke R, Martinez PP, Smith RL, Gibson LL, Achenbach CJ, McFall S, et al. Longitudinal Analysis of SARS-CoV-2 Vaccine Breakthrough Infections Reveals Limited Infectious Virus Shedding and Restricted Tissue Distribution. *Open Forum Infect Dis*. 2022;9(7):ofac192. <https://doi.org/10.1093/ofid/ofac192> PMID: [35791353](https://pubmed.ncbi.nlm.nih.gov/35791353/)
64. Hannon WW, Roychoudhury P, Xie H, Shrestha L, Addetia A, Jerome KR, et al. Narrow transmission bottlenecks and limited within-host viral diversity during a SARS-CoV-2 outbreak on a fishing boat. *Virus Evol*. 2022;8(2):veac052. <https://doi.org/10.1093/ve/veac052> PMID: [35799885](https://pubmed.ncbi.nlm.nih.gov/35799885/)
65. Lin G-L, Drysdale SB, Snape MD, O'Connor D, Brown A, MacIntyre-Cockett G, et al. Distinct patterns of within-host virus populations between two subgroups of human respiratory syncytial virus. *Nat Commun*. 2021;12(1):5125. <https://doi.org/10.1038/s41467-021-25265-4> PMID: [34446722](https://pubmed.ncbi.nlm.nih.gov/34446722/)