Check for updates

Predominance of antibody-resistant SARS-CoV-2 variants in vaccine breakthrough cases from the San Francisco Bay Area, California

Venice Servellita[1,2], Mary Kate Morris[3], Alicia Sotomayor-Gonzalez[1,2], Amelia S. Gliwa[1,2], Erika Torres[4], Noah Brazer[1,2], Alicia Zhou[4], Katherine T. Hernandez[5], Madeline Sankaran[5], Baolin Wang[1,2], Daniel Wong[1,2], Candace Wang[1,2], Yueyuan Zhang[1,2], Kevin R. Reyes[1,2], Dustin Glasner ●[1,2], Xianding Deng[1,2], Jessica Streithorst[1,2], Steve Miller[1,2], Edwin Frias[6], Mary Rodgers[6], Gavin Cloherty[6], John Hackett Jr.[6], Carl Hanson[3], Debra Wadford ●[3], Susan Philip[5], Scott Topper[4], Darpun Sachdev[5] and Charles Y. Chiu ●[1,2,7,8] ✉

Associations between vaccine breakthrough cases and infection by different SARS coronavirus 2 (SARS-CoV-2) variants have remained largely unexplored. Here we analysed SARS-CoV-2 whole-genome sequences and viral loads from 1,373 persons with COVID-19 from the San Francisco Bay Area from 1 February to 30 June 2021, of which 125 (9.1%) were vaccine breakthrough infections. Vaccine breakthrough infections were more commonly associated with circulating antibody-resistant variants carrying ≥1 mutation associated with decreased antibody neutralization (L452R/Q, E484K/Q and/or F490S) than infections in unvaccinated individuals (78% versus 48%, $P = 1.96 \times 10^{-8}$). Differences in viral loads were non-significant between unvaccinated and fully vaccinated cases overall ($P = 0.99$) and according to lineage ($P = 0.09–0.78$). Symptomatic vaccine breakthrough infections had comparable viral loads ($P = 0.64$), whereas asymptomatic breakthrough infections had decreased viral loads ($P = 0.023$) compared with infections in unvaccinated individuals. In 5 cases with serial samples available for serologic analyses, vaccine breakthrough infections were found to be associated with low or undetectable neutralizing antibody levels attributable to an immunocompromised state or infection by an antibody-resistant lineage. Taken together, our results show that vaccine breakthrough infections are overrepresented by antibody-resistant SARS-CoV-2 variants, and that symptomatic breakthrough infections may be as efficient in spreading COVID-19 as unvaccinated infections, regardless of the infecting lineage.

Vaccines targeting the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) have been highly effective in preventing symptomatic illness and in reducing hospitalizations and deaths from coronavirus disease 2019 (COVID-19)[1–7]. Previous studies have also suggested that vaccination may reduce viral loads in persons with breakthrough SARS-CoV-2 infection who have received at least one dose[7–9], thus decreasing infectiousness and mitigating transmission. However, most of these studies were done before the emergence of 'antibody-resistant' SARS-CoV-2 variants of concern/variants of interest (VOCs/VOIs) carrying key mutations that have been shown to decrease antibody (Ab) neutralization (L452R/Q, E484K/Q and/or F490S), including the Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Epsilon (B.1.427/B.1.429) and Lambda (C.37), but not the Alpha (B.1.1.7) variants[10–14]. Breakthrough infections have been reported in a small proportion of vaccine recipients[3,15–24], yet little is known regarding the relative capacity of different circulating variants to escape vaccine-induced immunity and facilitate ongoing spread within highly vaccinated communities.

In San Francisco County, a sharp decline in COVID-19 cases following a 2020–2021 winter outbreak of the Epsilon variant in California[10,25] preceded mass vaccination efforts (Fig. 1a). From February to June 2021, the number of cases per day continued to gradually decrease, despite a nationwide outbreak from the Alpha variant in the United States[26] and the continual introduction of other VOCs/VOIs into the community[25]. In late June, there was an uptick in Delta variant cases presaging a surge of infections from this variant in the county and nationwide[25].

Here we performed whole-genome sequencing and viral load measurements of nasal swabs in conjunction with retrospective medical chart review from COVID-19 infected persons over a 5-month period to investigate dynamic longitudinal shifts in the distribution of SARS-CoV-2 variants over time and identify potential correlates of breakthrough infections in a progressively vaccinated community.

Results

We performed whole-genome sequencing of available remnant mid-turbinate nasal, nasopharyngeal and/or oropharyngeal

[1]Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA, USA. [2]UCSF-Abbott Viral Diagnostics and Discovery Center, San Francisco, CA, USA. [3]Viral and Rickettsial Disease Laboratory, California Department of Public Health, Richmond, CA, USA. [4]Color Genomics, Inc., Burlingame, CA, USA. [5]San Francisco Department of Public Health, San Francisco, CA, USA. [6]Abbott Laboratories, Abbott Park, IL, USA. [7]Department of Medicine, University of California San Francisco, San Francisco, CA, USA. [8]Innovative Genomics Institute, University of California Berkeley, Berkeley, CA, USA. ✉e-mail: charles.chiu@ucsf.edu
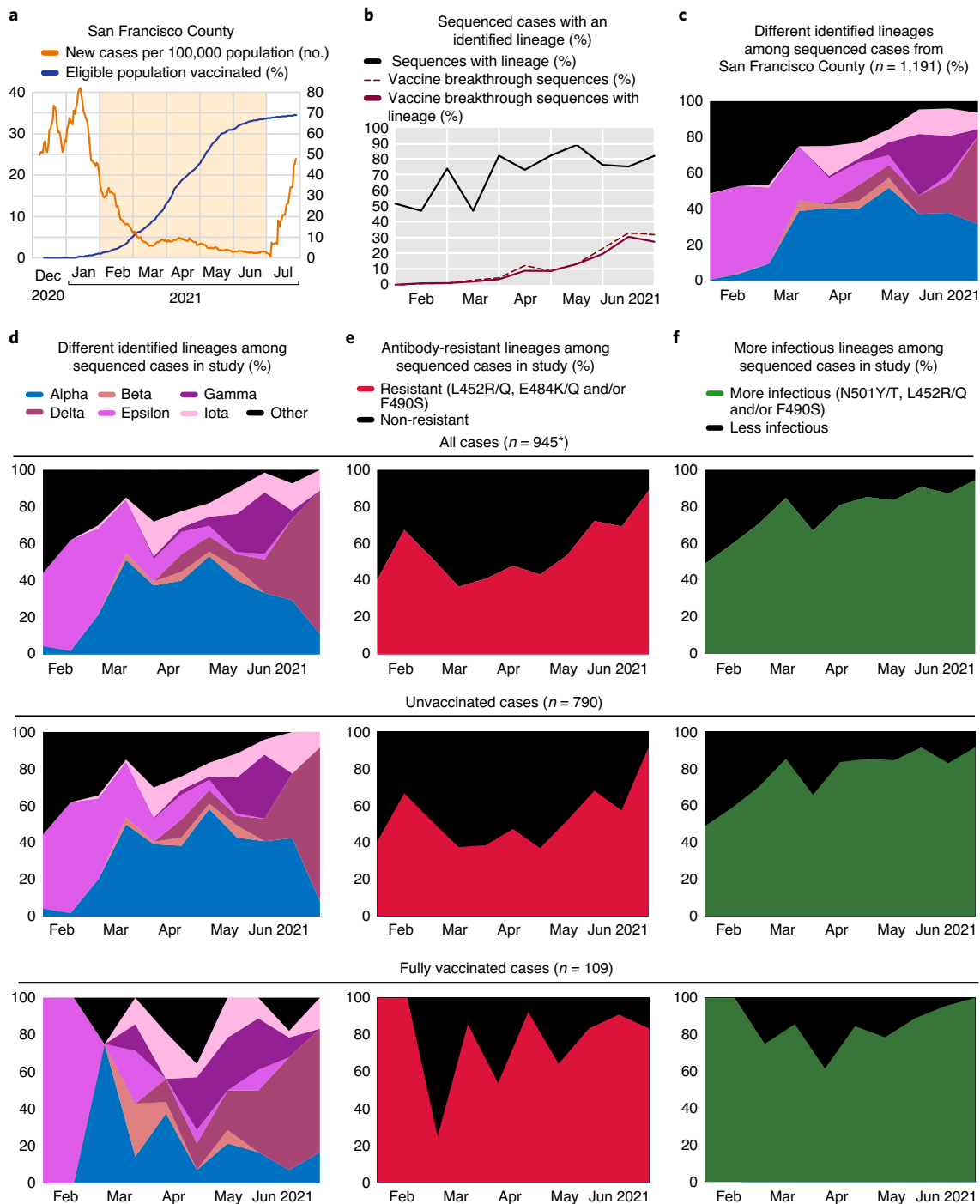
Fig. 1 | Overview of vaccination and SARS-CoV-2 whole-genome sequencing data from the San Francisco Bay Area. a, Plot showing the percentage of eligible individuals in San Francisco County who had received an FDA-authorized vaccine from the beginning of mass vaccine rollout until 30 June 2021 (blue line) and the number of new SARS-CoV-2 positive cases per 100,000 people in San Francisco County (orange line). The peach-coloured area denotes the study timeframe for sample collection (1 February to 30 June 2021). **b**, Plot showing the proportion of sequenced genomes with identified lineages from all cases (black line) and from fully vaccinated breakthrough cases (dark red line). The dotted line shows the proportion of sequenced genomes from breakthrough cases, regardless of whether a lineage was identified. **c**, Plot showing the distribution of SARS-CoV-2 lineages identified using Pangolin algorithm[31] from sequenced COVID-19 cases from San Francisco County, aggregated biweekly from 1 February to 30 June 2021. **d**, Plots showing the distribution of Pangolin-identified SARS-CoV-2 lineages from all sequenced (top), unvaccinated (middle) and fully vaccinated breakthrough (bottom) cases. **e**, Plots showing the proportion of sequenced genomes carrying mutations associated with antibody resistance (L452R/Q, E484K/Q and/or F490S) among all sequenced (top), unvaccinated (middle) and fully vaccinated breakthrough (bottom) cases. **f**. Plots showing the proportion of sequenced genomes carrying mutations associated with enhanced infectivity (L452R/Q, N501Y/T and/or F490S) among all sequenced (top), unvaccinated (middle) and fully vaccinated breakthrough (bottom) cases. Proportions were calculated relative to the total number of sequenced cases and aggregated biweekly from 1 February to 30 June 2021.

swab samples collected from 1,373 polymerase chain reaction (PCR)-positive COVID-19 cases from San Francisco County from 1 February to 30 June 2021. During this study period, the percentage of eligible persons vaccinated in the county increased from 2 to 70%, while the number of new cases per 100,000 population declined from 23 to 2 (Fig. 1a). The cohort included COVID-19 patients seen in hospitals and clinics at University of California, San Francisco (UCSF, $n=598$, 43.6%) and infected persons identified by community testing in San Francisco County performed by a commercial laboratory (Color Genomics, $n=775$, 56.4%). Using the U.S. Centers for Disease Control and Prevention (CDC) definition of a vaccine breakthrough infection as a positive SARS-CoV-2 RNA or antigen test ≥ 14 d after completion of all recommended doses of a U.S. Food and Drug Administration (FDA)-authorized vaccine[27], 125 (9.1%) of infections in the cohort were vaccine breakthroughs (Extended Data Fig. 1). Of the 125 breakthrough cases, 122 (97.6%) were confirmed to have received the Pfizer-BioNTech (BNT162b2) COVID-19 mRNA, Moderna (mRNA-1273) COVID-19 mRNA, or Johnson & Johnson/Janssen (JnJ) COVID-19 viral vector (adenovirus) vaccine. The percentage of sequenced cases that were vaccine breakthroughs increased from 0% to 31.8% from February to June (Fig. 1b). Among the viruses sequenced from the 1,373 cases, 69% (945 of 1,373) were unambiguously assigned to a SARS-CoV-2 lineage (Fig. 1b and Extended Data Fig. 1). The remaining 31% were not assigned a lineage due to insufficient genome coverage, and hence were not used for variant identification or for phylogenetic analysis.

Multiple variant lineages were found to be circulating in San Francisco County during the 5-month study period (Fig. 1c,d). The distribution of study lineages among all (Fig. 1d, top) and unvaccinated (Fig. 1d, middle) cases reflected the community distribution based on all 1,191 available reference genomes in the Global Initiative on Sharing All Influenza Data (GISAID) database, which also includes SARS-CoV-2 sequences (Fig. 1c). In contrast, the distribution of study lineages among vaccinated cases was skewed, with overrepresentation of antibody-resistant lineages (those containing ≥ 1 of L452R/Q, E484K/Q and/or F490S mutation[10-14]), including Beta, Gamma, Delta, Epsilon and Iota, and a corresponding decreased number and proportion of Alpha variant cases (Fig. 1c, bottom). Among all and unvaccinated cases (Fig. 1e), the proportion of antibody-resistant variants increased from approximately 40% to 90%. In contrast, antibody-resistant lineages comprised a higher percentage of cases in fully vaccinated as compared with unvaccinated or all cases (Fig. 1e, bottom). The proportion of variants with increased infectivity (those containing ≥ 1 of L452R/Q, N501Y/T and/or F490S) mutations[10-14,28] increased over time to >95% in all, unvaccinated and fully vaccinated cases (Fig. 1f). By phylogenetic analysis, the genomes from fully vaccinated cases were found to be intermixed with those from unvaccinated cases and broadly distributed across all the major viral subclades (Fig. 2a,b), with proportionally more genomes assigned to subclades associated with antibody-resistant variants and no evidence of clustering by either time (Fig. 2a) or genetic distance (number of mutations)

(Fig. 2b). A multiple sequence alignment comparing 42 representative SARS-CoV-2 genomes, including 6 variant lineages and 1 non-VOC/VOI lineage, showed similar overlapping mutation patterns between vaccinated and unvaccinated cases (Extended Data Fig. 2). After stratifying by month, we identified statistically significant differences in the proportion of antibody-resistant variants between fully vaccinated and unvaccinated cases in April, May, and June 2021 (Fig. 2c), but not in February and March, when the numbers of vaccine breakthrough cases were very low.

Among unvaccinated cases, most viruses consisted of non-resistant variants (57% and 61% based on hospital and community testing, respectively) (Fig. 3a, left), in contrast to vaccinated cases, for which the proportions of non-resistant variants fell to 34% and 20%, respectively (Fig. 3a, right). Variant distribution in unvaccinated cases were predominated by Alpha and Epsilon cases, whereas infections by the Gamma and Delta variants, which cause more pronounced decreases in Ab neutralization than other VOCs[11,29], were increased in fully vaccinated breakthrough infections. The variant distribution in partially vaccinated cases was similar to that in unvaccinated cases (Extended Data Fig. 3). Overall, fully vaccinated cases were significantly more likely than unvaccinated cases to be infected by resistant variants (77.6% versus 47.7%, $P=1.96\times10^{-8}$), but not by variants associated with increased infectivity (84.7% versus 76.8%, $P=0.092$) (Fig. 3b, left and Extended Data Fig. 1). A proportionally increased number of infections from antibody-resistant variants in fully vaccinated cases was also observed for the subset of cases in immunocompetent patients (Fig. 3b, middle), which exhibited a similar distribution of variants (Fig. 3a, inner circles), but not in immunocompromised patients (Fig. 3b, right).
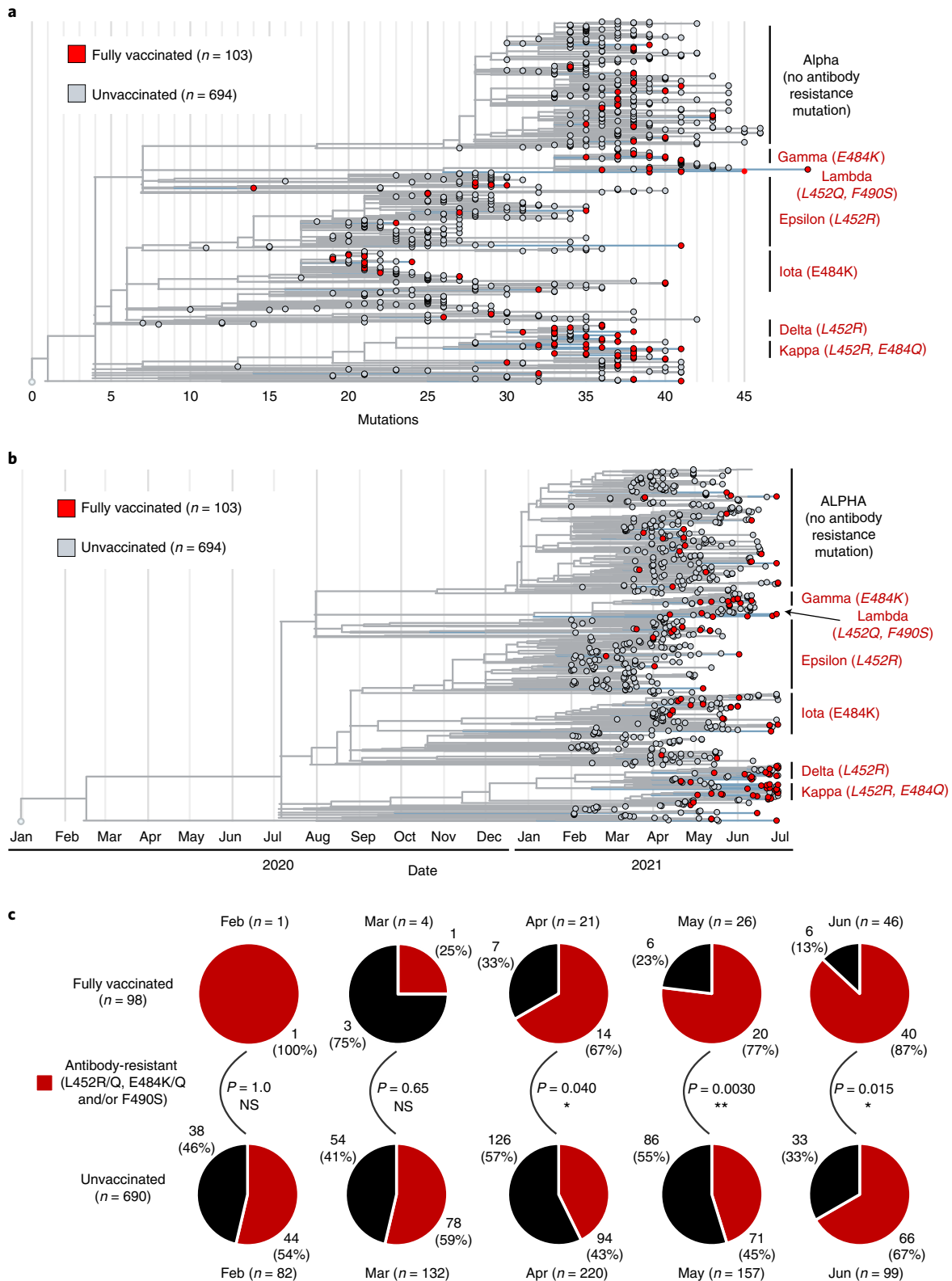
Viral RNA loads for infected persons in the full patient cohort, 125 of whom were vaccine breakthrough cases, were estimated by comparing differences in mean cycle threshold (Ct) values using quantitative PCR with reverse transcription (RT–qPCR) and use of a standard curve (Extended Data Fig. 4). There was no difference in viral RNA loads between fully vaccinated breakthrough and unvaccinated cases, either overall ($P=0.99$) or according to lineage ($P=0.09$–0.78) (Fig. 4a,b). Infections from variants of concern (VOCs/VOIs) had 2× viral loads compared with non-VOC/VOI lineages overall ($P=0.017$) (Fig. 4c). With respect to individual VOCs, higher viral RNA loads were observed for infections by the Gamma (4×, $P=0.00076$), Delta (3×, $P=0004$) and Epsilon (2×, $P=0.047$) variants, but not for the Alpha, Beta and Iota variants (Fig. 4d).
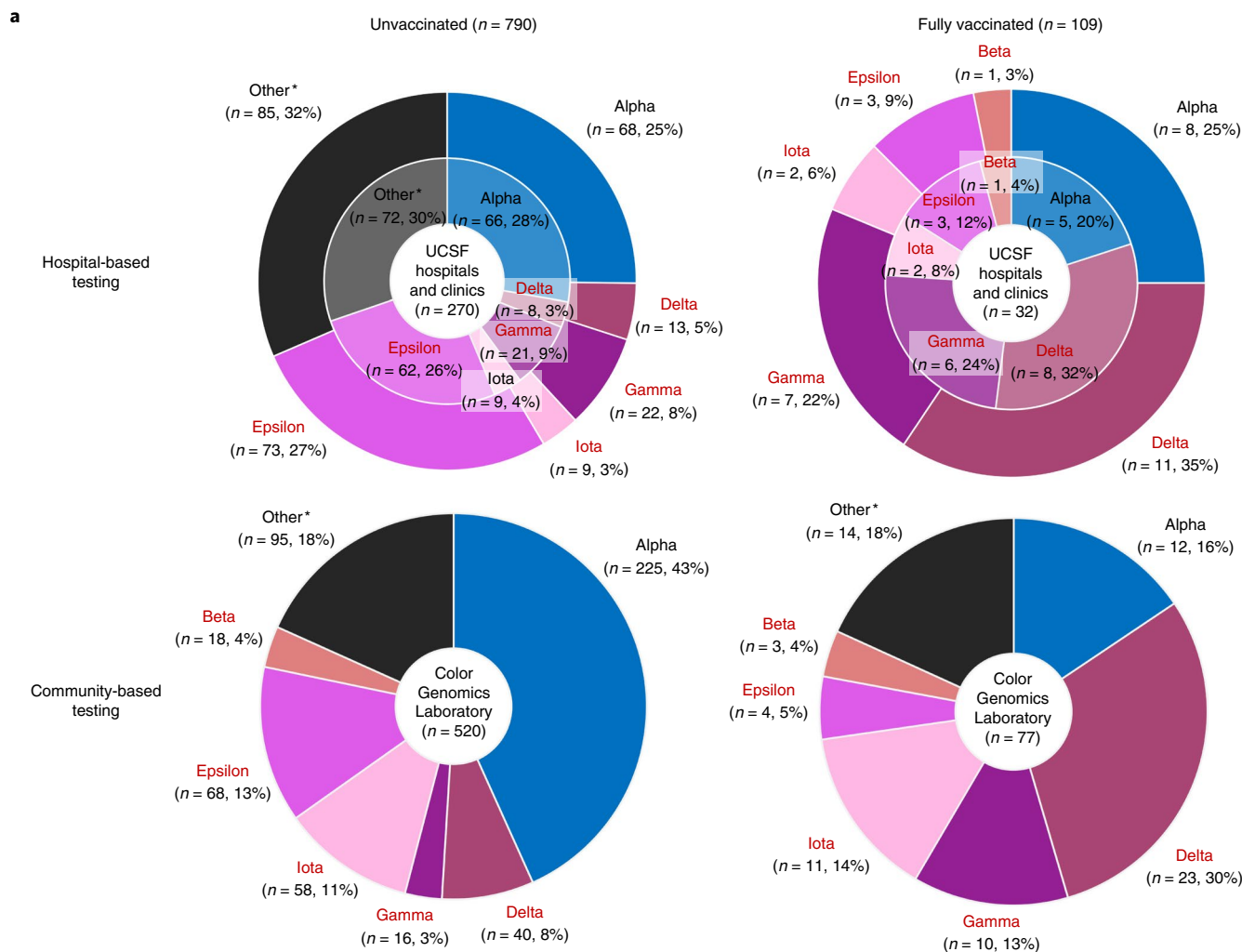
We investigated potential correlations between variant identification, clinical symptomatology, vaccine type and viral load in vaccine breakthrough infections. Retrospective medical chart review was performed for a subset of patients from UCSF hospitals and clinics with available clinical and demographic data ($n=598$) (Table 1). Among the 39 breakthrough infections out of 598 in this subset, the average age was 49 years (range 22 to 97), and the majority were women (54%). The median interval from completion of all doses of the vaccine and COVID-19 breakthrough infection was 73.5 d

Fig. 2 | Phylogenetic and temporal analyses of SARS-CoV-2 genomes. a, Phylogenetic tree showing the patterns of divergence and clade assignments of 797 sequenced SARS-CoV-2 positive genomes from California in the study, along with reference genomes from a globally representative dataset obtained from Nextstrain[47] (grey branches). Study genomes from fully vaccinated ($n=103$, red circles) or unvaccinated ($n=694$, grey circles) cases are shown. Clades corresponding to the major SARS-CoV-2 variants identified in the study are named, including those that carry antibody-resistant mutations (dark red boldface text). The length of each branch is proportional to the number of mutations occurring on that branch. **b**, The same phylogenetic tree, except that the length of each branch is determined by inference of the molecular clock phylogeny using TreeTime[50], which positions terminal nodes according to the sampling times and internal nodes at the most likely time of divergence. **c**, Pie charts showing the proportions of SARS-CoV-2 genomes containing identified mutations associated with antibody resistance (L452R/Q, E484K/Q and/or F490S) among fully vaccinated (top row) and unvaccinated (bottom row) cases per month over a 5-month period (from February to June 2021). The charts are shaded according to genomes carrying ≥ 1 mutation associated with antibody resistance (L452R/Q, E484K/Q and/or F490S) (red), or lacking an antibody-resistant mutation (black). Fisher's Exact test (two-tailed) was used to calculate P values. NS, non-significant; *$P<0.05$, **$P<0.01$.

(range 15 to 140). The Pfizer-BioNTech (BNT162b2) COVID-19 mRNA vaccine was administered to 20 (51%) of the vaccine breakthrough patients, while 12 (31%) received the Moderna (mRNA-1273) COVID-19 mRNA vaccine and 4 (10%) received the Johnson & Johnson/Janssen (JnJ) COVID-19 viral vector (adenovirus) vaccine (Extended Data Fig. 5).

Nine (23%) of the vaccine breakthrough patients were immunocompromised, while 28 (72%) were identified as symptomatic and 10 (26%) as asymptomatic. Among the symptomatic breakthrough infections, 6 patients (15.4%) were hospitalized for COVID-19 pneumonia, 1 patient (2.6%) required care in the intensive care unit (ICU) and 0 patients (0%) died. Among the unvaccinated infections

a

Unvaccinated (n = 790)

Fully vaccinated (n = 109)

Hospital-based testing

Community-based testing

*Includes some lineages carrying antibody-resistant mutations (Kappa and Lambda)

b

Antibody-resistant (L452R/Q, E484K/Q and/or F490S)

More infectious (N501Y/T, L452R/Q and/or F490S)
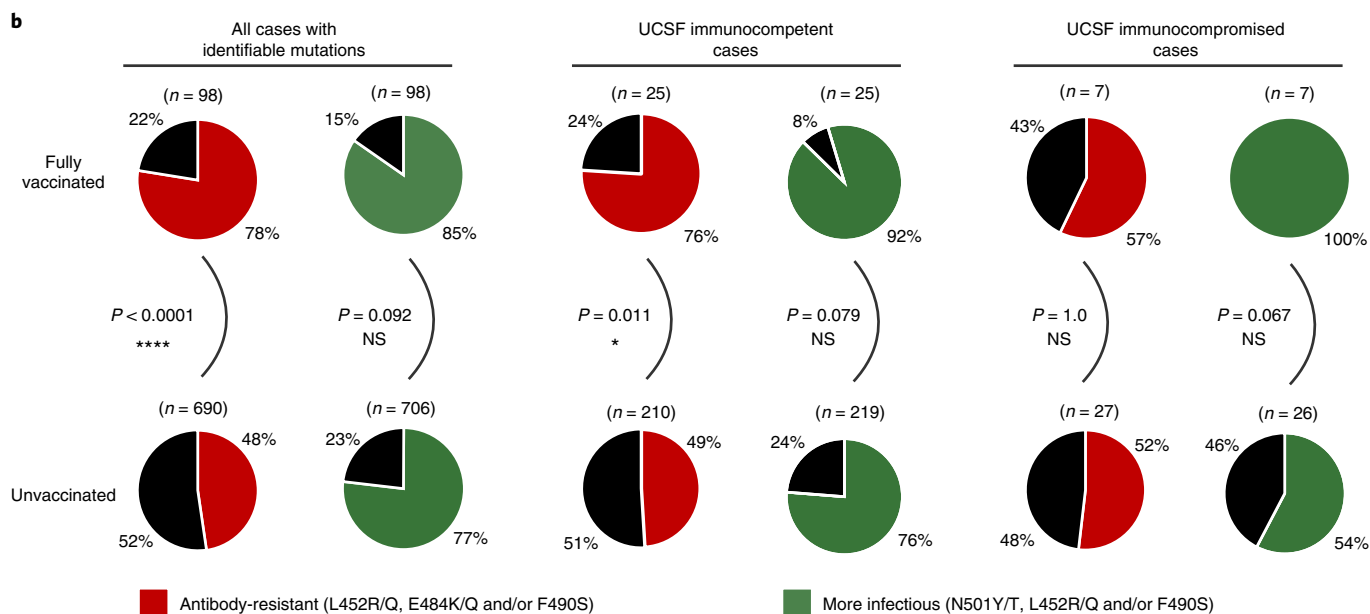
(n = 433), 287 (66%) patients presented with symptoms, while 132 (30%) patients reported no symptoms; 51 patients (11.8%) required hospitalization, 27 (6.2%) patients were admitted to the ICU and

5 (1.2%) patients died with COVID-19 reported as the primary cause of death. Among these clinical and demographic variables, only advanced age of >65 years was significantly associated with

Fig. 3 | Comparisons of lineage distribution and proportion of mutations associated with antibody resistance and increased infectivity in fully vaccinated and unvaccinated cases. a, Pie charts showing the distribution of SARS-CoV-2 variant lineages in fully vaccinated and unvaccinated cases from UCSF hospitals and clinics (top row) and from Color Genomics Laboratory (bottom row). For the UCSF charts, the lighter-shaded inner circles show cases in immunocompetent patients only, while the outer circles include both immunocompetent and immunocompromised individuals. Variants carrying mutations associated with antibody resistance are highlighted in dark red boldface text. **b**, Pie charts showing the proportions of SARS-CoV-2 genomes carrying mutations associated with antibody resistance (L452R/Q, E484K/Q and/or F490S) and increased infectivity (N501Y/T, L452R/Q and/or F490S) in fully vaccinated (top row) and unvaccinated cases (bottom row). The pie charts include genomes corresponding to all sequenced cases containing identifiable mutations (left) and immunocompetent (middle) or immunocompromised (right) patients from UCSF hospitals and clinics. The charts are shaded according to genomes carrying ≥1 mutation associated with antibody resistance (red), ≥1 mutation associated with increased infectivity (green), or neither type of mutation (black). Fisher's Exact test (two-tailed) was used to calculate P values. *P < 0.05, ****P < 0.0001.

vaccine breakthrough infections as compared with unvaccinated infections ($P = 0.035$, odds ratio 2.36 (95% confidence interval (CI) 0.97–5.45)). Viral RNA loads were significantly higher overall for symptomatic as compared with asymptomatic infections for both unvaccinated ($P = 0.0014$, $\Delta Ct = 8.8$ (95% CI 4.0–13.7), 4.5×) and vaccine breakthrough ($P = 9.8 \times 10^{-5}$, $\Delta Ct = 2.8$ (95% CI 1.42–4.21), 164×) cases (Fig. 3b). Differences in RNA viral loads for vaccine breakthroughs as compared with unvaccinated cases were non-significant for symptomatic cases ($P = 0.64$, $\Delta Ct = 0.6$ (95% CI −2.0 to 3.2), 1.6×) but were significant for asymptomatic cases ($P = 0.023$, $\Delta Ct = -5.4$ (95% CI -9.9 – 0.09), 0.043×), while viral RNA loads of hospitalized patients with COVID-19 did not differ significantly from those of outpatients (Extended Data Fig. 6).

We sought to understand the serologic basis behind some of the vaccine breakthrough infections in the study cohort. Plasma samples were available from 5 of 39 (12.8%) patients with clinical metadata for qualitative testing of nucleoprotein immunoglobulin G (IgG) and spike immunoglobulin M (IgM) levels and neutralizing Ab titre using a cytopathic effect (CPE) endpoint neutralization assay as previously described[10]. For 4 of the 5 patients, serially collected samples were available. Neutralization assays were tested for activity against cultures of a D614G-carrying non-VOC control virus and the Alpha, Beta, Gamma, Delta and Epsilon variants (Fig. 5 and Extended Data Fig. 7). Among the 4 cases out of 5 in immunocompromised patients, 3 patients (Fig. 5 and Extended Data Fig. 7, P1, P2 and P5) failed to mount detectable qualitative and neutralizing Ab responses to the vaccine, likely due to their immunocompromised status (Extended Data Fig. 7). The interpretation was indeterminate for one immunocompromised patient (Fig. 5 and Extended Data Fig. 7, P3), as only samples at day 11 post-breakthrough or later were available, by which time the patient had probably generated a robust antibody response to the breakthrough infection. The remaining vaccine breakthrough case out of 5 (Fig. 5 and Extended Data Fig. 7, P4) was an immunocompetent patient who had received the JnJ vaccine and had also been previously infected with COVID-19 before vaccination. This patient was negative for detection of qualitative nucleoprotein IgG and spike protein IgM Ab from plasma 2 d after testing SARS-CoV-2 positive from nasopharyngeal swab by RT–qPCR; however, strong positivity for spike protein IgG Ab and high neutralizing Ab titre

against D614G-carrying control virus suggested a robust antibody response to vaccination. Levels of neutralizing Ab were lowest for the Beta, Delta and Epsilon variants, consistent with the patient's breakthrough infection by Delta.

Discussion

Here we used variant identification by SARS-CoV-2 whole-genome sequencing, quantitative viral load analysis and antibody studies, along with retrospective medical chart review to compare vaccine breakthrough ($n = 125$, 9.1%) and unvaccinated ($n = 1,169$, 85.1%) cases in both community and hospitalized settings from northern California. Previous reports have shown that the distribution of VOCs/VOIs in breakthrough cases generally reflect the estimated community prevalence in the unvaccinated population[3,15–24]. These reports, however, investigated breakthrough cases over a limited timeframe during which only a single predominant lineage was typically circulating. The current study spanned 5 months as the study population became progressively vaccinated from 2% to >70% while undergoing 3 successive surges of infection from the Epsilon (February–March 2021)[10], Alpha (March–June 2021)[26] and Delta (June 2021) variants[25], with simultaneous circulation of multiple viral variants in the community. In contrast to these previous studies, we found that vaccine breakthrough infections are overrepresented by immunity-evading variants as compared with unvaccinated infections. Phylogenetic analyses revealed that the viruses in vaccine breakthrough and unvaccinated cases had similar genomes with assignment to subclades representing all major variants circulating in the community. The overrepresentation of infection from immunity-evading variants in vaccinated cases thus most likely arises from differential effectiveness of neutralizing antibodies against multiple circulating lineages. Notably, a decreased proportion of vaccine breakthrough infections from the Alpha variant was observed, despite its documented higher infectivity relative to all VOCs except Delta and Gamma[10,30]. Decreased Alpha infections are consistent with the higher effectiveness of available SARS-CoV-2 vaccines against Alpha relative to other VOCs[2,11,13,21].

The predominance of immune-evading variants among post-vaccination cases indicates possible selective pressure for antibody-resistant escape variants circulating locally over time in the vaccinated population. In particular, the Delta variant, which

Fig. 4 | Comparison of viral loads according to vaccination status, SARS-CoV-2 lineage and clinical symptom status. a, Grouped box-and-whisker plots and swarm plots showing the differences in mean cycle threshold (Ct) values and calculated viral loads in copies per ml (cp ml⁻¹) between fully vaccinated and unvaccinated cases overall (left) and stratified by lineage (right). For all comparisons, the difference in viral loads was not statistically significant ($P > 0.05$). For the lineage plots on the right, the mean Ct values and approximate viral loads corresponding to unvaccinated and fully vaccinated cases are shown as $\mu_{unvaccinated} \mid \mu_{vaccinated}$ and cp/ml$_{unvaccinated} \mid$ cp/ml$_{vaccinated}$, respectively, below the name. **b**, Box-and-whisker and swarm plots showing the differences in viral loads between symptomatic and asymptomatic cases, stratified by vaccination status. **c**, Box-and-whisker and swarm plots showing the differences in viral load between specific lineages identified as VOC/VOI and other lineages that were not designated VOCs or VOIs at the time of this study ('non-VOC/VOI'). Identified VOCs/VOIs included Alpha, Beta, Gamma, Delta, Epsilon and Iota variants, following the World Health Organization (WHO) nomenclature scheme. **d**, Box-and-whisker and swarm plots showing differences in viral loads between each VOC/VOI and non-VOC/VOI. For all box-and-whisker plots, the box outlines denote the interquartile range (IQR), the solid line inside the box denotes the median, the dashed line inside the box denotes the mean (μ) Ct value, and the whiskers outside the box extend to the minimum and maximum fold enrichment points. Welch's t-test was used for significance testing. A standard curve was used to determine the estimated viral load for a corresponding Ct value (Extended Data Fig. 4).
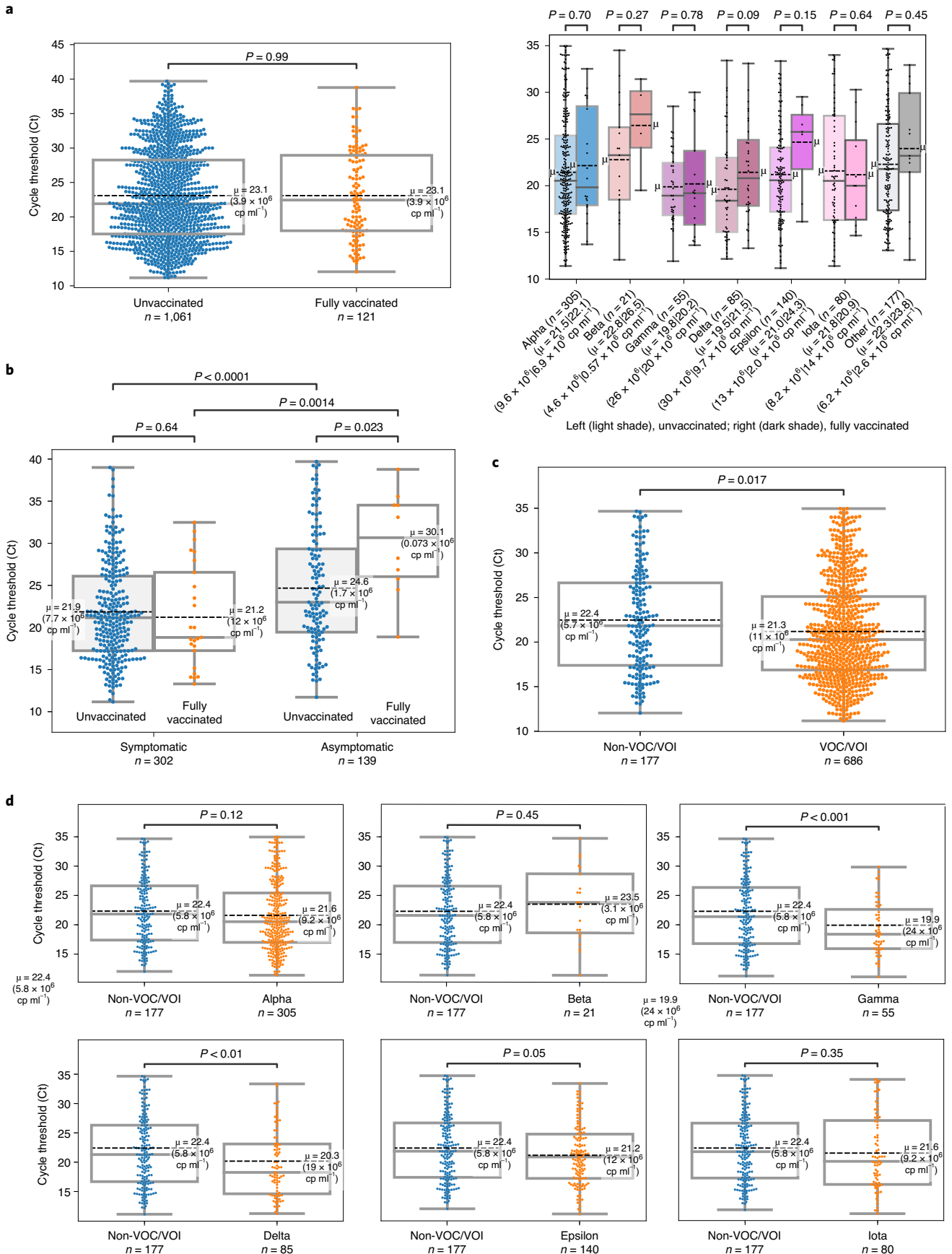
Table 1 | Clinical and demographic characteristics in vaccine breakthrough and unvaccinated cases

Characteristic		Fully vaccinated cases (no.)	Fully vaccinated cases (%)	Unvaccinated cases (no.)	Unvaccinated cases (%)	P value**	Odds ratio (95% confidence interval)
Gender	Female	21	53.85	241	55.66	0.87	0.93 (0.46–1.91)
	Male	18	46.15	192	44.34		
	Unknown	0	0	0	0		
Age	≤18 years old	0	0	187	43.19		
	>65 years old	11	28.21	35	8.08	0.035	2.36 (0.97–5.45)
	19–65 years old	28	71.79	211	48.73		
	Unknown	0	0	0	0		
Immunocompromised	Yes	9	23.08	50	11.55	0.073	2.26 (0.89–5.24)
	No	30	76.92	377	87.07		
	Unknown	0	0	6	1.39		
Symptoms	Symptomatic	28	71.79	289	66.74	0.59	1.28 (0.58–3.04)
	Asymptomatic	10	25.64	132	30.48		
	Unknown	1	2.56	12	2.77		
Hospitalized*	Yes	6	15.38	51	11.78	0.61	1.31 (0.43–3.4)
	No	26	66.67	311	71.82		
	non COVID-related	6	15.38	46	10.62		
	Unknown	1	2.56	25	5.77		
Deceased*	Yes	0	0	5	1.15	1	0 (n/a)
	No	38	97.44	418	96.54		
	Unknown	1	2.56	10	2.31		
Total		39	100	433	100		

*Patient hospitalized or deceased due to severe COVID-19 disease. **Two-tailed Fisher's Exact test was used to determine P values.

is the predominant circulating lineage in the United States as of July 2021, has been shown to be more resistant to vaccine-induced immunity as well as being more infectious than Alpha[13,21,29]. Although our data suggest that vaccination at levels below the threshold for achieving herd immunity may increase selection for antibody-resistant variants, it is notable that vaccine breakthrough infections comprised only a minority of total infections (9%, 125 out of 1,373 cases). These findings are consistent with previous reports showing that vaccination is effective in decreasing viral transmission[31–33], probably reducing the rate at which new variants emerge and spread in the community[34]. Among demographic and clinical factors associated with vaccine breakthrough infection, we only identified a significant association with age, consistent with the prioritized rollout of the vaccine in the elderly population[2]. Several studies have demonstrated that the vaccine remains highly effective against preventing symptomatic breakthrough infections resulting in serious illness leading to hospitalization and/or death[1–7]. Our findings are consistent with these other studies, as there were fewer hospital admissions and no deaths in vaccinated patients as compared with unvaccinated patients, although these differences were not statistically significant due to low case numbers.

We also found that differences in viral RNA loads (as estimated using Ct values) between vaccine breakthrough and unvaccinated infections were non-significant ($P=0.99$), regardless of lineage. A previous study of a community outbreak of Delta infections in the state of Massachusetts also found that viral loads were similar for both vaccinated and unvaccinated persons with COVID-19[17].

These findings probably formed the basis for revised indoor mask guidance in July 2021 from the US Centers for Disease Control and Prevention[35]. Our results show that comparably high viral loads in vaccine breakthrough infections are not confined to Delta alone; indeed, the highest viral loads were observed from Gamma. Notably, viral RNA loads in symptomatic vaccine breakthrough cases were approximately 164× higher as compared with asymptomatic cases ($P=0.0014$), and similar to those in unvaccinated cases ($P=0.64$). However, significantly lower viral RNA loads were observed in asymptomatic breakthrough cases as compared with unvaccinated cases (0.043×, $P=0.0023$). Taken together, these data suggest that symptomatic breakthrough cases are probably as infectious as symptomatic unvaccinated cases, and thus may contribute to ongoing SARS-CoV-2 transmission, even in a highly vaccinated community. These findings thus reinforce the importance of mask wearing recommendations in symptomatic persons to control community spread, regardless of vaccination status[35]. These also suggest that asymptomatic transmission of breakthrough cases may be less efficient given the lower viral loads. Contact tracing investigation of vaccine breakthroughs is likely needed to ascertain the role, if any, of asymptomatic transmission in vaccinated persons in SARS-CoV-2 spread.

Our antibody analyses, although performed on a small number of cases ($n=5$), show that vaccine breakthrough cases are generally associated with low or undetectable qualitative and neutralizing antibody levels in response to vaccination. These findings are consistent with studies that have correlated high antibody levels
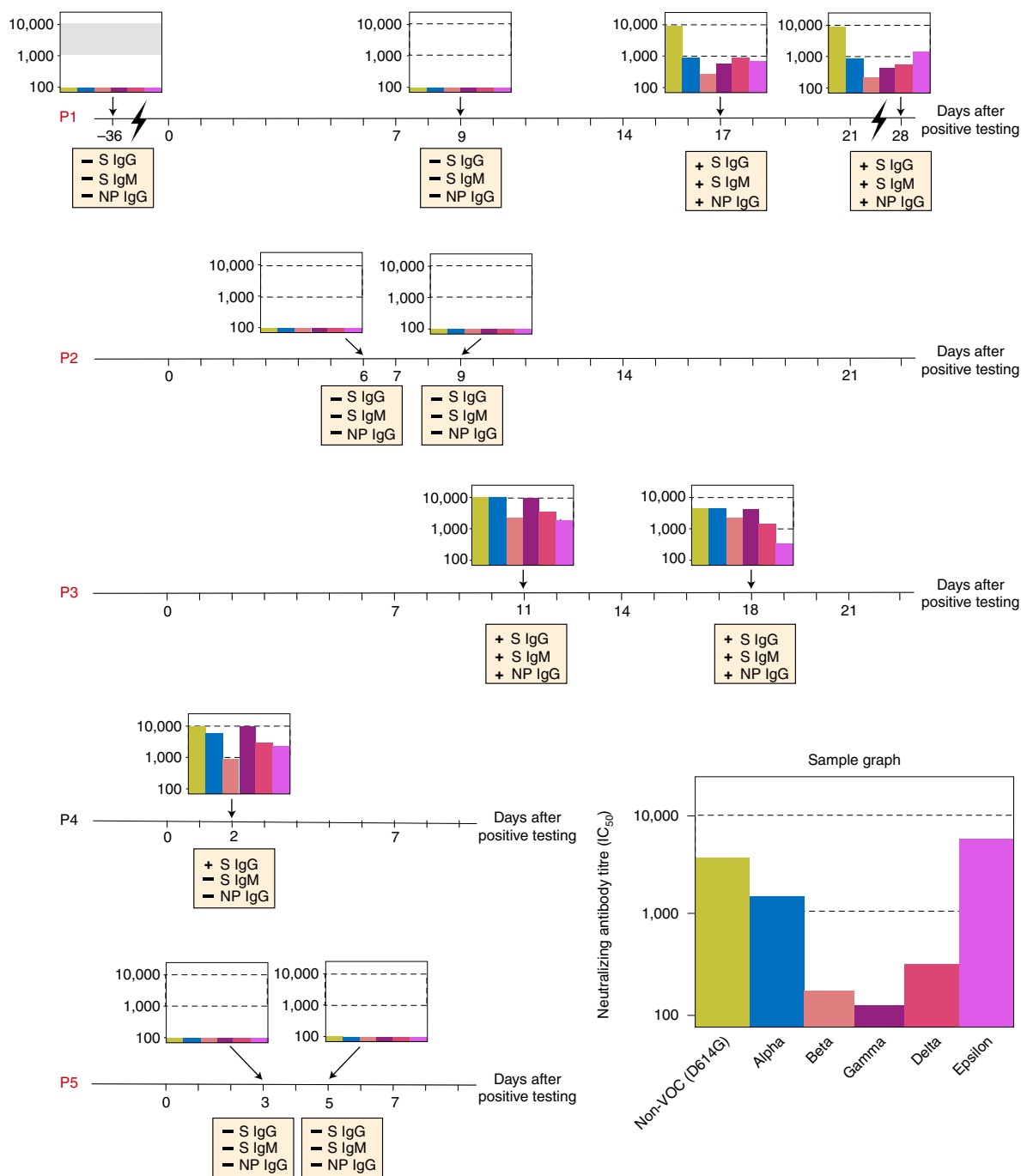
Fig. 5 | Qualitative and neutralizing antibody studies in vaccine breakthrough patients. Graphical timeline showing results of qualitative spike and nucleoprotein antibody testing and quantitative neutralizing antibody testing against 5 variants (Alpha, Gamma, Delta and Epsilon) and a 'non-VOC/VOI' D614G-carrying control strain at various timepoints relative to the date of positive SARS-CoV-2 testing ($t = 0$) for 5 patients with vaccine breakthrough infections (P1–P5). Immunocompromised patients (P1, P2, P3 and P5) are highlighted in red text. Qualitative antibody results are shown in peach-coloured boxes, while IC_{50} (50% inhibitory dose) neutralizing antibody titres against the 5 variants and D614 control virus are plotted as a bar graph on a log scale. The black zigzag marks a break in the timeline.

with vaccine efficacy[36]. We identified 3 cases of breakthrough infection from Alpha, all in immunocompromised patients who failed to mount detectable levels of neutralizing antibody to both wild-type and VOC SARS-CoV-2 lineages. The failure to mount adequate neutralizing antibody responses in a subset of immunocompromised patients may explain why a proportionally higher rate of infection by antibody-resistant variants was not observed in the current study. We also reported 1 case of Delta breakthrough

infection in a patient who had contracted COVID-19 in 2020 and had also received the JnJ vaccine. As the patient was asymptomatic, we do not know exactly when the breakthrough infection occurred, potentially explaining why a reduced but detectable neutralizing antibody response to Delta was observed. Although singular, this case demonstrates the likely inadequacy of convalescent antibodies generated from previous infection in protecting against future infection, especially against emerging antibody-resistant VOCs, and the reduced effectiveness of the JnJ vaccine relative to the mRNA vaccines against the Delta variant[37].

There are several limitations to our study. First, breakthrough infections in this study were identified by testing persons presenting to a tertiary hospital and clinic or as part of community-based testing by a commercial laboratory, hence sampling bias may be present. This limitation is mitigated by our results showing similar variant distributions and viral load comparisons across two separate test cohorts. Second, the total number of vaccinated persons with breakthrough infections was relatively small at 125, of which detailed clinical and epidemiologic metadata were only available for 39. Third, the sequencing depth of 5-fold or greater was not sufficient to identify intra-host diversity and viral quasi-species subpopulations. Fourth, clinical data were obtained by retrospective medical chart review and thus may have had missing data or may have been subject to inaccurate reporting. Finally, in the absence of contact tracing metadata, we were unable to assess transmission and secondary attack rates from vaccinated persons to exposed contacts.

In summary, our results reveal that vaccine breakthrough infections are overrepresented by immune-evading variants such as Gamma[11,38] and Delta[11,13,29], probably due to selection pressure in a highly vaccinated community (>71% fully vaccinated as of early August 2021), and that high-titre symptomatic post-vaccination infections may be a key contributor to viral spread. Waning immunity resulting in decreased effectiveness of the vaccine in preventing symptomatic infection over time[39], relaxation of COVID-19 restrictions and complacency due to 'pandemic fatigue', and in particular, the emergence of the Delta variant with both higher infectivity and antibody resistance[13,21,29,40,41] may explain the steep rise in COVID-19 cases in San Francisco County (Fig. 1)[42] and nationwide[25] in July–August 2021. Targeted booster vaccinations to increase protective neutralizing antibody levels against antibody-resistant variants[33,36], potentially guided by monitoring of immune correlates of vaccine efficacy[36], will probably be needed in the near future to control viral spread in the community.

Methods

Human sample collection, ethics statement and public health surveillance data. Remnant nasopharyngeal and/or oropharyngeal samples and plasma samples from laboratory confirmed SARS-CoV-2 positive patients were retrieved from the UCSF Clinical Laboratories and stored in a biorepository until processed. Remnant samples were biobanked and retrospective medical chart reviews for relevant clinical and demographic metadata were performed under a waiver of consent and according to protocols approved by the UCSF Institutional Review Board (protocol numbers 10-01116 and 11-05519).

De-identified samples from community COVID-19 testing were obtained from Color Genomics Laboratory as part of a research collaboration. Vaccine breakthrough data corresponding to the de-identified samples from Color Genomics were obtained from the San Francisco Department of Public Health. Approval for sequencing and analysis of these de-identified samples and metadata was obtained from the UCSF Institutional Review Board (protocol number 11-05519).

Data regarding the number of COVID-19 cases in San Francisco County during the study period per 100,000 population and the percentage of the eligible persons in the county who were vaccinated were obtained from publicly available records[42,43].

Viral whole-genome sequencing. For primary nasopharyngeal and/or oropharyngeal swab samples from UCSF hospitals and clinics, remnant samples collected in UTM/VTM were diluted with DNA/RNA shield (Zymo Research, R1100-250) in a 1:1 ratio (100 μl primary sample + 100 μl shield). The Omega BioTek MagBind Viral DNA/RNA Kit (Omega Biotek, M6246-03) and the

KingFisher Flex Purification System with a 96 deep-well head (ThermoFisher, 5400630) were then used for viral RNA extraction. For mid-turbinate nasal swab samples sent to Color Genomics for commercial laboratory testing, dry swabs were collected and transported to the laboratory with no added media. At the laboratory, the swabs were resuspended in 1.3 ml lysis buffer and RNA was extracted using the Chemagic 360 system (Perkin-Elmer). Remnant RNA was then aliquoted for viral whole-genome sequencing.

Extracted RNA was reverse transcribed to complementary DNA and tiling multiplexed amplicon PCR was performed using SARS-CoV-2 primers version 3 according to a published protocol[44]. Adapter ligation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, E7645L). Libraries were barcoded using NEBNext Multiplex Oligos for Illumina (96 unique dual-index primer pairs) (New England Biolabs, E6440L) and purified with AMPure XP (Beckman-Coulter, A63881). Amplicon libraries were then sequenced on either Illumina NextSeq 550 or Novaseq 6000 as 1×300 single-end reads (300 cycles).

SARS-CoV-2 viral genome assembly and analyses. SARS-CoV-2 viral genome reads were assembled and variants were identified using an in-house bioinformatics pipeline as previously described[45]. BCL files generated by Illumina sequencers (NextSeq 550 or NovaSeq 6000) were simultaneously demultiplexed and converted to FASTQ files. Raw FASTQ files were first screened for SARS-CoV-2 sequences using BLASTn (BLAST + package 2.9.0) and were aligned against the Wuhan-Hu-1 SARS-CoV-2 reference genome (National Center for Biotechnology Information (NCBI) GenBank accession number NC_045512.2). Reads containing adapters, the ARTIC primer sequences, and low-quality reads were filtered using BBDuk (version 38.87), and then mapped to the NC_045512.2 reference genome using BBMap (version 38.87). Variants were called with CallVariants (version 38.87) and a depth cutoff of 5 was used to generate the final assembly. Pangolin software (version 3.0.2) was used to identify the lineage[46]. Using a custom in-house script (code available at Zenodo, doi: 10.5281/zenodo.5207242), consensus FASTA files generated by the genome assembly pipeline were scanned to confirm the presence/absence of resistance-associated (L452R, L452Q, E484K and/or F490S)[10–14], and infectivity-associated (L452R/Q, N501Y/T and/or F490S) mutations[10–14,28]. Only genomes with defined lineages were included in this analysis. Phylogenetic analyses were performed using Nextstrain v9 (cli version 3.0.3)[47], which runs the Augur (version 13.0.2) bioinformatics pipeline consisting of MAFFT v7.453[48] for alignments, IQTREE v.1.6 for estimating maximum likelihood phylogenies[49], TreeTime (version 0.8.4) for dating and ancestral inference[50], and Auspice (version 2.31.0) for tree visualization. Multiple sequence alignment of SARS-CoV-2 genomes was performed using the MAFFT aligner v7.388[48] as implemented in Geneious v11.1.5[51].

RT–qPCR and viral load analysis. The TaqPath COVID-19 Combo kit (ThermoFisher) was used to determine Ct values. This multiplex real-time RT–qPCR assay detects the nucleoprotein (N) gene, spike (S) gene, and orf1ab genes. For simplicity, only the N gene Ct value was used for quantitative analysis of RNA viral loads in this study. A standard curve was generated by serially diluting known concentrations of SARS-CoV-2 positive control in triplicates and the N gene Ct values of each concentration were determined. The log of each known concentration was plotted against the Ct value. The data were fitted to a regression curve and the correlation coefficient was calculated. The viral load measurements in copies per ml were interpolated from the derived standard curve.

Antibody assays. SARS-CoV-2-specific antibodies were determined using the Abbott ARCHITECT SARS-CoV-2 IgG (N-based), AdviseDx SARS-CoV-2 IgM (spike receptor-binding domain (RBD)-based), and AdviseDx SARS-CoV-2 IgG II (spike RBD-based) tests according to the manufacturer's specifications.

CPE endpoint neutralization assays using a VOC lineage virus. CPE endpoint neutralization assays were done following the limiting dilution model[52] and using P1 stocks of D614G-carrying control, B.1.1.7, B.1.617.2, B.1.429, B.1.351 and P.1 lineages. Convalescent patient plasma was diluted 1:10 and heat inactivated at 56 °C for 30 min. Serial 2-fold dilutions of plasma were made in BSA-PBS. Plasma dilutions were mixed with 100 TCID50 of each virus diluted in BSA-PBS at a 1:1 ratio (160 μl plasma dilution and 160 μl virus input) and incubated for 1 h at 37 °C. Final plasma dilutions in plasma–virus mixtures ranged from 1:100 to 1:12,800. Plasma–virus mixtures (100 μl) were inoculated on confluent monolayer of Vero-81 cells in 96-well plates in triplicate and incubated at 37 °C with 5% CO_2 incubator. After incubation, 150 μl MEM containing 5% FCS was added to the wells, and plates were incubated at 37 °C with 5% CO_2 until consistent CPE was seen in virus control (no neutralizing plasma added) wells. Positive and negative controls were included as well as cell control wells and a viral back titration to verify TCID50 viral input. Individual wells were scored for CPE as having a binary outcome of 'infection' or 'no infection', and the IC_{50} was calculated using the Spearman–Karber method. All steps were done in a Biosafety Level 3 lab using approved protocols.

Statistical analyses. Statistical analyses were performed using Python scipy package (version 1.5.2) and rstatix package (version 0.7.0) in R (version 4.0.3).

For comparisons of the mean Ct values, significance testing was done using Welch's *t*-test as implemented in Python (version 3.7.10). Fisher's Exact test was used to assess the association of demographics and clinical variables with vaccination status. Box-and-whisker and swarm plots were generated using Python matplotlib (version 3.3.2) and seaborn (version 0.11.0) packages. All statistical tests were conducted as two-sided at the 0.05 significance level.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Assembled SARS-CoV-2 genomes in this study have been uploaded to GISAID (accession numbers in Supplementary Table) and can be visualized in NextStrain[47]. FASTA files, Newick phylogenetic tree files and the Supplementary Table have been uploaded to a Zenodo data repository (https://doi.org/10.5281/zenodo.5207242).

Code availability
Custom scripts and code for data analyses and visualization are available in a Zenodo data repository (https://doi.org/10.5281/zenodo.5207242).

References
1. Baden, L. R. et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021).
2. Chemaitelly, H. et al. mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar. *Nat. Med.* https://doi.org/10.1038/s41591-021-01446-y (2021).
3. Haas, E. J. et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* **397**, 1819–1829 (2021).
4. Pilishvili, T. et al. Interim estimates of vaccine effectiveness of Pfizer-BioNTech and Moderna COVID-19 vaccines among health care personnel – 33 U.S. Sites, January–March 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 753–758 (2021).
5. Tang, L. et al. Asymptomatic and symptomatic SARS-CoV-2 infections after BNT162b2 vaccination in a routinely screened workforce. *JAMA* **325**, 2500–2502 (2021).
6. Tenforde, M. W. et al. Effectiveness of Pfizer-BioNTech and Moderna vaccines against COVID-19 among hospitalized adults aged ≥65 Years – United States, January–March 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 674–679 (2021).
7. Thompson, M. G. et al. Prevention and attenuation of Covid-19 with the BNT162b2 and mRNA-1273 vaccines. *N. Engl. J. Med.* **385**, 320–329 (2021).
8. Levine-Tiefenbrun, M. et al. Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nat. Med.* **27**, 790–792 (2021).
9. McEllistrem, M. C., Clancy, C. J., Buehrle, D. J., Lucas, A. & Decker, B. K. Single dose of an mRNA severe acute respiratory syndrome Coronavirus 2 (SARS-Cov-2) vaccine is associated with lower nasopharyngeal viral load among nursing home residents with asymptomatic Coronavirus Disease 2019 (COVID-19). *Clin. Infect. Dis.* **73**, e1365–e1367 (2021).
10. Deng, X. et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell* **184**, 3426–3437.e8 (2021).
11. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2523 (2021).
12. Motozono, C. et al. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* **29**, 1124–1136 e1111 (2021).
13. Planas, D. et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* https://doi.org/10.1038/s41586-021-03777-9 (2021).
14. Wang, R., Chen, J., Gao, K. & Wei, G. W. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries. *Genomics* **113**, 2158–2170 (2021).
15. Magalis, B. R. et al. SARS-CoV-2 infection of BNT162b2(mRNA)-vaccinated individuals is not restricted to variants of concern or high-risk exposure environments. Preprint at medRxiv 10.1101/2021.05.19.21257237 (2021).
16. Bergwerk, M. et al. Covid-19 breakthrough infections in vaccinated health care workers. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2109072 (2021).
17. Brown, C. M. et al. Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings - Barnstable County, Massachusetts, July 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 1059–1062 (2021).
18. Dougherty, K., Mannell, M., Naqvi, O., Matson, D. & Stone, J. SARS-CoV-2 B.1.617.2 (Delta) variant COVID-19 outbreak associated with a gymnastics facility - Oklahoma, April–May 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 1004–1007 (2021).
19. Farinholt, T. et al. Transmission event of SARS-CoV-2 delta variant reveals multiple vaccine breakthrough infections. *BMC Med.* **19**, 255 (2021).
20. Kustin, T. et al. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals. *Nat. Med.* https://doi.org/10.1038/s41591-021-01413-7 (2021).
21. Lopez Bernal, J. et al. Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N. Engl. J. Med.* **385**, 585–594 (2021).
22. McEwen, A. E. et al. Variants of concern are overrepresented among post-vaccination breakthrough infections of SARS-CoV-2 in Washington State. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciab581 (2021).
23. Olsen, R. J. et al. Trajectory of growth of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) variants in Houston, Texas, January through May 2021, based on 12,476 genome sequences. *Am. J. Pathol.* **191**, 1754–1773 (2021).
24. Vignier, N. et al. Breakthrough infections of SARS-CoV-2 Gamma variant in fully vaccinated gold miners, French Guiana, 2021. *Emerg. Infect. Dis.* **27**, 2673–2676 (2021).
25. Gangavarapu, K. et al. *outbreak.info* (2020, accessed 7 April 2021); https://outbreak.info/
26. Washington, N. L. et al. Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* **184**, 2587–2594.e7 (2021).
27. CDC. COVID-19 vaccine breakthrough infections reported to CDC – United States, January 1–April 30, 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 792–793 (2021).
28. Fiorentini, S. et al. First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. *Lancet Infect. Dis.* **21**, e147 (2021).
29. *Delta Variant: What We Know about the Science* (US Centers for Disease Control and Prevention, accessed 11 August 2021); https://www.cdc.gov/coronavirus/2019-ncov/variants/delta-variant.html
30. Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* https://doi.org/10.1126/science.abg3055 (2021).
31. Layan, M. et al. Impact of BNT162b2 vaccination and isolation on SARS-CoV-2 transmission in Israeli households: an observational study. Preprint at *medRxiv* https://doi.org/10.1101/2021.07.12.21260377 (2021).
32. de Gier, B. et al. Vaccine effectiveness against SARS-CoV-2 transmission and infections among household and other close contacts of confirmed cases, the Netherlands, February to May 2021. *Euro Surveill.* https://doi.org/10.2807/1560-7917.ES.2021.26.31.2100640 (2021).
33. Harris, R. J. et al. Effect of vaccination on household transmission of SARS-CoV-2 in England. *N. Engl. J. Med.* **385**, 759–760 (2021).
34. Cobey, S., Larremore, D. B., Grad, Y. H. & Lipsitch, M. Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination. *Nat. Rev. Immunol.* **21**, 330–335 (2021).
35. *Interim Public Health Recommendations for Fully Vaccinated People* (US Centers for Disease Control and Prevention, accessed 11 August 2021); https://www.cdc.gov/coronavirus/2019-ncov/vaccines/fully-vaccinated-guidance.html
36. Gilbert, P. B. et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy trial. Preprint at *medRxiv* https://doi.org/10.1101/2021.08.09.21261290 (2021).
37. Tada, T. et al. Comparison of Neutralizing Antibody Titers Elicited by mRNA and Adenoviral Vector Vaccine against SARS-CoV-2 Variants. Preprint at *bioRxiv*, https://doi.org/10.1101/2021.07.19.452771 (2021).
38. Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021).
39. Thomas, S. J. et al. Six month safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. Preprint at *medRxiv* https://doi.org/10.1101/2021.07.28.21261159 (2021).
40. Puranik, A. et al. Comparison of two highly-effective mRNA vaccines for COVID19 during periods of Alpha and Delta variant prevalence. Preprint at *medRxiv* https://doi.org/10.1101/2021.08.06.21261707 (2021).
41. Scobie, H. M. et al. Monitoring incidence of COVID-19 cases, hospitalizations, and deaths, by vaccination status – 13 U.S. Jurisdictions, April 4–July 17, 2021. *MMWR Morb. Mortal. Wkly Rep.* **70**, 1284–1290 (2021).
42. *Covid-19 Cases and Deaths in San Francisco* (City and County of San Francisco, accessed 11 August 2021); https://sf.gov/data/covid-19-cases-and-deaths
43. *U.S. COVID Risk & Vaccine Tracker* (CovidActNow, accessed 11 October 2021); https://covidactnow.org/us/california-ca/county/san_francisco_county/?s=24131326
44. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
45. Deng, X. et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science* **369**, 582–587 (2020).
46. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol* **5**, 1403–1407 (2020).

47. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
48. Katoh, K. & Standley, D. M. MAFFT: iterative refinement and additional methods. *Methods Mol. Biol.* **1079**, 131–146 (2014).
49. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
50. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
51. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
52. Wang, S., Sakhatskyy, P., Chou, T. H. & Lu, S. Assays for the assessment of neutralizing antibody activities against Severe Acute Respiratory Syndrome (SARS) associated coronavirus (SCV). *J. Immunol. Methods* **301**, 21–30 (2015).

Author contributions

C.Y.C., S.P., S.T. and D.S. conceived and designed the study. C.Y.C. and V.S. coordinated the sequencing efforts and laboratory studies. V.S., M-K.M., A.S-G., E.T., B.W., D.W., C.W., Y.Z., N.B., K.R.R., D.R.G., X.D. and E.F. performed experiments. C.Y.C. and V.S. performed genome assembly, viral mutation and phylogenetic analyses. C.Y.C., V.S., M.-K.M., A.S.-G., A.S.G., K.T.H., M.S., N.B., J.H.Jr and C.H. analysed data. V.S., A.S.-G., A.S.G., E.T., N.B., A.Z., D.W., S.T., D.A.W., K.R.R., J.S. and S.M. collected samples. C.Y.C. and V.S. wrote the manuscript and prepared the figures. C.Y.C., V.S., M.K.M., N.B., B.W., D.W., C.W., Y.Z., X.D., J.H.Jr, C.H., D.A.W. and S.T. edited the manuscript. C.Y.C. and V.S. revised the manuscript. All authors read the manuscript and agreed to its contents.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-021-01041-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-021-01041-4.

Correspondence and requests for materials should be addressed to Charles Y. Chiu.

Peer review information *Nature Microbiology* thanks Elitza Theel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Characteristic		Fully vaccinated breakthrough* cases (#)	Fully vaccinated breakthrough* cases (%)	Unvaccinated cases (#)	Unvaccinated cases (%)	Other cases** (#)	Other cases** (%)	All cases	p-value***	Odds ratio [95% confidence interval]
# of samples		125	9.1	1169	85.1	79	5.8	1373		
# of samples with identifiable lineage		109	11.5	790	83.6	46	4.9	945		
Antibody-resistant mutations	Resistant	76	77.5	329	47.7				**1.96E-08**	3.78 [2.27 - 6.55]
	Nonresistant	22	22.4	361	52.3					
Total assessed*		98	100	690	100					
Infectivity mutations	More infectious	83	84.7	542	76.8				0.092	1.67 [0.93 - 3.21]
	Less infectious	15	15.3	164	23.2					
Total assessed#		98	100	706	100					

*The use of the term "breakthrough" is used here and throughout this paper in compliance with the current CDC definition[29]; current vaccines are designed to prevent severe COVID-19 disease leading to hospitalizations and death, and not to prevent mild or asymptomatic infection.

#Total assessed refers to the number of sequences for which the presence or absence of antibody-resistant or infectivity mutations can be reliably determined based on lineage classification and/or the presence or absence of key single nucleotide polymorphisms (SNPs) associated with these coding mutations.

**Other cases include partially vaccinated (n=44), unknown vaccination status (n=12), and cases confirmed to be SARS-CoV-2 RT-PCR false positives (n=23) after further investigation.

***Two-tailed Fisher's Exact test was used to determine p-values.

Extended Data Fig. 1 | Characteristics of vaccine breakthrough and unvaccinated cases.
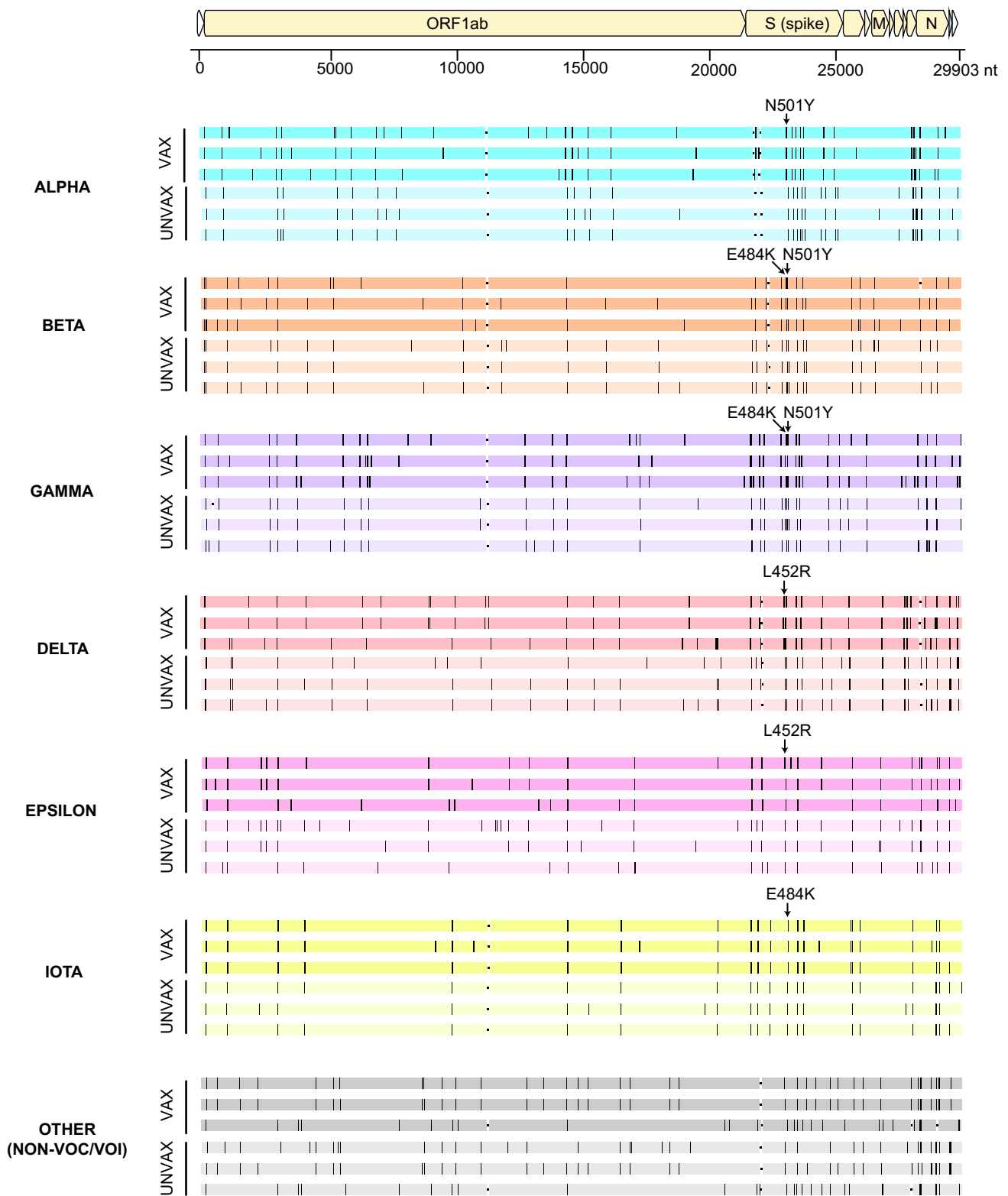
Extended Data Fig. 2 | Mutational analysis of 42 representative SARS-CoV-2 genomes. A multiple sequence alignment of 42 representative viral genomes consisting of the Alpha, Beta, Gamma, Delta, Epsilon, Iota, and D614G-carrying non-VOC/non-VOI lineages and including 3 genomes each from unvaccinated and unvaccinated cases per lineage was performed. Mutations including single nucleotide polymorphisms (SNPs) and deletions (black lines) are shown using the Wuhan Hu-1 genome (NC_045512.2) as a reference. Arrows denote the locations of key mutations associated with decreased antibody neutralization (L452R and E484K) and increased infectivity (L452R and N501Y). The alignments are color-coded by lineage, with darker shades corresponding to genomes from vaccinated cases and lighter shades to genomes from unvaccinated cases.

FULLY VACCINATED   **UNVACCINATED**   **PARTIALLY VACCINATED**
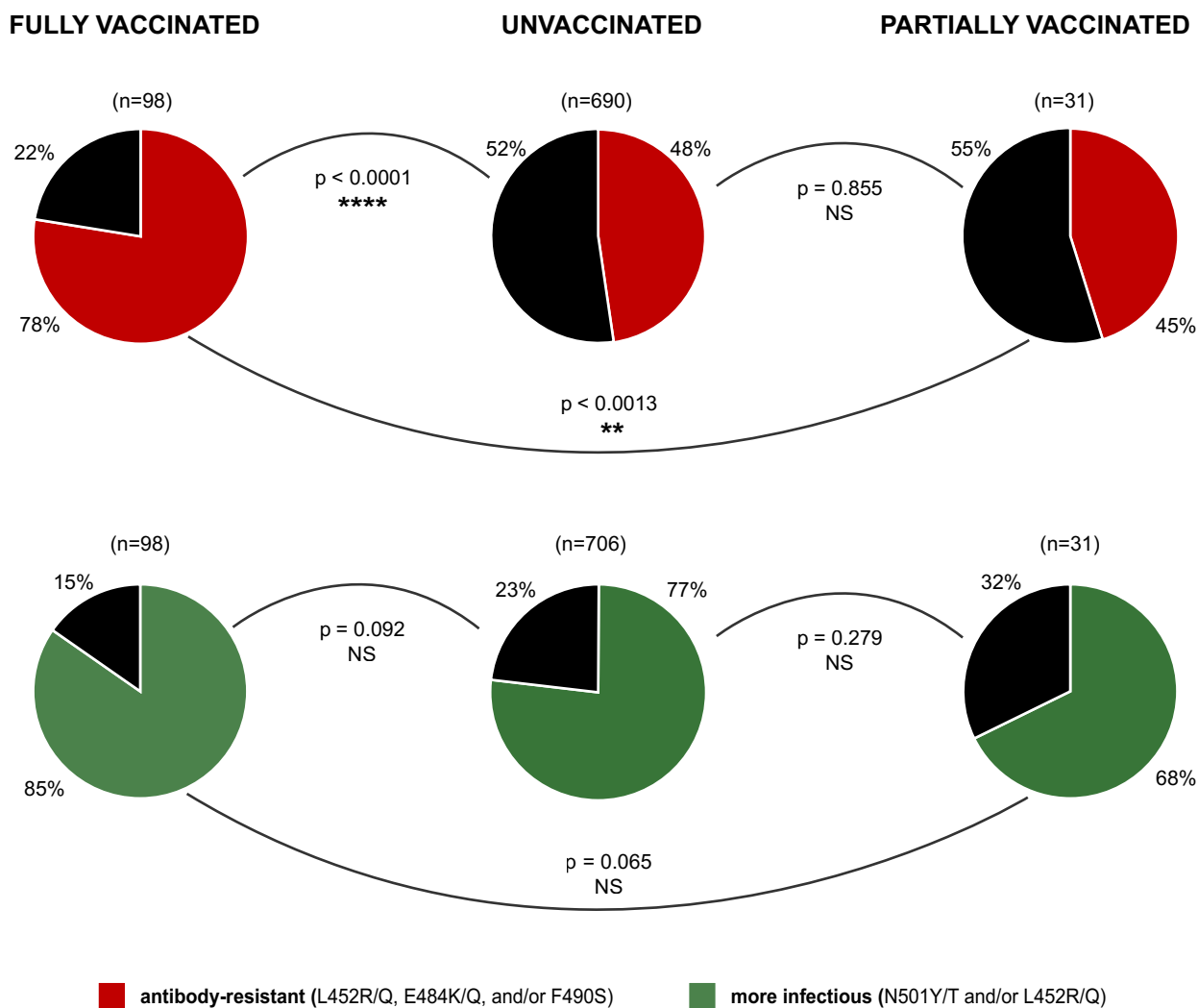
Extended Data Fig. 3 | Proportion of SARS-CoV-2 genomes carrying mutations associated with antibody resistance and/or increased infectivity in cases from UCSF hospitals and clinics. Pie charts showing the proportions of SARS-CoV-2 genomes carrying identifiable mutations associated with antibody resistance (L452R/Q, E484K/Q, and/or F490S) and increased infectivity (N501Y/T, L452R/Q, and/or F490S) in fully vaccinated (left), unvaccinated (middle), and partially vaccinated (right) cases. The pie charts are shaded according to genomes carrying ≥1 mutation associated with antibody resistance (red), ≥1 mutation associated with increased infectivity (green), or neither type of mutation (black). Fisher's Exact test (two-tailed) was used to calculate p-values. Abbreviations: NS, non-significant; *, $p < 0.05$; **, $p < 0.01$.

$$y = -1.757\ln(x) + 37.341$$
$$R^2 = 0.9995$$
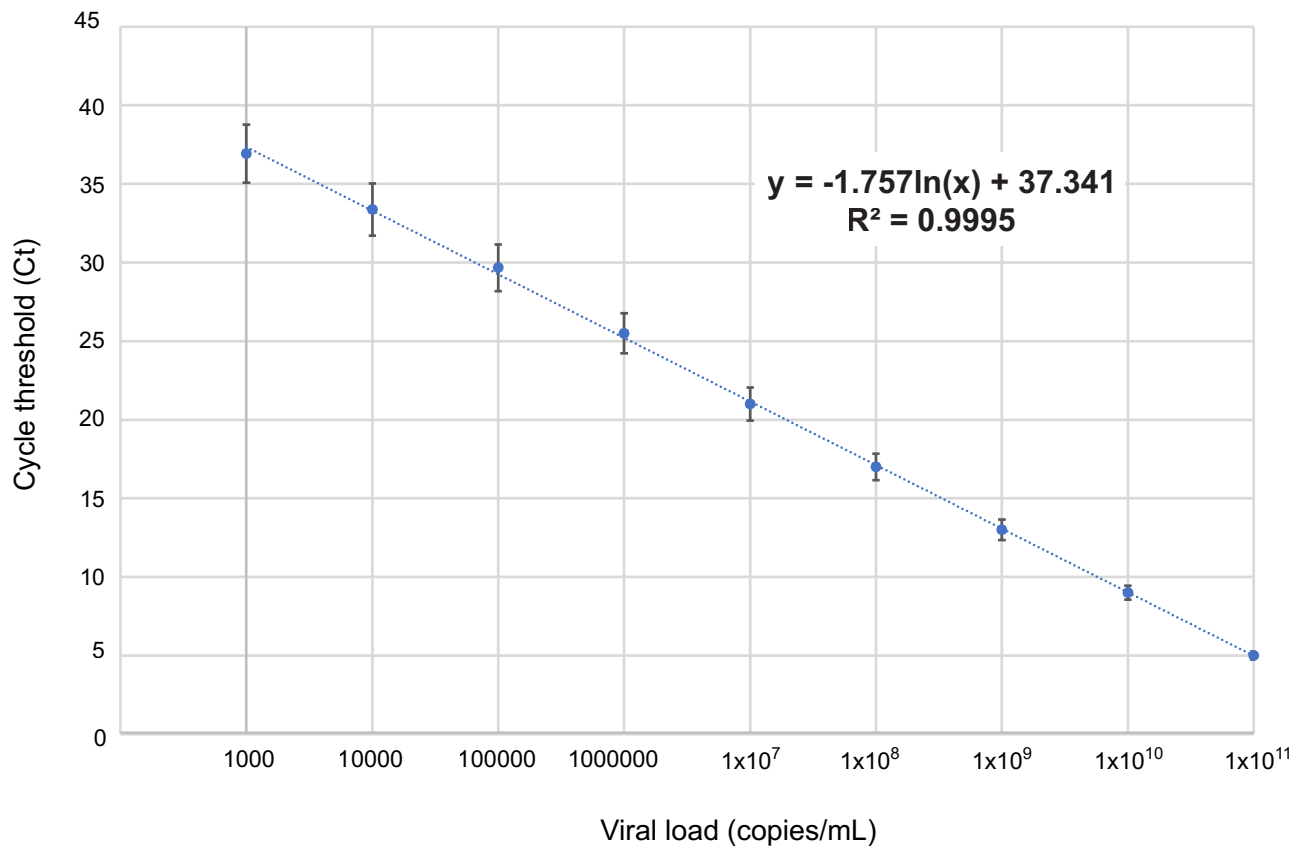
Extended Data Fig. 4 | Quantitative RT-PCR standard curve. A semi-log regression line plot of the Ct value vs. the log of input nucleic acid (copies/ mL). A dilution series of known template concentrations was used to determine the initial target measurements. The R^2 value represents the calculated correlation coefficient. The error bars indicate the standard error of the mean (SEM) determined from the results of three replicates of RT-PCR reactions.

Patient*	# of days from completion of vaccine doses to positive COVID-19 diagnostic test	vaccine type
1	37	Pfizer
2	49	Pfizer
3	48	Pfizer
4	23	Pfizer
5	44	Pfizer
6	17	Pfizer
7	59	Pfizer
8	42	Moderna
9	42	Moderna
10	20	not specified
11	94	Pfizer
12	83	Pfizer
13	27	JnJ
14	84	Pfizer
15	30	JnJ
16	78	Moderna
17	117	Pfizer
18	18	JnJ
19	69	Pfizer
20	81	Moderna
21	129	Pfizer
22	116	Pfizer
23	47	JnJ
24	31	Moderna
25	118	Moderna
26	140	Pfizer
27	58	Moderna
28	not specified	not specified
29	102	Pfizer
30	133	Moderna
31	36	not specified
32	123	Moderna
33	15	Moderna
34	104	Pfizer
35	114	Pfizer
36	131	Moderna
37	81	Pfizer
38	112	Pfizer
39	127	Moderna

*For the remaining 86 vaccine breakthrough cases in the study, identified by community testing from Color Genomics, each individual had received the Pfizer, Moderna, or JnJ vaccine, although details regarding the exact vaccine that was administered are not available.

Extended Data Fig. 5 | Vaccine received and number of days from completion of vaccine to COVID-19 infection for 39 confirmed vaccine breakthrough cases* in the study.
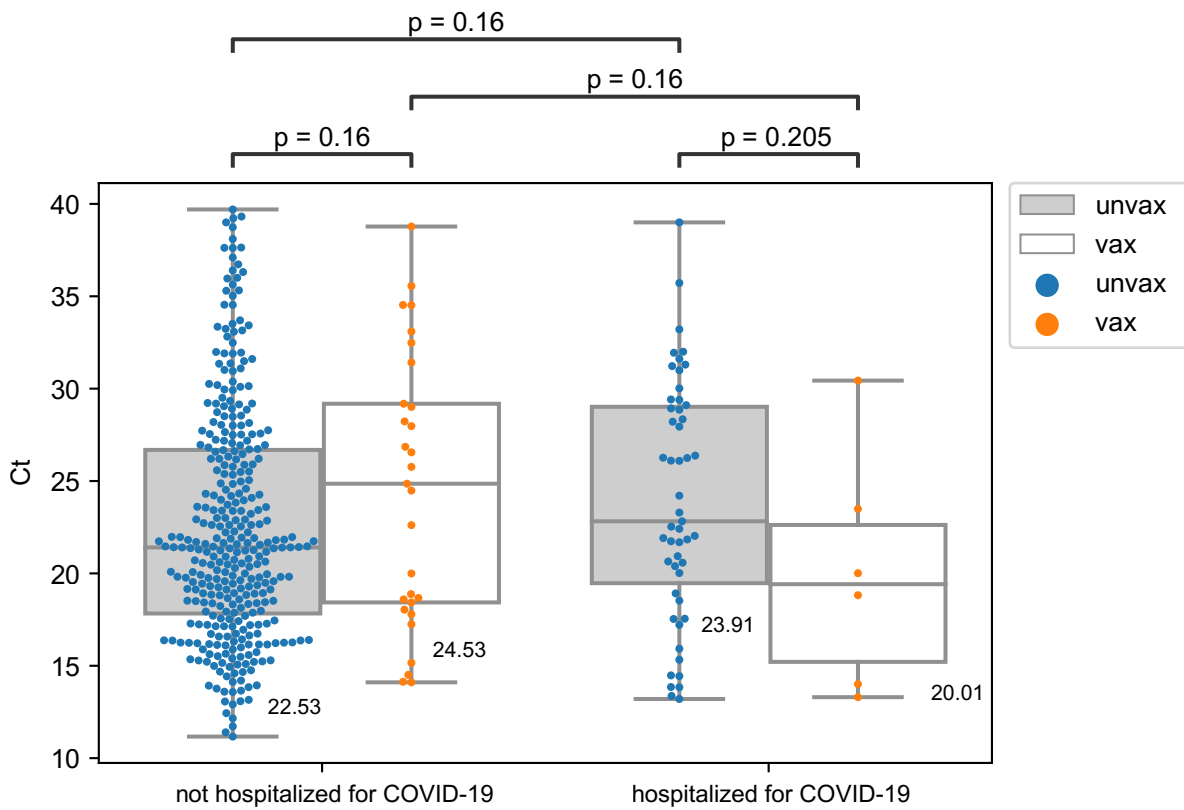
Extended Data Fig. 6 | Comparison of viral loads between vaccinated and unvaccinated cases based on hospitalization status. Grouped box-and-whisker plots and swarm plots showing the differences in mean cycle threshold (Ct) values between vaccinated (n = 32) and unvaccinated (n = 362) cases overall given hospitalization due to COVID-19 (n = 57) or nonhospitalized status (n = 337). There were no significant differences in viral loads (inversely proportional to the Ct value) in the pairwise comparisons. For the box-and-whisker plots, the box outlines denote the interquartile ratio (IQR), the solid line inside the box denotes the median, the dotted line inside the box denotes the mean (μ) Ct value, and the whiskers outside the box extend to the minimum and maximum fold enrichment points. Welch's t-test was used for significance testing.

Patient #	Infecting lineage	Vaccine received	Immunocompromised?	Clinical presentation of breakthrough	Plasma sample collection day*	SARS-CoV-2 IgG Ab (spike RBD)	SARS-CoV-2 IgG Ab (nucleoprotein)	SARS-CoV-2 IgM Ab (spike)	Antibody neutralization						Interpretation (reason for vaccine failure)
									D614G (control)	B.1.1.7 (alpha)	B.1.351 (beta)	P.1 (gamma)	B.1.617.2 (delta)	B.1.429 (epsilon)	
P1	non-VOC (B.1)	Pfizer/BioNTech BNT162b2	yes, immunosuppression status post liver tranplant	hospitalized with COVID-19 pneumonia	-36 days	− (0.0)	− (1.0)	− (0.02)	<100	<100	<100	<100	<100	<100	no Ab response to vaccine, likely due to immunocompromised state
					+9 days	+ (157.62)	− (0.41)	+ (1.58)	<100	<100	<100	<100	<100	<100	
					+17 days	+ (14258.34)	+ (4.27)	+ (17.11)	9051	898	356	566	898	713	
					+28 days	+ (52626.99)	+ (8.12)	+ (5.41)	9051	898	224	449	566	1425	
P2	B.1.1.7 (alpha)	Pfizer/BioNTech BNT162b2	yes, diffuse large B-cell lymphoma status post CAR-T immunotherapy	hospitalized with COVID-19 pneumonia	+6 days	− (55.3)	− (0.23)	− (0.01)	<100	<100	<100	<100	<100	<100	no Ab response to vaccine, likely due to immunocompromised state
					+9 days	− (41.9)	− (0.26)	− (0.01)	<100	<100	<100	<100	<100	<100	
P3	B.1.1.7 (alpha)	Moderna mRNA-1273	yes, immunosuppression status post kidney transplant	hospitalized with COVID-19 pneumonia	+11 days	+ (13889.68)	+ (4.94)	+ (58.97)	11404	11404	2263	11404	3592	1796	indeterminate (no pre-breakthrough sample available)
					+18 days	+ (23622.01)	+ (5.02)	+ (45.66)	4525	4525	2263	4525	1425	356	
P4#	B.1.617.2 (delta)	Johnson and Johnson JNJ-78436735.	no	asymptomatic	+2 days	+ (13889.68)	− (0.88)	− (0.73)	11404	5702	898	11404	2851	2263	failure to mount adequate Ab response to vaccine; decreased Ab response to delta variant
P5	P.1 (gamma)	Moderna mRNA-1273	yes, immunosuppression status post lung transplant	symptomatic (fever, malaise, and myalgia)	+3 days	− (0.0)	− (0.04)	− (0.03)	<100	<100	<100	<100	<100	<100	no Ab response to vaccine, likely due to immunocompromised state
					+5 days	− (0.0)	− (0.03)	− (0.03)	<100	<100	<100	<100	<100	<100	

*Relative to day of first positive respiratory swab PCR

#Patient had a prior infection from COVID-19

Abbreviations: PCR, polymerase chain reaction; RBD, receptor binding domain; IgG, immunoglobulin G; IgM, immunoglobulin M; Ab, antibody.

Extended Data Fig. 7 | Qualitative and neutralizing antibody studies in vaccine breakthrough patients.

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based.

All Submitters of data may be contacted directly via www.gisaid.org

Authors are sorted alphabetically.

Accession ID	Originating Laboratory	Submitting Laboratory	Authors
EPI_ISL_2797922, EPI_ISL_2798047, EPI_ISL_2825482, EPI_ISL_2826335, EPI_ISL_2826349, EPI_ISL_2826382, EPI_ISL_2853656, EPI_ISL_2853657, EPI_ISL_2853659, EPI_ISL_2853759, EPI_ISL_2855068, EPI_ISL_2858305, EPI_ISL_2858366, EPI_ISL_2881874, EPI_ISL_2882006, EPI_ISL_2882337, EPI_ISL_2882772, EPI_ISL_2882814, EPI_ISL_2882846, EPI_ISL_2889906, EPI_ISL_2889908, EPI_ISL_2890074, EPI_ISL_2890075, EPI_ISL_2890085, EPI_ISL_2890372, EPI_ISL_2890765, EPI_ISL_2890830, EPI_ISL_2891048, EPI_ISL_2891049, EPI_ISL_2891144, EPI_ISL_2891405, EPI_ISL_2891406, EPI_ISL_2892133, EPI_ISL_2892150, EPI_ISL_2892192, EPI_ISL_2921753, EPI_ISL_2921754, EPI_ISL_2921755, EPI_ISL_2921889, EPI_ISL_2922063, EPI_ISL_2922128, EPI_ISL_2923411, EPI_ISL_2923673, EPI_ISL_2934163, EPI_ISL_2934196, EPI_ISL_2977148, EPI_ISL_2977426, EPI_ISL_3244299, EPI_ISL_3669511, EPI_ISL_3671155, EPI_ISL_3671297, EPI_ISL_3671412, EPI_ISL_3695196, EPI_ISL_4510994, EPI_ISL_4511029, EPI_ISL_4511308, EPI_ISL_4511324, EPI_ISL_4511328, EPI_ISL_4511334	CDPH VBL	California Department of Public Health	CDPH-COVIDNet; Emily Smith on behalf of CDPH-COVIDNet and UCI Genome Sciences Center/GHTF; Emily Smith on behalf of CDPH-COVIDNet and UCSD EXCITE lab; SwabSeq/UCLA; UCLA Technology Center for Genomics & Bioinformatics; UCSD EXCITE lab; UCSF Center for Advanced Technology
see above			
EPI_ISL_2234121, EPI_ISL_2234122, EPI_ISL_2234123, EPI_ISL_2234124, EPI_ISL_2234125, EPI_ISL_2234313, EPI_ISL_2234333, EPI_ISL_2537412, EPI_ISL_2537451, EPI_ISL_2537452, EPI_ISL_2537453, EPI_ISL_2842807, EPI_ISL_2861257, EPI_ISL_2861258, EPI_ISL_2892405, EPI_ISL_2892421, EPI_ISL_2892422, EPI_ISL_2892454, EPI_ISL_2892455, EPI_ISL_2892456, EPI_ISL_2892511, EPI_ISL_2892517, EPI_ISL_2892537, EPI_ISL_2892538, EPI_ISL_2892564, EPI_ISL_2892539, EPI_ISL_3536689, EPI_ISL_3536715, EPI_ISL_3536743, EPI_ISL_3536744, EPI_ISL_3536787, EPI_ISL_3536804, EPI_ISL_3536812, EPI_ISL_3536874, EPI_ISL_3536880, EPI_ISL_3536899, EPI_ISL_3536931, EPI_ISL_3536932, EPI_ISL_3537009, EPI_ISL_3537043, EPI_ISL_3537049, EPI_ISL_3537058, EPI_ISL_3537095	California Department of Public Health	California Department of Public Health	CDPH IDLB COVIDNet et al
EPI_ISL_3134153	California Department of Public Health Valencia Branch Laboratory	California Department of Public Health	CDPH-COVIDNet; UCSF Center for Advanced Technology
EPI_ISL_2778790, EPI_ISL_2778811, EPI_ISL_2778832, EPI_ISL_2837493, EPI_ISL_2837502, EPI_ISL_2837630, EPI_ISL_2837874	California Department of Public Health Valencia Branch Laboratory (CDPH VBL)	California Department of Public Health	CDPH-COVIDNet; UCLA Technology Center for Genomics & Bioinformatics
EPI_ISL_1824301, EPI_ISL_1824302, EPI_ISL_1824303, EPI_ISL_1824304, EPI_ISL_1824307, EPI_ISL_1824309, EPI_ISL_1824310, EPI_ISL_1824313, EPI_ISL_1824322, EPI_ISL_1824323, EPI_ISL_1824324, EPI_ISL_1824329, EPI_ISL_1824330, EPI_ISL_1824331, EPI_ISL_1824332, EPI_ISL_1824335, EPI_ISL_1824336, EPI_ISL_1824338, EPI_ISL_1824339, EPI_ISL_1824343, EPI_ISL_1824344, EPI_ISL_1824348, … (continued) … EPI_ISL_4262068, EPI_ISL_4262072, EPI_ISL_4262073, EPI_ISL_4262076, EPI_ISL_4262078, EPI_ISL_4262082, EPI_ISL_4262083, EPI_ISL_4262084	Chiu Laboratory, University of California, San Francisco	Chiu Laboratory, University of California, San Francisco	Alicia Sotomayor-Gonzalez; Baolin Wang; Candace Wang; Charles Chiu; Daniel Wong; Dustin Glasner; Venice Servellita; Xianding (Wayne) Deng; Yueyuan Zhang
see above			
EPI_ISL_1824783, EPI_ISL_1824785, EPI_ISL_1824786, … (continued) … EPI_ISL_4300677, EPI_ISL_4300678, EPI_ISL_4300679, EPI_ISL_4300680	Color Genomics	Chiu Laboratory, University of California, San Francisco	Alicia Sotomayor-Gonzalez; Alicia Zhou; Baolin Wang; Candace Wang; Charles Chiu; Daniel Wong; Erika Torres; Madyson Wanner; Noah Brazer; Scott Topper; Venice Servellita; Yueyuan Zhang
see above			
EPI_ISL_1318697, EPI_ISL_1318701, EPI_ISL_1318702, EPI_ISL_1318704, EPI_ISL_1318705, … (continued) … EPI_ISL_2801557, EPI_ISL_2801558, EPI_ISL_2801559, EPI_ISL_2801560, EPI_ISL_2801561, EPI_ISL_2801562, EPI_ISL_2801563, EPI_ISL_3099533	DeRisi Lab, University of California, San Francisco	DeRisi Lab, University of California, San Francisco	Amy Kistler; Anthea Mitchell; Carina Marquez; Diane Havlir; Diane Jones; Douglas Black; Eric Chow; Gabriel Chamie; Genay Pilarowski; IDseq Team; Jackie Martinez; James Peng; Jamin Liu; Jon Jacobo; Lucy Li; Luis Rubio; Manu Vanaerschot; Matthew Laurie; Matthias Hapte-Selassie; Maya Petersen; Patrick Ayscue; Sabrina Mann; Sara Sunshine; Susana Rojas; Susy Rojas; Valerie Tulier-Laiwa; and Joseph DeRisi
EPI_ISL_1060075, EPI_ISL_1060116	Innovative Genomics Institute, UC Berkeley	Innovative Genomics Institute, UC Berkeley	Alison Cling; Haridha Shivram; Liana Lareau; Netravathi Krishnappa; Phil Frankino; Stacia Wyman
EPI_ISL_2009392	Lurie Children's Hospital of Chicago	Northwestern University - Ozer Lab	Egon A. Ozer; Judd F. Hultquist; Lacy M. Simons; Larry K. Kociolek; Michael G. Ison; Ramon Lorenzo-Redondo; Taylor J. Dean; William J. Muller; Xiaotian; Zheng
EPI_ISL_2663528	NYU Langone Health	Departments of Pathology and Medicine, New York University School of Medicine	Adriana Heguy; Christian Marier; Dacia Dimartino; Emily Guzman; Gael Westby; Guiqing Wang; Paolo Cotzia; Paul Zappile; Peter Meyn; Sitharam Ramaswami; Yutong Zhang
EPI_ISL_2009353	Northwestern Memorial Hospital	Northwestern University - Ozer Lab	Chad J. Achenbach; Chao Qi; Egon A. Ozer; Judd F. Hultquist; Lacy M. Simons; Lawrence J. Jennings; Michael G. Ison; Ramon Lorenzo-Redondo; Taylor J. Dean

Extended Data Fig. 8 | Laboratories submitting sequences from San Francisco County to the GISAID reference database from February 1 to June 30, 2021.

nature portfolio

Corresponding author(s):   Charles Chiu

Last updated by author(s):   November 22, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For NP swabs, RNA was extracted using the Omega BioTek MagBind Viral DNA/RNA Kit (Omega Biotek, # M6246-03) and the KingFisherTM Flex Purification System with a 96 deep-well head. For dry nasal swabs, lysis buffer was added and RNA was extracted using the Chemagic 360 system. Extracted RNA was reverse transcribed to complementary DNA and tiling multiplexed amplicon PCR was performed using SARS-CoV-2 primers version 3 according to a published protocol. Adapter ligation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, # E7645L). Libraries were barcoded using NEBNext Multiplex Oligos for Illumina (96 unique dual-index primer pairs) (New England Biolabs, # E6440L) and purified with AMPure XP (Beckman-Coulter, #. Amplicon libraries were then sequenced on either Illumina NextSeq 550 or Novaseq 6000 as 1x300 single-end reads (300 cycles). RT-PCR was performed on all available RNA extracts to determine cycle threshold(Ct).
Data analysis	SARS-CoV-2 viral genome reads were assembled and variants were identified using an in-house bioinformatics pipeline as previously described. BCL files generated by Illumina sequencers (NextSeq 550 or NovaSeq 6000) were simultaneously demultiplexed and converted to FASTQ files. Raw FASTQ files were first screened for SARS-CoV-2 sequences using BLASTn (BLAST+ package 2.9.0) alignment against viral reference genome Wuhan-Hu-1 (NC_045512). Reads containing adapters, the ARTIC primer sequences, and low-quality reads were filtered using BBDuk (version 38.87), and then mapped to the NC_045512 reference genome using BBMap (version 38.87). Variants were called with CallVariants (version 38.87) and a depth cutoff of 5 was used to generate the final assembly. Pangolin software (version 3.0.6) was used to identify the lineage. Using a custom in-house script (doi: 10.5281/zenodo.5207242), consensus FASTA files generated by the genome assembly pipeline were scanned to confirm the presence/absence of resistance-associated (L452R, L452Q, E484K, and/or F490S), and infectivity-associated (L452R/Q, F490S, and/or N501T/Y) mutations. Only genomes with defined lineages were included in the analysis. Phylogenetic analyses were performed using Nextstrain v9 (cli version 3.0.3), which runs the Augur (version 13.0.2) bioinformatics pipeline consisting of MAFFT v7.453 for alignments, IQTREE v.1.6 for estimating maximum likelihood phylogenies, TreeTime (version 0.8.4) for dating and ancestral inference, and Auspice (version 2.31.0) for tree visualization. Multiple sequence alignment of SARS-CoV-2 genomes was performed using the MAFFT aligner v7.388 as implemented in Geneious v11.1.5.

Differences in mean cycle thresholds were calculated and significance was determined using Welch's t-test as implemented in Python (version 3.7.10). A standard curve was used to calculate viral loads.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Assembled SARS-CoV-2 genomes in this study have been uploaded to GISAID (accession numbers in Supplementary Table) and can be visualized in NextStrain. FASTA files, Newick phylogenetic tree files, and supplementary data tables have been uploaded to a Zenodo data repository (doi: 10.5281/zenodo.5207242).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Sample size was determined by the total number of SARS-CoV-2 PCR positive samples that are available and accessible for sequencing. Sample size is deemed adequate based on significant differences between groups. A total of 1373 laboratory-confirmed SARS-CoV-2 positive nasopharyngeal swab samples were sequenced (598 samples were retrieved from the UCSF Clinical Microbiology Laboratory, while 775 samples were from Color Genomics Laboratory). 945 of the 1373 samples had identifiable lineages based on Pangolin algorithm. Longitudinal plasma samples from five SARS-CoV-2 positive cases were obtained from the UCSF Clinical Laboratories.
Data exclusions	No experimental data were excluded. There were inclusion and exclusion criteria for some analyses as described in the manuscript. For instance, only genomes with Pangolin-identified lineages are included in the viral mutation and phylogenetic analyses.
Replication	Viral samples were sequenced only a single time, but sequence analysis was done on two independent sample cohorts, a hospital-based cohort from University of California, San Francisco (UCSF) hospitals and clinics and a community-based cohort from Color Genomics. Using the same analysis pipeline, the two different cohorts showed similar variant distributions.
Randomization	Randomization is not applicable in this study since all available PCR-positive samples were sequenced.
Blinding	Researchers were blinded to vaccination status during data collection and data processing. Samples were unblinded during analysis to determine differences in viral loads, lineage distributions and presence/absence of resistance/infectivity associated mutations.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	ATCC Vero CCL-81
Authentication	The cell line was not authenticated.
Mycoplasma contamination	The cell line tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	none

Human research participants

Policy information about studies involving human research participants

Population characteristics	See Table 1 for a summary of Demographics and Clinical Characteristics and Supplementary Table for a summary of sample characteristics. Among the 1373, only 598 samples (samples collected from UCSF Hospitals and Clinics) had available clinical metadata. See Supplementary Table for summary and additional metadata of sequenced samples used in the study.
Recruitment	598 samples or 43.6% of the total samples were obtained from patients at the University of California San Francisco (UCSF) hospitals and clinics between February 1, 2021 to June 30, 2021. The study only used residual nasal swab samples after a gold-standard diagnostic PCR laboratory testing was performed. 775 samples or 56.4% of the total samples were obtained from Color Genomics Laboratory. Sampling bias may be present because of the two different cohorts but this is mitigated by results showing similar distributions.
Ethics oversight	Remnant nasopharyngeal and/or oropharyngeal (NP/OP) samples and plasma samples from laboratory confirmed SARS-CoV-2 positive patients were retrieved from the UCSF Clinical Laboratories and stored in a biorepository until processed. Remnant samples were biobanked and retrospective medical chart review for relevant clinical and demographic metadata were performed under a waiver of consent and according to protocols approved by the UCSF Institutional Review Board (protocol numbers 10-01116 and 11-05519). De-identified samples from community COVID-19 testing were obtained from Color Genomics Laboratory as part of a research collaboration. Vaccine breakthrough data corresponding to the de-identified samples from Color Genomics were obtained from the San Francisco Department of Public Health. Approval for sequencing and analysis of these de-identified samples and metadata was obtained from the UCSF Institutional Review Board (protocol number 11-05519). All experimental methods followed guidelines established by the Helsinki Declaration.

Note that full information on the approval of the study protocol must also be provided in the manuscript.