Med

CellPress

Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination

Co-circulation in USA

Fraction of infections

Nov 2021 — Study period — Feb 2022

Delta variant

P1000L   T19R 156/157del   I82T

ORF1A   ORF1B   S   M

Omicron variant

P3395H   I1566V   214EPEins N969K A63T

ORF1A   ORF1B   S   M

Sequencing and analysis of 29,719 SARS-CoV-2 samples

CO-INFECTIONS | RECOMBINANTS

Collection & RNA extraction

Sequencing

Visualization

AAF   1.0 / 0.5 / 0.0   5' — 3'

AAF   1.0 / 0.5 / 0.0   Breakpoint   5' — 3'

18 co-infections | **2 recombinants**

Bolze et al. sequenced 29,719 SARS-CoV-2 samples collected during co-circulation of the Delta and Omicron variants in the United States. Co-infections were detected in 18 samples, with one co-infection containing a recombinant viral population. Two distinct clonal Delta-Omicron recombinant infections were also confirmed.

Translation to Patients

Alexandre Bolze, Tracy Basler, Simon White, ..., Magnus Isaksson, William Lee, Shishi Luo

shishi.luo@helix.com

Highlights

Whole-genome sequencing of SARS-CoV-2 when Delta and Omicron variants co-circulated

18 samples (<0.1%) were confirmed to be Delta-Omicron co-infections

Recombination of Delta and Omicron detected in one of the co-infection samples

Infections by clonal Delta-Omicron recombinants was rare, only 2 samples

Med

CellPress

Clinical and Translational Report

Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination

Alexandre Bolze,[1] Tracy Basler,[1] Simon White,[1] Andrew Dei Rossi,[1] Dana Wyman,[1] Hang Dai,[1] Pavitra Roychoudhury,[2] Alexander L. Greninger,[2] Kathleen Hayashibara,[3] Mark Beatty,[4] Seema Shah,[4] Sarah Stous,[4] John T. McCrone,[1] Eric Kil,[1] Tyler Cassens,[1] Kevin Tsan,[1] Jason Nguyen,[1] Jimmy Ramirez,[1] Scotty Carter,[1] Elizabeth T. Cirulli,[1] Kelly Schiabor Barrett,[1] Nicole L. Washington,[1] Pedro Belda-Ferre,[1] Sharoni Jacobs,[1] Efren Sandoval,[1] David Becker,[1] James T. Lu,[1] Magnus Isaksson,[1] William Lee,[1] and Shishi Luo[1,5,6,*]

SUMMARY

Background: Between November 2021 and February 2022, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Delta and Omicron variants co-circulated in the United States, allowing for co-infections and possible recombination events.
Methods: We sequenced 29,719 positive samples during this period and analyzed the presence and fraction of reads supporting mutations specific to either the Delta or Omicron variant.
Findings: We identified 18 co-infections, one of which displayed evidence of a low Delta-Omicron recombinant viral population. We also identified two independent cases of infection by a Delta-Omicron recombinant virus, where 100% of the viral RNA came from one clonal recombinant. In the three cases, the 5′ end of the viral genome was from the Delta genome and the 3′ end from Omicron, including the majority of the spike protein gene, though the breakpoints were different.
Conclusions: Delta-Omicron recombinant viruses were rare, and there is currently no evidence that Delta-Omicron recombinant viruses are more transmissible between hosts compared with the circulating Omicron lineages.
Funding: This research was supported by the NIH RADx initiative and by the Centers for Disease Control Contract 75D30121C12730 (Helix).

INTRODUCTION

Recombination is one way DNA viruses and RNA viruses can evolve to acquire a new combination of mutations.[1] For example, recombinations have played an important role in the evolution of adenoviruses.[2] Recombinations have also played an important role in the evolution of the RNA viruses HIV-1 and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).[3,4] In humans, the analysis of the first 87,695 SARS-CoV-2 genomes shared on GISAID in 2021 identified 225 sequences of likely recombinant origins.[5] However, tracking recombinations for SARS-CoV-2 remains challenging because of the relatively low diversity of the genomes. Moreover, without the underlying sequencing data or orthogonal confirmation, it is difficult to determine whether recombinant sequences are real or due to contamination, technical artifacts, or naturally occurring mutations shared by multiple variants.

The Omicron variant, first detected by scientists in Botswana and South Africa in early November 2021,[6] has rapidly spread across the globe. The Omicron variant

CONTEXT AND SIGNIFICANCE

Viruses can gain new properties by recombination. Recombination between SARS-CoV-2 Delta and Omicron variants can lead to a new set of mutations with unknown impact on transmissibility and severity of the virus. Here, the authors sequenced SARS-CoV-2-positive samples between November 2021 and February 2022, when Delta and Omicron were both present in the United States. After validation, they found 18 samples (<0.1%) infected by the two variants, with evidence that recombination occurred in at least one of these samples. They found only two samples where 100% of the virus was a Delta-Omicron recombinant. The methods and strategy used in this study could be used to track co-infections and new recombinations in the future.

is characterized by a large number of mutations in the spike protein.[6] In the United States, the first Omicron case was reported on December 1, 2021,[7] when the Delta variant was dominant.[8,9] Omicron BA.1 was the dominant Omicron sublineage during the study period, with BA.2 beginning to rise in the United States in February 2022. We hypothesized that during the subsequent period when Delta and Omicron co-circulated, some cases of co-infections would occur and potentially result in the emergence of a new SARS-CoV-2 variant resulting from the recombination of a Delta variant and an Omicron variant. This, in turn, would result in a new combination of mutations with unknown properties.

The aims of this study were to (1) look for cases co-infected with Delta and Omicron variants and (2) study SARS-CoV-2 recombination in humans by analyzing co-infection samples, as well as looking for clonal infections caused by a Delta and Omicron recombinant.

RESULTS

Co-circulation of Delta and Omicron variants in the United States

We sequenced and assigned a lineage to 29,719 samples positive for SARS-CoV-2 collected by the Helix laboratory across the United States (Table S1) between November 22, 2021, and February 13, 2022. These samples came from anterior nasal swabs of different individuals, with one viral sequence assay performed per person infected (similar to a cross-sectional analysis). The large majority of samples were collected at a national retail pharmacy. Samples from San Diego County were collected as part of community testing organized by San Diego County. The individuals tested represented a diverse range of race, age, and gender groupings (details in Table S1). We observed that the Omicron variant quickly grew to explain >99% of cases as of the week of January 17 (Figure 1; Table S2). Delta and Omicron variants therefore co-circulated (each representing >1% of infections) from December 6, 2021, to January 16, 2022, represented by 14,214 sequences in our dataset (Figure 1; Table S2). During that time, the overall number of cases in the United States remained high, above 150,000 new cases per day and above a 7-day case rate of 250 per 100,000 individuals.[10] The possibility of a co-infection by two distinct variants was therefore also high during this period.

Co-infection with Delta and Omicron variants

When a person is infected by two distinct variants/viruses, multiple copies of the full genome of each variant are present in the sample. A fraction (x%) of the total extracted SARS-CoV-2 RNA will come from variant A, and the remaining fraction (100-x%) of the RNA will come from variant B. Sequencing at a high enough coverage will lead to calling mutations that define both variant A and variant B, but each mutation will only be supported by a fraction of the reads overlapping the given position; the mutations specific to variant A should be called with ∼x% of the reads overlapping the position, whereas the mutations specific to variant B should be called with (100-x)% (Figure 2A). In order to identify this co-infection signature, we selected a list of mutations specific to the Delta variant and a list of mutations specific to the Omicron variant (Table S3). All mutations selected had a call (not an "N") in >95% of the samples between November 2021 and February 2022 (Table S3).

We identified 21 samples that were very likely to be co-infected with Delta and Omicron variants by filtering for samples where the median alternative allele fraction is less than 0.85 (STAR methods). For 19 samples (two samples were unable to be retested), we validated the results by RNA re-extraction and re-sequencing. For a few

[1]Helix, San Mateo, CA 94401, USA

[2]Department of Laboratory Medicine and Pathology, University of Washington Medical Center, Seattle, WA 98195, USA

[3]Thermo Fisher Scientific, South San Francisco, CA 94080, USA

[4]County of San Diego Health and Human Services, San Diego, CA 92110, USA

[5]Twitter: @shishiluo

[6]Lead contact

*Correspondence: shishi.luo@helix.com

https://doi.org/10.1016/j.medj.2022.10.002
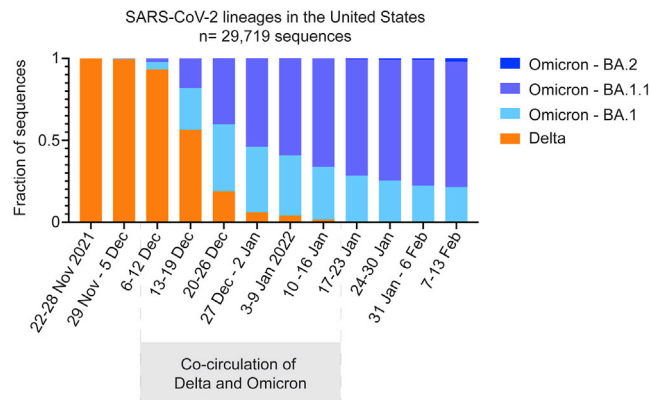
Med
Clinical and Translational Report

Figure 1. Co-circulation of Delta and Omicron variants in the United States
Fraction of lineages sequenced per week in the United States. Delta includes B.1.617.2 and all lineages starting with AY. Delta: orange. Omicron: light blue (BA.1), blue (BA.1.1), and navy blue (BA.2). The week of collection of the sample is on the x axis.
See also Tables S1 and S2.

samples, we also validated the results using an orthogonal genotyping assay (STAR methods). The results replicated in 18 of the 19 samples (Figures 2B and S1; Table 1), including two (HMIX1 and HMIX2) that have already been reported in a separate study.[11] A detailed list of the allele depths for each mutation for all of these samples is in Table S4. Overall, we estimate that, on average, ~1 in 800 (18/14,214; 95% confidence interval [CI]: 1/540 to 1/1,470, assuming a binomial distribution) positive samples between December 6, 2021, and January 16, 2022, had a co-infection.

Given how quickly Omicron displaced Delta, we hypothesized that in cases of co-infections, we would see, on average, more Omicron virions compared with Delta. Using the fraction of sequencing reads that mapped to mutations in either Delta or Omicron as a proxy, the fraction of Delta and Omicron virions in a given sample appeared similar (between 40% and 60%) in 8 out of 18 co-infections (Figures 2B and S1; Table S5). The Delta variant was higher than the Omicron variant in five co-infection samples, while the Omicron variant was higher than the Delta variant in the remaining five (Figures 2B and S1; Table S5). The fraction of Delta and Omicron virions in each sample was similar in the replicates (Figure 2C) despite re-extraction. These results did not support the hypothesis that the Omicron variant would outcompete the Delta variant when in the same host. There was also no correlation between the date of co-infection and the dominant variant (Figure 2D) or between the viral load in the nose and the dominant variant (Figure S2). We were not able to test the hypothesis that vaccination or prior infection by SARS-CoV-2 would better control Delta and lead to a higher fraction of Omicron in these co-infected samples. Our analysis is also limited by the fact that we do not have information on whether the exposure and seed infection by the two distinct variants happened at the same time or if they followed each other.

Within-host recombination of Delta and Omicron genomes
We hypothesized that a subset of host cells in a co-infection would inevitably contain both variants and, therefore, have the potential to generate recombinants. If these recombinants were replication competent and replicated to high enough titers, we would detect them in sequencing output, manifesting as a change in allele fraction of defining mutations near the recombination breakpoint.
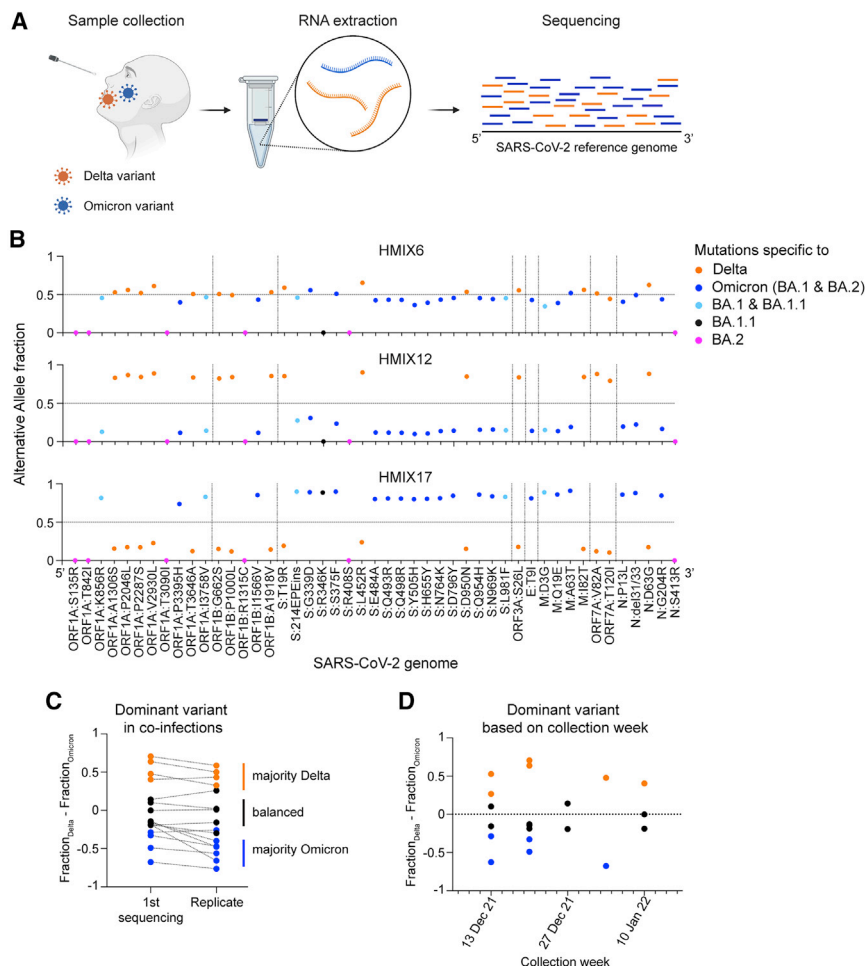
Figure 2. Co-infection with Delta and Omicron variants

(A) Schematic of a co-infection and the impact on the sequencing output.

(B) Three example graphs representing the alternative allele fraction for each mutation. Forty-seven mutations are plotted in order of their position on the SARS-CoV-2 genome from 5′ to 3′. Genes are separated by dashed vertical lines. Sixteen mutations specific to Delta are represented in orange. Nineteen mutations specific to Omicron and shared by its BA.1 and BA.2 sublineages are represented in blue. Five mutations specific to BA.1 (and BA.1.1) are in light blue. The defining BA.1.1 mutation is in black, and 6 mutations specific to BA.2 are in magenta. The graphs for all 18 co-infections are in Figure S1.

(C) The difference of Delta fraction and Omicron fraction is plotted for the initial sample and the replicate sequenced after a new RNA extraction. Dashed lines connect identical samples. Orange dots represent samples with a majority of Delta. Blue dots represent samples with a majority of Omicron.

(D) The difference of Delta fraction and Omicron fraction for the initial sample is plotted against the collection week of the sample. Orange dots represent samples with a majority of Delta. Blue dots represent samples with a majority of Omicron.

See also Figure S2 and Tables S3, S4, and S5.

Indeed, we find that HMIX16 (Figure 3A) exhibits precisely this characteristic. Alternative allele fractions for Delta mutations hover around 0.80 near the 5′ end of the genome but drop to around 0.50 near the beginning of the spike (S) gene and remain at this level until the 3′ end of the genome. This profile suggests the presence of a Delta-Omicron recombinant with a breakpoint preceding the S:214EPEins. Upon examination of read pairs sequenced from HMIX16 that spanned mutations

Med
Clinical and Translational Report

CellPress

Table 1. Samples with co-infections and clonal recombinant infections

Name	Collection week	State	COVID Cq	Unique SARS-CoV-2 reads	Median AAF	Lineage	Clade	Replication method	Dominant variant	Breakpoint
HMIX1	December 12, 2021	CA	18.1	480,979	0.71	none	21K (Omicron)	re-extraction and re-sequencing	Omicron	not applicable
HMIX2	December 20, 2021	NJ	16.6	625,507	0.53	none	21M (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX4	December 13, 2021	CA	19.5	826,167	0.61	none	21K (Omicron)	re-extraction and re-sequencing	Omicron	not applicable
HMIX5	December 13, 2021	PA	19.2	1,164,229	0.54	none	21M (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX6	December 13, 2021	CA	17.4	1,998,992	0.51	none	21M (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX7	December 13, 2021	PA	19.3	320,813	0.43	B.1.617.2	21J (Delta)	re-extraction and re-sequencing	Delta	not applicable
HMIX8	December 13, 2021	CA	24.1	31,419	0.77	BA.1	21K (Omicron)	re-extraction and re-sequencing	Omicron	not applicable
HMIX10	December 20, 2021	CA	24.4	20,698	0.55	none	21K (Omicron)	re-extraction and re-sequencing	Omicron	not applicable
HMIX11	December 20, 2021	FL	20.4	193,976	0.54	none	21K (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX12	December 20, 2021	GA	21.1	359,378	0.83	AY.25	21J (Delta)	re-extraction and re-sequencing	Delta	not applicable
HMIX13	December 20, 2021	MI	21.1	668,533	0.51	B.1.617.2	21J (Delta)	re-extraction and re-sequencing	Delta	not applicable
HMIX14	December 27, 2021	CA	17.5	2,540,144	0.50	none	21M (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX15	December 27, 2021	TX	22.9	41,983	0.51	none	21K (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX16	January 3, 2022	FL	18.2	103,983	0.51	none	21J (Delta)	re-extraction and re-sequencing	Delta	not applicable
HMIX17	January 3, 2022	FL	21.8	354,336	0.81	BA.1.1	21K (Omicron)	re-extraction and re-sequencing	Omicron	not applicable
HMIX18	January 10, 2022	IN	17.4	3,930,358	0.44	none	21M (Omicron)	re-extraction and re-sequencing	Delta	not applicable
HMIX19	January 10, 2022	OK	19.9	1,121,368	0.54	none	21M (Omicron)	re-extraction and re-sequencing	balanced	not applicable
HMIX20	January 10, 2022	GA	19.7	692,308	0.49	none	21M (Omicron)	re-extraction and re-sequencing	balanced	not applicable
RECOMB1	January 10, 2022	MA	22.5	8,639	1	none	21K (Omicron)	re-extraction and re-sequencing[a]	not Applicable	22,204 to 22,578
RECOMB2	February 7, 2022	MA	20.5	240,742	1	none	21K (Omicron)	re-extraction and re-sequencing	not Applicable	19,220 to 21,618

[a]Recombination in this sample was also confirmed by an orthogonal genotyping array and by long-read sequencing by an independent lab at the CDC.

unique for Delta and Omicron upstream of S:214EPEins, we found 4 read pairs that supported a Delta-Omicron recombinant, 7 read pairs that supported Delta only, and 10 that supported Omicron only (Figure 3B). The read pairs that supported a Delta-Omicron recombinant comprise the S:156/157del mutation of Delta on the 5′ end and the S:212del of Omicron on the 3′ end. The existence of these three unique mutation profiles presents compelling evidence that a recombinant virus was generated during co-infection, with a breakpoint region of 157 base pairs between nucleotide positions 22,036 and 22,193. We did not find read pairs supporting Delta-Omicron recombination in the same interval in the other co-infection samples, showing that these recombinations remain a rare event.

Infection with clonal Delta-Omicron recombinants
Having established both the presence of co-infections and evidence of recombination *in vivo*, we then looked for samples that are composed entirely of recombinant
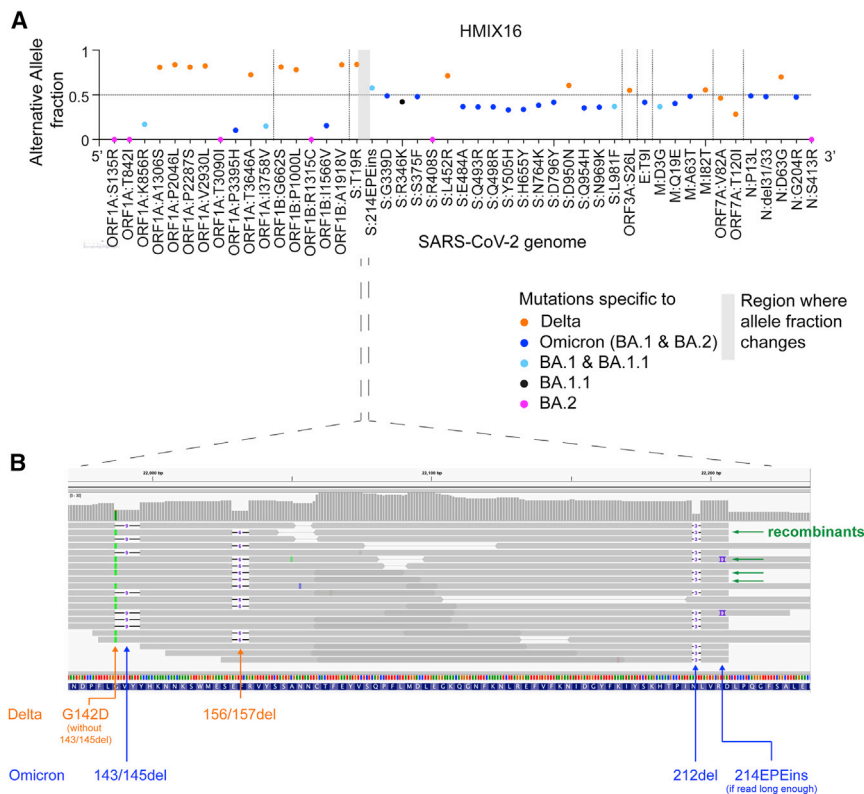
Figure 3. Evidence of Delta, Omicron, and recombinant read pairs in HMIX16

(A) Graph representing the alternative allele fraction for each mutation of sample HMIX16. Forty-seven mutations are plotted in order of their position on the SARS-CoV-2 genome from 5' to 3'. Genes are separated by dashed vertical lines. Sixteen mutations specific to Delta are represented in orange. Nineteen mutations specific to Omicron and shared by its sublineages BA.1 and BA.2 are represented in blue. Five mutations specific to BA.1 (and BA.1.1) are in light blue. One BA.1.1 mutation is in black, and 6 mutations specific to BA.2 are in magenta. The gray box represents the region where the alternative allele fraction changes.

(B) An IGV[12] view of the alignments for HMIX16, subsampled to only include read pairs where the first in pair covers the S:156/157del position and the second in pair covers the S:212del position. Read pairs representing three mutation profiles are present: (1) supporting Delta mutations only (7 read pairs), (2) supporting Omicron mutations only (10 read pairs), and (3) supporting a Delta-Omicron recombinant (4 read pairs) marked with a green arrow. Mutations specific to Delta are represented in orange. Mutations specific to Omicron are represented in blue. Read pairs that do not span these mutations are not shown.

virus. In such samples, we expect that all mutations called would be supported by ~100% of the reads because the viral population in the sample is composed of multiple copies of the same variant rather than a mixture of two. We first looked for recombinants with one breakpoint where all mutations identified on the 5' end of the breakpoint should be characteristic of one variant (e.g., variant A) and all mutations on the 3' end of the breakpoint should be characteristic of the other variant (e.g., variant B) (Figure 4A). We identified seven samples that had Delta-specific ORF1A:A1306 S at the 5' end of the genome and Omicron-specific N:P13L at the 3' end. One sample had Omicron-specific ORF1A:P3395 H at the 5' end and Delta-specific N:D63G at the 3' end. Further analysis of these eight genomes showed that only two genomes, RECOMB1 and RECOMB2, had multiple consecutive Delta mutations at the 5' end, while the 3' end of the genome had all of the Omicron mutations but none of the Delta mutations (Figure 4B). Four of the six other
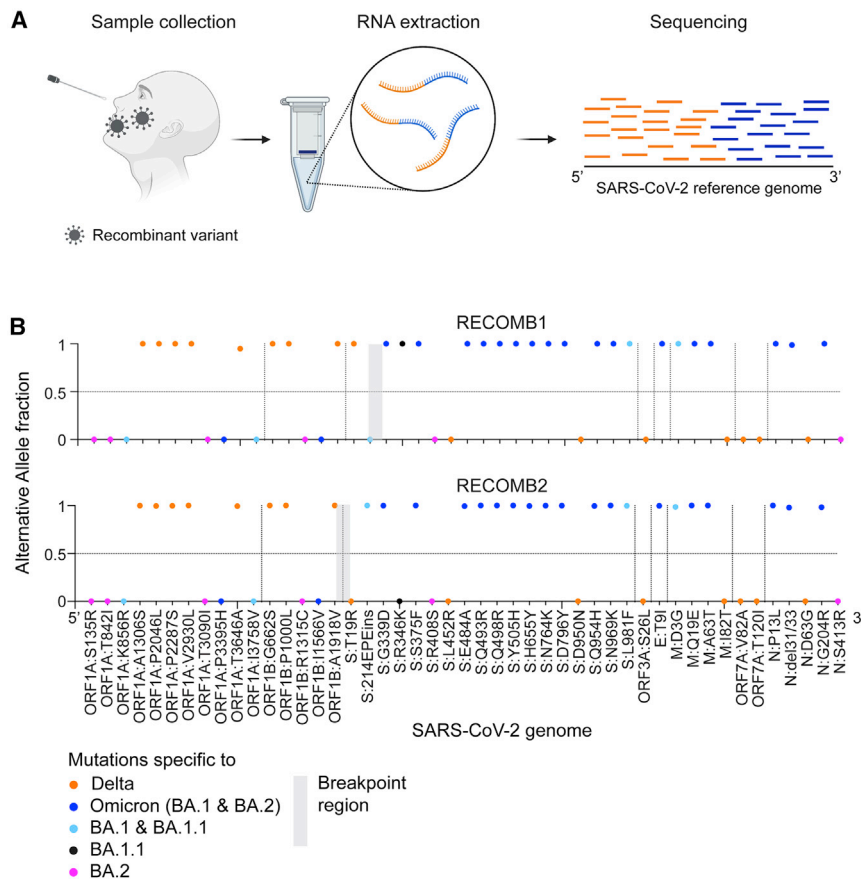
Med
Clinical and Translational Report

CellPress

Figure 4. Infection with clonal Delta-Omicron recombinants

(A) Schematic of the impact of an infection from a virus resulting from the recombination of Delta and Omicron on the sequencing output.

(B) Graphs representing the alternative allele fraction for each mutation. Forty-seven mutations are plotted in order of their position on the SARS-CoV-2 genome from 5′ to 3′. Genes are separated by dashed vertical lines. Sixteen mutations specific to Delta are represented in orange. Nineteen mutations specific to Omicron and shared by its sublineages BA.1 and BA.2 are represented in blue. Five mutations specific to BA.1 (and BA.1.1) are in light blue. One BA.1.1 mutation is in black, and 6 mutations specific to BA.2 are in magenta. The gray box represents the homologous region where the breakpoint of the recombination is.

See also Figure S3 and Tables S6 and S7.

genomes had all (5′ to 3′ of the genome) Omicron-specific mutations and the additional Delta ORF1A:A1306S, which was probably acquired independently. The remaining genomes had all of the Delta-specific mutations, with one containing an additional Omicron N:P13L and the other containing Omicron ORF1A:P3395H. These were probably also acquired independently.

Given that amplicon primer-based artifacts, in conjunction with laboratory contamination, have previously led to spurious signatures of recombination,[13] we were careful to check the quality of the two possible recombinant viruses. First, the library preparation method used by Helix's viral sequencing protocol is hybridization-based capture, not amplicon. Hybrid capture is less susceptible to artifacts due to mutations at primer sites, which has been a recurring issue with possible recombinant viruses observed in GISAID.[14] We have also seen far less drop out in sequences generated via hybrid capture compared with amplicon in our sequencing data

(Figure S3). Second, the Cq values for these two infections were low, $Cq_{(RECOMB1)} = 22.5$ and $Cq_{(RECOMB2)} = 20.5$, implying a high source viral load that would be less susceptible to contamination. Third, we replicated these results for RECOMB1 and RECOMB2 by re-extracting RNA from the collected sample and re-sequencing. In the initial and replicate samples, the median alternative allele fraction was 1, indicating that the majority of mutations were supported by 100% of the reads. Fourth, we performed a manual review of the alignments using IGV to make sure the reads supporting the mutations were of high quality. Fifth, we were able to validate our results for RECOMB1 by running a genotyping assay looking at Delta: C21618G (S:T19R at the protein level) and Omicron BA.1: G8393A, T13195C, C23202A (S:T547K at the protein level). The results showed the presence of Delta C21618G and Omicron BA.1C23202A in the sample but the absence of Omicron BA.1: G8393A and T13195C. These results confirm that the 5′ end of the genome was from Delta and the 3′ end from Omicron. Lastly, a RECOMB1 sample was also sent to the US Centers for Disease Control and Prevention (CDC) laboratory, and results were confirmed via another sequencing technology.[15] Together, these experiments provide evidence that the two independent infections were caused by viruses resulting from the recombination of Delta and Omicron.

The sequences of the two recombinant viruses differ slightly. The breakpoint region of RECOMB1 is 374 bases between nucleotide positions 22,204 and 22,578, while the breakpoint region of RECOMB2 is 2,398 bases between nucleotide positions 19,220 and 21,618 (Figure 4B). Of interest, there is a private mutation T19404C in RECOMB2 inside the breakpoint region. RECOMB1 is a recombination between Delta sublineage AY.119 and Omicron sublineage BA.1.1. The 5′ Delta end of RECOMB2 is too short for sublineage classification, but the 3′ end is Omicron sublineage BA.1. The full list of mutations including the presence of unlabeled mutations (not shared by a large fraction of genomes in the same lineage) are in Tables S6 (RECOMB1) and S7 (RECOMB2) in a VCF-like format. These two samples were both collected in Massachusetts (USA), but the difference in sequence suggests they are unrelated. Overall, infections from a recombinant Delta-Omicron virus remain rare: 2 out of 10,742 sequences between January 10 and February 13, 2022. Eight other sequences similar to RECOMB1 have been reported by the CDC from samples collected in the United States from December 31, 2021, to February 12, 2022.[15] We did not identify any recombinant with Omicron on the 5′ end and Delta on the 3′ end or any recombinant with two breakpoints in our dataset.

DISCUSSION

In this study, we identified and validated 18 cases of co-infection with the Delta and Omicron variants. While contamination could lead to the same output as a co-infection, several pieces of evidence discount contamination: (1) re-extraction and re-sequencing these samples led to the same results; (2) the fraction of reads supporting each variant was high in all cases (at least 15%); (3) samples that showed a co-infection were collected and processed on different days, and other samples sequenced on the same plates did not show co-infection; and (4) in one of these co-infections, we found evidence of recombinant virus at a low, but detectable, frequency, consistent with template switching during replication in a cell infected with two variants. Studies reporting on co-infections remain rare, likely due to the difficulty to differentiate true co-infections from contamination and the fact that these sequences are usually not uploaded to GISAID because of the specific quality controls they require. Nonetheless, at least two other studies reported on co-infections with Delta and Omicron.[11,16] These studies did not report any recombination event

identified within these co-infections. More emphasis on identifying co-infections followed by deep sequencing with long reads would allow the identification of rare recombination events in those samples similar to the one observed in sample HMIX16. In turn, the systematic assessment of recombination events in co-infections would further our understanding of recombination rates, hot spots, and the general mechanism of template switching in coronaviruses.[17,18]

We also identified two cases infected by a virus resulting from the recombination of Delta and Omicron. Our data again support chimeric sequences being the cause rather than technical artifacts: (1) we were able to replicate the result for both samples after re-extracting RNA; (2) our sequencing protocol is based on hybrid capture and is less prone to amplicon-based artifacts (Figure S3); (3) the recombination events were not limited to the S protein, which is where many amplicon-based primer artifacts have been detected;[13] and (4) eight other sequences identical or nearly identical to RECOMB1 were identified in the United States.[15] In parallel, other studies around the world have identified and proved the existence of Delta-Omicron recombinants.[19–21] Others have identified recombination between two Omicron variants (BA.1 and BA.2).[22] As of August 2022, none of these recombinant variants became dominant nor represented a sizable fraction of infections globally. However, there were enough sequences for some of them to merit their own Pango lineage such as XD, XE, or XJ.[23]

The mechanism by which a recombinant virus comes to dominate an infection remains somewhat of a puzzle. One possibility is that the two infections that contain only recombinant virus were themselves seeded by a recombinant virus. This implies that in their respective ancestral co-infections, the two recombinant viruses each rose to a high enough fraction to be transmitted during an exposure and were able to establish an infection in a new host. Yet, despite transmitting to a new host at least once, the transmission chain was not sustained; neither RECOMB1 nor RECOMB2 have led to large clusters of cases. The other possibility is that these two infections began as co-infections and that the recombinant viral population then completely outcompeted the Delta and Omicron populations within the host. Yet, it seems unlikely that Delta and Omicron can be completely cleared from a host while leaving the recombinant virus population intact. In either case, the recombinant did not appear to have an increased ability to transmit between hosts compared with co-circulating Omicron (BA.1, BA.2) variants. There are parallels here with HIV-1, where chronic infection and host immune response leads to extensive within-host diversity of the virus, but the genotypes of the virus that ultimately seed new infections are from a much narrower set of viral types.[24,25]

With more diversity in circulating SARS-CoV-2 genomes, it will now be possible to track recombinations, characterize the rate of recombination, and identify hot spots for breakpoints.[18] One way to detect and visualize these recombinants is the strategy we used. Another is to review every instance where a sample has good sequencing metrics but where methods like Nextclade[26] or Pangolearn[23] have difficulty attributing a clade or a lineage to the sequence. Specialized methods have also been developed to detect recombination in viruses.[5,27,28] With a better understanding of SARS-CoV-2 recombination, and by drawing parallels with recombination in other unsegmented positive-strand RNA viruses, as well as other viruses in general,[17] we can be better prepared to anticipate new variants or combinations of mutations of SARS-CoV-2 that may arise in the future.

Limitations of the study

One limitation to our study is that we did not consider the possibility that Delta sequences presenting with one Omicron-defining variant, or that Omicron sequences presenting with one Delta-defining variant, were the result of a recombination. We assumed that these single mutations were more likely to be the result of independent appearance rather than the result of recombination. There is evidence for convergent evolution and for specific mutations to appear multiple times independently. For example, a study in immunodeficient patients showed that E:T30I was the most recurrent occurring mutation in persistent infections.[29] However, it is still possible that single mutations could be the result of recombination. Moreover, our analysis of recombination between Delta and Omicron worked particularly well because of the high number of differences between Delta and Omicron BA.1. More sophisticated algorithms may be necessary to identify recombinations between variants with fewer differences. Lastly, our study was also limited in size and scope. We did not investigate co-infections and recombinations between two different Omicron variants. We also did not study potential recombinations between strains circulating in different hosts.[30]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Human subjects
  - Ethics statement
- METHOD DETAILS
  - Helix COVID-19 test data and sample selection
  - SARS-CoV-2 sequencing and consensus sequence generation
  - Viral lineage designation
  - Genotyping
  - Relative fraction of each variant in co-infections
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.medj.2022.10.002.

Med
Clinical and Translational Report

CellPress

REFERENCES

1. Bujarski, J.J. (1999). Recombination of viruses. In Encyclopedia of Virology, Second Edition, A. Granoff and R.G. Webster, eds. (Elsevier), pp. 1446–1453. https://doi.org/10.1006/rwvi.1999.0334.

2. Dehghan, S., Seto, J., Liu, E.B., Walsh, M.P., Dyer, D.W., Chodosh, J., and Seto, D. (2013). Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. Virology 443, 197–207. https://doi.org/10.1016/j.virol.2013.05.014.

3. Fischer, W., Giorgi, E.E., Chakraborty, S., Nguyen, K., Bhattacharya, T., Theiler, J., Goloboff, P.A., Yoon, H., Abfalterer, W., Foley, B.T., et al. (2021). HIV-1 and SARS-CoV-2: patterns in the evolution of two pandemic pathogens. Cell Host Microbe 29, 1093–1110. https://doi.org/10.1016/j.chom.2021.05.012.

4. Jackson, B., Boni, M.F., Bull, M.J., Colleran, A., Colquhoun, R.M., Darby, A.C., Haldenby, S., Hill, V., Lucaci, A., McCrone, J.T., et al. (2021). Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. Cell 184, 5179–5188.e8. https://doi.org/10.1016/j.cell.2021.08.014.

5. Varabyou, A., Pockrandt, C., Salzberg, S.L., and Pertea, M. (2021). Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. Genetics 218, iyab074. https://doi.org/10.1093/genetics/iyab074.

6. Viana, R., Moyo, S., Amoako, D.G., Tegally, H., Scheepers, C., Althaus, C.L., Anyaneji, U.J., Bester, P.A., Boni, M.F., Chand, M., et al. (2022). Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature 603, 679–686. https://doi.org/10.1038/s41586-022-04411-y.

7. Vaziri, A., and Allday, E. (2021). First U.S. Omicron Case Found in San Francisco. Here's what We Know (Chronicle).

8. Bolze, A., Cirulli, E.T., Luo, S., White, S., Wyman, D., Dei Rossi, A., Cassens, T., Jacobs, S., Nguyen, J., Ramirez, J.M., III, et al. (2021). Rapid Displacement of SARS-CoV-2 Variant B.1.1.7 by B.1.617.2 and P.1 in the United States. Preprint at bioRxiv. https://doi.org/10.1101/2021.06.20.21259195.

9. Earnest, R., Uddin, R., Matluk, N., Renzette, N., Siddle, K.J., Loreth, C., Adams, G., Tomkins-Tinch, C.H., Petrone, M.E., Rothman, J.E., et al. (2021). Comparative Transmissibility of SARS-CoV-2 Variants Delta and Alpha in New England. Preprint at medRxiv. https://doi.org/10.1101/2021.10.06.21264641.

10. CDC (2020). COVID Data Tracker (Centers for Disease Control and Prevention). https://covid.cdc.gov/covid-data-tracker/.

11. Roychoudhury, P., Luo, S., Hayashibara, K., Hajian, P., Mills, M.G., Lozach, J., Cassens, T., Wendm, S.T., Arnould, I., Becker, D., et al. (2022). Identification of omicron-delta coinfections using PCR-based genotyping. Microbiol. Spectr. 10, e0060522. https://doi.org/10.1128/spectrum.00605-22.

12. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754.

13. Kreier, F. (2022). Deltacron: the story of the variant that wasn't. Nature 602, 19. https://doi.org/10.1038/d41586-022-00149-9.

14. Sanderson, T., and Barrett, J.C. (2021). Variation at Spike position 142 in SARS-CoV-2 Delta genomes is a technical artifact caused by dropout of a sequencing amplicon. Preprint at bioRxiv. https://doi.org/10.1101/2021.10.14.21264847.

15. Lacek, K.A., Rambo-Martin, B.L., Batra, D., Zheng, X.-Y., Sakaguchi, H., Peacock, T., Keller, M., Wilson, M.M., Sheth, M., Davis, M.L., et al. (2022). Identification of a novel SARS-CoV-2 delta-omicron recombinant virus in the United States. Preprint at bioRxiv. https://doi.org/10.1101/2022.03.19.484981.

16. Rockett, R.J., Draper, J., Gall, M., Sim, E.M., Arnott, A., Agius, J.E., Johnson-Mackinnon, J., Fong, W., Martinez, E., Drew, A.P., et al. (2022). Co-infection with SARS-CoV-2 Omicron and Delta variants revealed by genomic surveillance. Nat. Commun. 13, 2745. https://doi.org/10.1038/s41467-022-30518-x.

17. Simon-Loriere, E., and Holmes, E.C. (2011). Why do RNA viruses recombine? Nat. Rev. Microbiol. 9, 617–626. https://doi.org/10.1038/nrmicro2614.

18. de Klerk, A., Swanepoel, P., Lourens, R., Zondo, M., Abodunran, I., Lytras, S., MacLean, O.A., Robertson, D., Kosakovsky Pond, S.L., Zehr, J.D., et al. (2021). Conserved recombination patterns across coronavirus subgenera. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.21.469423.

19. Ou, J., Lan, W., Wu, X., Zhao, T., Duan, B., Yang, P., Ren, Y., Quan, L., Zhao, W., Seto, D., et al. (2022). Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. Signal Transduct. Target. Ther. 7, 138. https://doi.org/10.1038/s41392-022-00992-2.

20. Simon-Loriere, E., Montagutelli, X., Lemoine, F., Donati, F., Touret, F., Bourret, J., Prot, M., Munier, S., Attia, M., Conquet, L., et al. (2022). Rapid Characterization of a Delta-Omicron SARS-CoV-2 Recombinant Detected in Europe. https://doi.org/10.21203/rs.3.rs-1502293/v1.

21. Wang, L., Zhou, H.-Y., Li, J.-Y., Cheng, Y.-X., Zhang, S., Aliyari, S., Wu, A., and Cheng, G. (2022). Potential intervariant and intravariant recombination of Delta and Omicron variants. J. Med. Virol. 94, 4830–4838. https://doi.org/10.1002/jmv.27939.

Med
Clinical and Translational Report

22. Lindh, E., Smura, T., Blomqvist, S., Liitsola, K., Vauhkonen, H., Savolainen, L., Ikonen, J., Ronkainen, J., Taskila, J., Taskila, T., et al. (2022). Genomic and epidemiological report of the recombinant XJ lineage SARS-CoV-2 variant, detected in northern Finland, January 2022. Euro Surveill. *27*. https://doi.org/10.2807/1560-7917.ES.2022.27.16.2200257.

23. Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., and Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. *5*, 1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

24. Joseph, S.B., Swanstrom, R., Kashuba, A.D.M., and Cohen, M.S. (2015). Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. Nat. Rev. Microbiol. *13*, 414–425. https://doi.org/10.1038/nrmicro3471.

25. Sagar, M., Laeyendecker, O., Lee, S., Gamiel, J., Wawer, M.J., Gray, R.H., Serwadda, D., Sewankambo, N.K., Shepherd, J.C., Toma, J., et al. (2009). Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. J. Infect. Dis. *199*, 580–589. https://doi.org/10.1086/596557.

26. Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics *34*, 4121–4123. https://doi.org/10.1093/bioinformatics/bty407.

27. Martin, D.P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. *1*, vev003. https://doi.org/10.1093/ve/vev003.

28. Samson, S., Lord, É., and Makarenkov, V. (2021). SimPlot++: a Python application for representing sequence similarity and detecting recombination. Preprint at arXiv. https://doi.org/10.1093/bioinformatics/btac287.

29. Wilkinson, S.A.J., Richter, A., Casey, A., Osman, H., Mirza, J.D., Stockton, J., Quick, J., Ratcliffe, L., Sparks, N., Cumley, N., et al. (2022). Recurrent SARS-CoV-2 mutations in immunodeficient patients. Virus Evol. *8*, veac050. https://doi.org/10.1093/ve/veac050.

30. Wang, L., Didelot, X., Bi, Y., and Gao, G.F. (2021). Assessing the extent of community spread caused by mink-derived SARS-CoV-2 variants. Innovation *2*, 100128. https://doi.org/10.1016/j.xinn.2021.100128.

31. Lai, E., Becker, D., Brzoska, P., Cassens, T., Davis-Turak, J., Diamond, E., Furtado, M., Gandhi, M., Gort, D., Greninger, A.L., et al. (2022). A method for variant agnostic detection of SARS-CoV-2, rapid monitoring of circulating variants, detection of mutations of biological significance, and early detection of emergent variants such as Omicron. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.08.22268865.

Med
Clinical and Translational Report

 CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Helix ® COVID-19 test	Helix	EUA201636
TaqPath COVID-19 Combo Kit	Thermo Fisher	Cat#A49868
MagMAX Viral/Pathogen II Nucleic Acid Isolation Kit	Thermo Fisher	Cat#A48383
Rapid RNA Library Kit	Swift Biosciences	Cat# R2384
RNA Library Prep Kit	Integrated DNA Technologies	Cat# 10010146
xGen COVID-19 Capture Panel	Integrated DNA Technologies	Cat# 10006764
Respiratory Virus Research Panel	Twist Biosciences	Cat#103068
NovaSeq 6000 Sequencing system S1 flow cell	Illumina	
NovaSeq 6000 Sequencing System S1 Reagent Kit v1.5 (300 cycles)	Illumina	Cat#20028317
Deposited data		
Raw SARS-CoV-2 genomes	GISAID	Virus name includes 'CDC-STM' and/or 'originating lab field' is 'Helix'
SARS-CoV-2 samples BAM files	NCBI SRA	BioProject PRJNA804575
Software and algorithms		
Bcl2fastq	Illumina	N/A
Klados-fastagenerator	Helix	N/A
BWA-MEM	https://github.com/lh3/bwa	N/A
Haplotyper algorithm	Sentieon, Inc	N/A
Pangolin v3.1.11	https://github.com/cov-lineages/pangoLEARN	N/A
PRISM v8	Graphpad	N/A
Other		
Code to make alternate allele fraction plots	This paper	https://data.mendeley.com/datasets/gvx4bwygdz/2

RESOURCE AVAILABILITY

Lead contact
Further information and requests for data, and resources should be directed to and will be fulfilled by the lead contact, Shishi Luo (shishi.luo@helix.com).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- All samples with a qc_status of 'pass' were uploaded to GISAID. There are two ways to find them on GISAID. One way is to download all of the samples that have a collection date within the time period studied in this paper and filter for 'Helix' in the 'Originating lab' field. The other way is to search for all samples with 'CDC-STM' in the 'Virus name' field as only Helix uses this nomenclature for the name of the virus (and filter by collection date as above).
- Identifiers for RECOMB1, GISAID: hCoV-19/USA/MA-CDC-STM-HZEBR92XC/2022, EPI_ISL_9088187
- Identifiers for RECOMB2, GISAID: hCoV-19/USA/MA-CDC-STM-SP94WR2RW/2022, EPI_ISL_10114799

- BAMs of co-infection samples and recombinant samples are available at SRA STUDY: PRJNA804575. Link: https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA804575&o=acc_s%3Aa
- SRA BioSample accessions:
- HMIX16: SAMN26527328, bioproject: PRJNA804575 (https://www.ncbi.nlm.nih.gov/biosample/26527328)
- RECOMB1: SAMN26527329, bioproject: PRJNA804575 (https://www.ncbi.nlm.nih.gov/biosample/26527329)
- RECOMB2: SAMN26527330, bioproject: PRJNA804575 (https://www.ncbi.nlm.nih.gov/biosample/26527330)
- The raw data used to generate all figures is available in the supplemental tables.
- The code used to make the alternative allele fraction plots is deposited in Mendeley Data: https://data.mendeley.com/datasets/gvx4bwygdz/2, and is in Data S1.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODELS AND SUBJECT DETAILS
Human subjects
This paper is based on the study of SARS-CoV-2 found and collected from individuals in the United States. The detailed demographics about these individuals can be found in Table S1 of this paper.

Ethics statement
The Helix data analyzed and presented here were obtained through IRB protocol WIRB#20203438, which grants a waiver of consent for a limited dataset for the purposes of public health under section 164.512(b) of the Privacy Rule (45 CFR § 164.512(b)). All samples were de-identified before receipt by the study investigators.

METHOD DETAILS
Helix COVID-19 test data and sample selection
All viral samples in this investigation were collected by Helix through its diagnostic testing laboratory. The Helix COVID-19 test is run on specimens collected across the US, and results are obtained as part of our standard test processing workflow using specimens from anterior nares swabs. The Helix COVID-19 Test is based on the Thermo Fisher TaqPath COVID-19 Flu A, Flu B Combo Kit, which targets three respiratory pathogens (SARS-CoV-2, Influenza A, and Influenza B). Swabs are transported in saline and sample tubes are heat inactivated upon receipt at the lab. Test results from positive cases, together with a limited amount of metadata (including sample collection date, state, and qRT-PCR Cq values for all targets), were used to build the research database used here.

SARS-CoV-2 sequencing and consensus sequence generation
Sequencing was performed by Helix as part of the SARS-CoV-2 genomic surveillance program in partnership with the Centers for Disease Control and Prevention (CDC). In the Helix workflow, RNA is extracted from 400 μL of patient anterior nares sample using the MagMAX Viral/Pathogen kit (ThermoScientific). Sequencing libraries were generated with total RNA library preparation (5μL RNA input volume) using the Rapid RNA Library Kit protocol (Swift Biosciences/Integrated DNA Technologies). SARS-CoV-2 genome capture was accomplished using hybridization kit xGen COVID-19 Capture Panel (Integrated DNA Technologies). Samples were sequenced using the NovaSeq 6000 Sequencing system S1 flow cell, with S1 Reagent Kit v1.5 (300 cycles). Note: Specimens identified as co-infections or recombinants through

Med
Clinical and Translational Report

CellPress

the surveillance program were verified by reprocessing from the original specimen. The process was replicated as described above; however, the hybridization probe panel was substituted. The IDT COVID-19 Capture Panel was replaced with the Respiratory Virus Research Panel (Twist Biosciences), while all other reagents remained the same.

Bioinformatic processing of this sequencing output is as follows. The flow cell output is demultiplexed with bcl2fastq (Illumina) into per-sample FASTQ sequences that are then run through the Helix fastagenerator pipeline to produce a sequence FASTA file. First, reads are aligned to a reference comprising the SARS-CoV-2 genome (NCBI accession NC_045512.2) and the human transcriptome (GENCODE v37) using BWA-MEM. Reads are then marked for duplicates before proceeding to variant calling using the Haplotyper algorithm (Sentieon, Inc). Finally, the per-base coverage from the alignment file (BAM) and per-variant allele depths from the variant call format (VCF) file are used to build a consensus sequence according to the following criteria: coverage from at least 5 unique reads is required with at least 80% of the reads supporting the allele. Otherwise, that base is considered uncertain, and an N is reported.

Alternative allele fraction is the number of reads supporting an alternative allele (i.e. a mutation) divided by the total number of reads covering the position. The median alternative allele fraction is calculated as the median value of alternative allele fractions at sites where at least 15% of the reads support a mutation.

Quality control (QC) of the viral sequences occurs primarily at two levels: sample and plate. A sample-level QC status of 'pass' indicates a sample is unlikely to have been contaminated and has a sufficiently complete consensus sequence to be assigned a lineage. For a qc_status of 'pass', a sample required a median alternative allele fraction of at least 0.8 for its variants (any variant VCF record in the Haplotyper VCF file) and a consensus sequence containing at most 30% N bases. At the plate level, our QC criteria are designed to flag potential reagent issues or sample swaps that would require an entire plate to be re-processed (this is extremely rare).

Viral lineage designation
Viral sequences were assigned a Pango lineage[23] using pangoLEARN (https://github.com/cov-lineages/pangoLEARN). For this analysis, pangoLEARN version 2022-02-02 with Pangolin software version 3.1.11 was used. We sequenced and were able to attribute a lineage to 29,719 sequences from samples collected between November 22, 2021 and February 13, 2022 for genomic surveillance purposes.

Genotyping
The detailed genotyping method as well as the validation of the method used in this study are previously described.[31] The four specific markers used were:

- Delta: C21618G
- Omicron (BA.1): G8393A
- Omicron (BA.1): T13195C
- Omicron (BA.1): C23202A

Relative fraction of each variant in co-infections
Number of RNA copies and coverage does vary across the SARS-CoV-2 genome. The density of mutations specific to Delta or Omicron also varies. There are many

more Omicron-specific mutations in the spike protein. To try to minimize some biases, we took 4 Delta-specific mutations and 4 Omicron-specific mutations spread across SARS-CoV-2 genome to calculate the mean Delta-allele fraction and the mean Omicron-allele fraction in each sample. The 4 Delta-specific mutations are: ORF1A:P2046L, ORF1B:P1000L, S:T19R, M:I82T. The 4 Omicron-specific mutations are: ORF1A:P3395H, ORF1B:I1566V, S:N969K, M:A63T. Of note, ORF1A:P2046L is not a defining mutation of Delta (Nextstrain clade 21A) as it is present in only 90.47% of the Delta sequences based on https://covariants.org/variants/21A.Delta. It is present in the majority of Delta sub-lineages including Nexstrain clade 21J, but it is not present in the Nexstrain clade 21I. Importantly it is present in all co-infection samples from our study and can therefore be used for the calculation of the relative fraction.

The results are in Table S5. We considered Delta to be the dominant variant if the Delta fraction was above 60% and the Omicron fraction was below 40%. We considered Omicron to be the dominant variant if the Omicron fraction was above 60% and the Delta fraction was below 40%. Other samples were considered balanced. To decide which variant was dominant for each sample for Table 1, we used the results of the initial sequencing.

QUANTIFICATION AND STATISTICAL ANALYSIS

The 95% Confidence Interval for the frequency of co-infections during co-circulation was calculated using the Python statsmodels.stats.proportion.proportion_confint package.

```
ci_low, ci_up = sm.stats.proportion_confint(18, 14214, alpha=0.05, method='normal')
```