Computational and Structural Biotechnology Journal 20 (2022) 4015–4024

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Genomic evidence for divergent co-infections of co-circulating SARS-CoV-2 lineages

Hang-Yu Zhou [a,b,1], Ye-Xiao Cheng [a,b,c,1], Lin Xu [a,b,c,1], Jia-Ying Li [a,b,1], Chen-Yue Tao [a,b,d], Cheng-Yang Ji [a,b], Na Han [a,b], Rong Yang [a,b], Hui Wu [a,b,c], Yaling Li [e], Aiping Wu [a,b,*]

[a] Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China
[b] Suzhou Institute of Systems Medicine, Suzhou 215123, China
[c] School of Life Science and Technology, China Pharmaceutical University, Nanjing 211100, China
[d] Imperial College London, London SW7 2AZ, UK
[e] Zhejiang Lab, Hangzhou 310000, China

ARTICLE INFO

Article history:
Received 26 March 2022
Received in revised form 25 July 2022
Accepted 25 July 2022
Available online 28 July 2022

Keywords:
SARS-CoV-2
Co-infection
Co-circulation
Recombination

ABSTRACT

Co-infection of RNA viruses may contribute to their recombination and cause severe clinical symptoms. However, the tracking and identification of SARS-CoV-2 co-infection persist as challenges. Due to the lack of methods for detecting co-infected samples in a large amount of deep sequencing data, the lineage composition, spatial–temporal distribution, and frequency of SARS-CoV-2 co-infection events in the population remains unclear. Here, we propose a hypergeometric distribution–based method named Cov2Coinfect with the ability to decode the lineage composition from 50,809 deep sequencing data. By resolving the mutational patterns in each sample, Cov2Coinfect can precisely determine the co-infected SARS-CoV-2 variants from deep sequencing data. Results from two independent and parallel projects in the United States achieved a similar co-infection rate of 0.3–0.5 % in SARS-CoV-2 positive samples. Notably, all co-infected variants were highly consistent with the co-circulating SARS-CoV-2 lineages in the regional epidemiology, demonstrating that the co-circulation of different variants is an essential prerequisite for co-infection. Overall, our study not only provides a robust method to identify the co-infected SARS-CoV-2 variants from sequencing samples, but also highlights the urgent need to pay more attention to co-infected patients for better disease prevention and control.

1. Introduction

Since its initial appearance in late 2019, the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has rapidly evolved into a global pandemic [1–2]. The widespread transmission and geographical isolation of SARS-CoV-2 have greatly promoted its genetic diversity. By March 22th 2022, over one thousand lineages had already been clearly defined by the Pangolin nomenclature [3]. Viruses within a defined lineage often share several common mutations and have similar biological properties. Until July 2022, five "variants of concern" (VoCs) have been identified by the World Health Organization (WHO). Among them, Alpha variant (B.1.1.7 and descendent lineages) was estimated to have greater than 50 % enhanced transmissibility [4]. Beta variant (B.1.351 and descendent lineages) and Gamma variant (P.1 and descendent lineages) showed the capacity to evade inhibition by neutralizing antibodies [5]. Delta variant (B.1.617.2 and descendent lineages) caused greatly increased numbers of infections in India early in 2021 and became the dominant epidemic strain in global until late 2021 [6–7], while B.1.1.529 and its descendent lineages (the Omicron variant) spread at an unprecedented rate. Studies have shown that the Omicron variant can escape the majority of existing SARS-CoV-2 neutralizing antibodies [8–10].

Currently, the re-infection of SARS-CoV-2 has been extensively discussed [11–12]. In addition, accumulated evidence in viral homologous recombination [13–15] implied that co-infection events caused by different SARS-CoV-2 lineages may occur frequently. However, due to the lack of effective identification methods, reports on viral co-infection of divergent lineages are relatively rare [16–20]. The co-infection of SARS-CoV-2 lineages should be given more attention. Previous reports have indicated

* Corresponding author at: Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China.
E-mail address: wap@ism.cams.cn (A. Wu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2022.07.042

that viral co-infection may cause severe clinical symptoms. For instance, human immunodeficiency virus (HIV) co-infection contributes to rapid disease progression [21–22], increased viral load, and requires antiretroviral treatment effective against both HIV variants [23]. Also, co-infection may contribute to SARS-CoV-2 recombination and accelerate the generation of recombinant viruses since coronaviruses have relatively high recombination rates [24–26]. It has been reported that recombination between coronavirus occurs frequently. The emerging virus through recombination could have ability to infect new species [27–28], increase cross-species transmission [29–30], and gain resistance to antivirals [31]. Thus, with the increasing diversity of SARS-CoV-2 and the co-existence of multiple regional lineages globally, it is significant to clarify the frequency of co-infection in the population and the exact compositional lineages of co-infection in individuals.

In theory, genomic evidence should be available in deep sequencing data if a patient has been co-infected with two or more SARS-CoV-2 lineages. Like other RNA viruses, the identified SARS-CoV-2 genomes in patients exist as quasi-species with many within-host variations [18,32–33]. Thus, in a co-infection sample, viruses from each SARS-CoV-2 lineage would retain the same number of variations. It could be inferred that at least three criterions should be met, including 1) featured mutations in the inferred candidate lineages should be detected in the sample, 2) frequencies of featured mutations in the same candidate lineage should be kept at similar levels, 3) the sum of frequencies of all the detected lineages should be nearly 100 %.

Based on these criterions, here we propose a hypergeometric distribution-based method (Cov2Coinfect) to identify the co-infected SARS-CoV-2 lineages from next-generation sequencing (NGS) sequencing data. In Cov2Coinfect, hypergeometric distribution was applied to search candidate lineages based on the detected mutation patterns in a sample. To provide an example of this application, we collected and analyzed 50,809 SARS-CoV-2 positive samples with paired-end deep sequencing data that were generated with the Illumina platform from two parallel projects obtained from the National Center for Biotechnology Information (NCBI). All these samples had detailed metadata and were collected from the United States between January 1st and September 7th, 2021. Among all the samples, we have identified 195 potential co-infection samples of divergent SARS-CoV-2 variants, with the co-infection rate in PRJNA716985 and PRJNA720050 as 0.38 % and 0.46 %, respectively. Apart from 192 samples co-infected by two lineages, three samples were co-infected by three lineages. The co-circulation of multiple dominant viral lineages in the same region is the main cause of these co-infection events.

2. Material and methods

2.1. Sample collection

In total, 46,465 and 4,344 SRA runs in Projects PRJNA716985 and PRJNA720050 were collected from the NCBI (https://www.ncbi.nlm.nih.gov), respectively. These samples were collected in the United States from January 2021 to September 2021 and sequenced with the Illumina platform. Samples in these projects have been retained with complete meta information, including the collection date, isolated region, and sex and age of the patient.

2.2. Calling variants

The collected samples were primarily transformed into FASTQ files using sra-tools. Since the samples were sampled and sequenced following the ARTIC version 3 protocol, all the sra files were treated with a recommended workflow (https://dockstore.org/workflows/github.com/iwc-workflows/sars-cov-2-pe-illumina-artic-variant-calling/COVID-19-PE-ARTIC-ILLUMINA:main?tab=info) to detect intra-host single nucleotide variants (iSNVs). This workflow is specifically designed for samples sequenced with the ARTIC version 3 protocol and can reliably detect iSNVs and low-frequency mutations. The detected nucleotide mutations were further converted into amino acid variations using a homemade Python script.

2.3. Identification of lineages-defined feature variations

The lineage-defined feature variations were defined as shared lineage-specific signature variations of strains belonging to the same lineage. In general, the lineage-defined feature variations were set as the nonsynonymous mutations shared by at least 75 % of viral strains in a specific lineage (https://outbreak.info/situation-reports/methods#characteristic). However, given the rapid divergence of SARS-CoV-2, many sub-lineages have been formed and share the same feature variations at the 75 % level, which could not distinguish viral strains belonging to similar lineages. Therefore, in this study, we further introduced the mutations shared by at least 10 % of viruses to distinguish the neighboring lineages with similar feature variations at the 75 % level. In total, more than 2.5 million SARS-CoV-2 consensus genomes were collected from Global Initiative on Sharing All Influenza Data (GISAID) database [34–35]. All variations that caused nonsynonymous mutations were identified for each viral genome. The lineage of each virus was derived with the Pango nomenclature [3]. A homemade Python script was applied to extract the mutations shared by at least 75 % of all the viruses in one lineage as the 75 % feature variations (FV-75). Similarly, mutations shared by at least 10 % of all the viruses in one lineage were extracted as 10 % feature variations (FV-10). To avoid overfitting, the lineages with few viral genomes globally (<0.01 % of all 2.5 million SARS-CoV-2 genomes, or < 250 genomes) were discarded.

2.4. Hypergeometric distribution-based method for detecting SARS-CoV-2 lineages

Files contain the iSNV of each sample and the Lineage Defining Variation of each lineage were used as the input files. The detection of co-infection could be divided into three steps. Firstly, all the samples were sent for a hypergeometric distribution test, for which the formula is:

$$P(X = K) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Here, N is the total number of nonsynonymous mutations that occur in all SARS-CoV-2 consensus genomes, K is the number of feature variations of a SARS-CoV-2 lineage, n is the number of remaining undefined mutations of sample, and k is the number of remaining undefined mutations that occur in both the sample and lineage feature variations.

A list of candidate lineages with *P*-value were generated. All mutations in the screened sample were assigned into each candidate lineage and were labelled as lineage unique mutations and lineage shared mutations. Then, the consistency of lineage unique mutations was evaluated by standard deviation. All the candidate lineages were tested and lineages with low mutation consistency were dropped. The frequencies of the reserved lineages were calculated as the average frequencies of all the lineage unique mutations. For each sample every lineage frequency was summed up to test if that total was approximately equal to 100 %. Finally,

Fig. 1. The overview of Cov2Coinfect. The algorithm of identifying co-infected SARS-CoV-2 lineages consists of three steps. Firstly, the input data (both NGS sequencing data and Lineage defining variation list) are sent for a Hypergeometric distribution test to calculate the *P*-value of every candidate lineage. Secondly, mutations in each candidate lineage are evaluated for their consistency. Lineages with consistently featured mutations were reserved. Thirdly, if the sum of the lineage frequencies of all the reversed candidate lineages is approximately equal to 100%, the sample will be identified as co-infection sample. The orange triangle points to the mutations shared by multiple lineages. This algorithm could be easily applied in finding lineage composition of a co-infection sample, and in tracking the spatial–temporal distribution and frequencies of co-infections in population. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

single-lineage infections, multi-lineage co-infections, and other situations were determined and outputted as three individual files.

3. Results

Under the quasi-species hypothesis, we designed the hypergeometric distribution–based method (Cov2Coinfect) to decode the infected SARS-CoV-2 lineage(s) in a sequencing sample (Fig. 1, Fig. S1 and Methods). In summary, the combination of mutations in each sample was compared with feature variations (mutations) of all defined SARS-CoV-2 lineages. For each lineage, a hypergeometric test was used to compute the probability (*P*-value) of observed successes (mutations that occurred in both the sample and lineage feature variations) under the "null hypothesis," i.e., the hypothesis that there is nothing special about the lineage. If the *P*-value is sufficiently low, we can reject the null hypothesis as impossible and conclude that the sample is highly correlated with the tested lineage, and the candidate lineage could be considered to investigate. Then, mutations in each candidate lineage were evaluated together for their consistency. Lineages with featured mutations of similar frequency were kept and the frequency of lineage(s) in the detected sample was further calculated. Finally, the co-infection event was determined, and the co-infected lineages were recognized. Using Cov2Coinfect, any dataset containing over 50,000 samples could be screened for the possible co-infection samples. Furthermore, the co-infected pattern, spatiotemporal distribution, and the frequency in population of SARS-CoV-2 co-infection could be inferred as well.

In the 50,809 samples (which account for over 30 % collected samples in the United States from Jan. to Sep. 2021) from two independent projects, 46,465 samples were collected from project no. PRJNA716985 with an average sequencing depth of 50x, whilst the other 4,344 samples were collected from project no. PRJNA716985 with an average sequencing depth of 300x. Since samples from these two projects were collected and sequenced in parallel and their collection dates have some overlap during Feb. 2021 to Mar. 2021, the co-infection results between these two projects could be mutual verification. The NGS raw data were treated following a ready-to-use ARTIC workflow [36], which can guarantee the robustness of both high-and low-frequency iSNVs. Of all the samples, most of them were identified to be infected with only one SARS-CoV-2 lineage as expected. As shown in Fig. 2A, the pattern of feature variations for a typical single-lineage infection is easily determined. Namely, most of the feature variations belonging to a specific lineage could be detected in a sample. Besides, the feature variations have a similar frequency of reads in each site, demonstrating good genomic homogeneity within a single lineage. In addition, few variations that do not match any lineage-defined feature variations were observed and could be recognized as *de novo* mutations. Furthermore, in this case, the identified Alpha (B.1.1.7) lineage was the dominant lineage in the place where the samples were collected (Fig. 2B), which confirmed the rationality of identifying a lineage using its lineage-defined feature variations from deep sequencing data.

In project no. PRJNA716985, 172 (0.37 %) samples were clearly classified as co-infected by SARS-CoV-2 strains from two different lineages, whilst project no. PRJNA720050 has 20 samples (0.46 %). Fig. 2C shows a typical example for co-infection by two SARS-CoV-

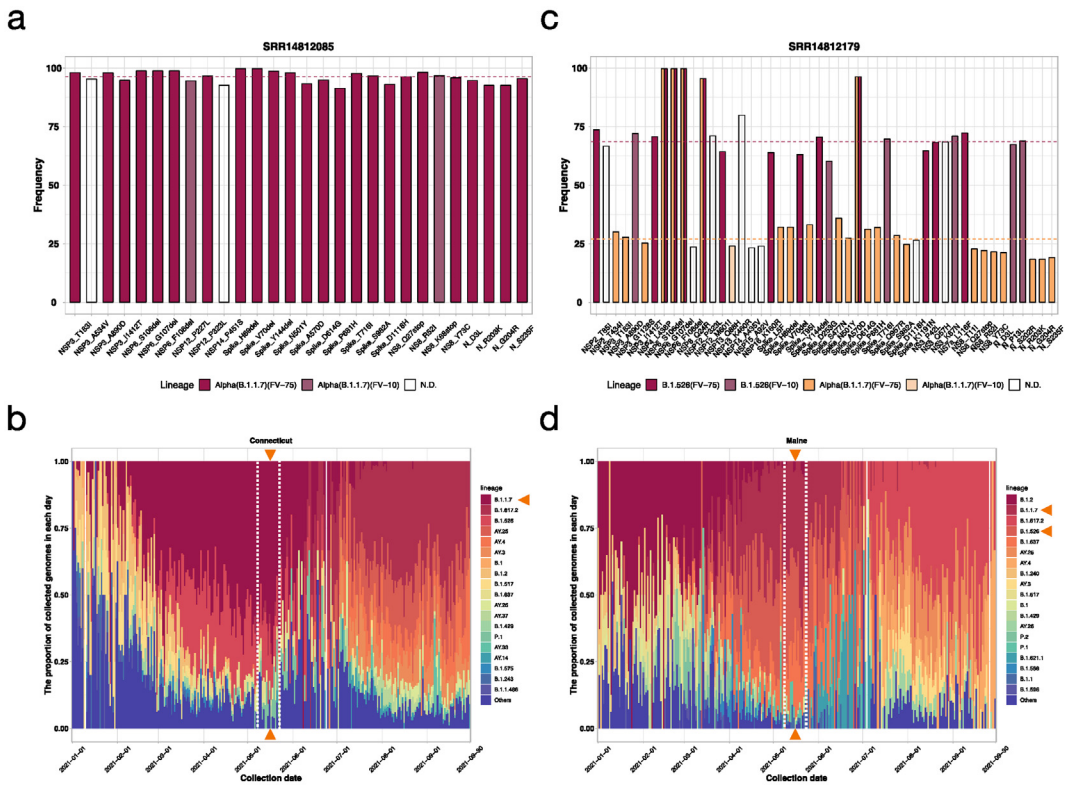

Fig. 2. Patterns for single-lineage infection and two-lineage co-infections. a. A sample infected by one specific SARS-CoV-2 lineage. Most of the feature variations of the identified Alpha lineage (FV-75 and FV-10) were detected at the same level. Non-determined variations are shown as a white column. b. The lineage ratio of SARS-CoV-2 lineages isolated in Connecticut from January 1 to September 30, 2021, including the location and time point of the representative sample used in a, i.e., Connecticut and May 17, 2021 (the date is signed with orange arrows). c. A sample co-infected by two SARS-CoV-2 lineages. Most of the feature variations of the two identified lineages (B.1.526 and Alpha) are shown in purple and orange. Two shared variations are shown as both purple and orange. d. The lineage ratio of SARS-CoV-2 lineages isolated in Maine from January 1 to September 30, 2021. The sample used in c was isolated in Maine on May 16, 2021. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2 lineages. Alpha (B.1.1.7) and B.1.526 were identified as the two lineages in this sample. One hundred percent (24/24) of feature variations existed in more than 75 % (FV-75) of the Alpha lineage, and 100 % (14/14) of the FV-75 feature variations of the B.1.526 lineage were detected in this sample. Meanwhile, the average frequency of Alpha lineage–specific variations was ∼ 28 %, while that of B.1.2 lineage–specific variations was ∼ 70 %, and the average frequencies of the five variations shared by Alpha and B.1.526 lineages, including NSP12_P323L, Spike_D614G, and deletions in NSP6, were all nearly 100 %. These observed facts exactly matched with the three hypothesized pieces of genomic evidence inferred from the quasi-species hypothesis. The co-infection of Alpha and B.1.526 lineages were also consistent with the epidemiological background of regional SARS-CoV-2. As shown in Fig. 2D, at the collection date (May 16, 2021), the Alpha lineage was the dominant lineage in the U.S. state of Maine, while B.1.526 was the second dominant epidemic lineage around the collection date.

Apart from co-infection by two lineages, we unexpectedly identified three samples co-infected with three lineages (0.006 %) from project no. PRJNA716985. The sample in Fig. 2 is a typical example and was collected in the U.S. state of Connecticut on May 17, 2021. The three hypothesized pieces of genomic evidence (Fig. 1) could be observed in this sample clearly (Fig. 3A). First, most lineage-specific feature variations of Alpha, B.1.526, and Gamma (P.1) could be identified at their own levels, respectively. The Alpha lineage was identified to occupy ∼ 55 % of all strains, while B.1.526 and Gamma occupied ∼ 25 % and ∼ 15 %, respectively. Second, the frequency of three feature variations (Spike_N501Y, N_R203K, and N_G204R) shared by Gamma and Alpha totaled nearly 70 %, which was almost equal to the sum of the mean frequencies of Alpha and Gamma. Finally, the frequencies of five feature variations (NSP12_P323L, Spike_D614G, and deletions in NSP6) shared by all three lineages were all nearly 100 %. The detection of these three lineages was also consistent with the epidemiological patterns of SARS-CoV-2 lineages in the sampling location (i.e., Connecticut) (Fig. 3B).

The metadata of all the SARS-CoV-2 co-infected samples (Table 1) were further analyzed. Of all the 195 co-infected samples, 91 were from male individuals and 104 were from female individuals. The average age was 35 years old (median, 32 years) for all patients, where the youngest patient was one year old, and the oldest patient was 85 years old. No obvious spatial–temporal bias was found in these samples. We evaluated the viral load with the diagnostic polymerase chain reaction cycle threshold (Ct) value. Compared to the average Ct value of all samples (18.70), the average Ct value of co-infected samples was at a similar level (20.47). Additionally, in these samples, we found some samples with falsely identified lineages. For instance, a sample had been wrongly classified as a B.1 infection (Fig. S2), but we found all the feature variations belonged to two identified lineages (Alpha and B.1.526) divided as ∼ 50 % each.

Since 195 co-infected samples were obtained, we made the effort to answer the question of whether the co-infected SARS-CoV-2 lineages have lineage tendentiousness by designating each pair of co-infected lineages as having a connection to build up a comprehensive network (Fig. 4A). In the co-infected network, the Alpha (B.1.1.7) lineage and Delta (B.1.617.2) lineage successively became the centers of co-infection (Fig. S3). However, from June 2021 onward, when the Delta lineage grew to become the domi-

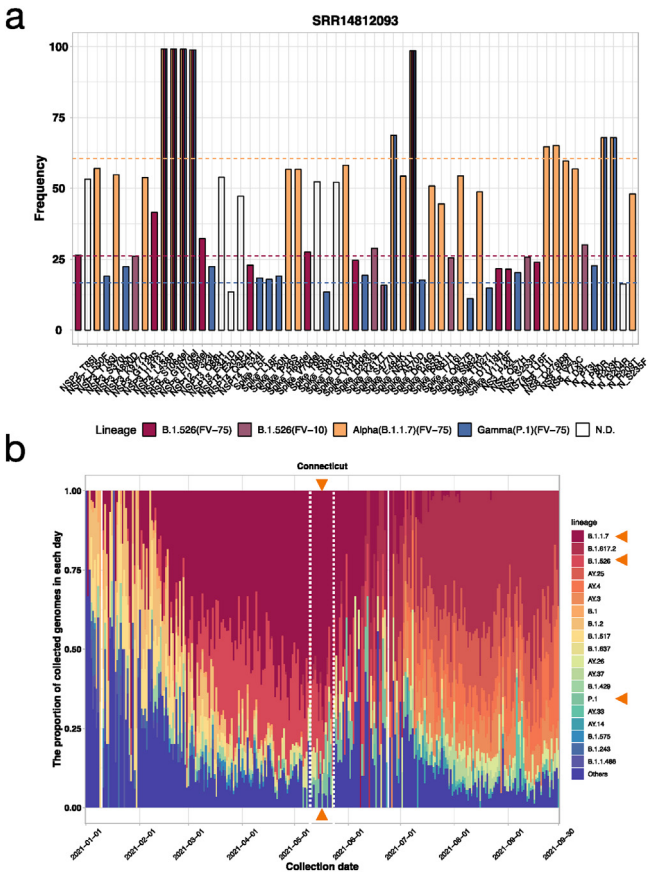

Fig. 3. Co-infection of three SARS-CoV-2 lineages. a. An identified sample co-infected by three SARS-CoV-2 lineages. The feature variations of the three identified lineages (B.1.526, Alpha, and Gamma) are shown in purple, orange, and blue, respectively. b. The lineage ratio of SARS-CoV-2 lineages isolated in Connecticut from January 1 to September 30, 2021. The sample used in a was isolated in Connecticut on May 17, 2021 (this day is denoted by orange arrows). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

nant lineage(s), an increasing number of regionally differentiated Delta descendant lineages emerged. The situation of co-infection was transferred from one lineage centered to multiple lineages centered (Fig. S4), which greatly increased the rate of co-infection (Fig. 4B). With the increasing number of co-infections of Delta lineage and its descendant lineages (Fig. 4C), the next variant of concern is likely to result from recombined viruses of the Delta lineage, and it is necessary to keep a close eye on the co-circulation of sub-lineages in the future.

4. Discussion

Recent studies have confirmed the high reliability of sequencing data in detecting within-host variations [37–39]. Benefiting from the worldwide rapid accumulation and open sharing of SARS-CoV-2 genomes, the available large-scale genomic dataset offers substantial support in detecting co-infection events even when they are very rare in the population. For most of the SARS-CoV-2–positive samples, whether they were infected by one lineage or by multiple lineages, the pattern of mutations in sequencing data fit well with their lineage-defined feature variations. In particular, we observed that the sum of the frequencies of lineage-unique variations was equal to the average frequencies of their shared variations, demonstrating the co-existence of these lineages within the same sample. Moreover, the epidemiological background of the detected co-infected SARS-CoV-2 samples was highly consistent with the identified lineages for their co-circulations around the sampling locations. The consistency between the hypothesis and observations provides strong evidence for the detected co-infection events.

One question to ask is whether we can infer the sources of a co-infection event from its genomic characteristics. When we assigned variations into lineage(s), we found there were always some undetermined variations. Further analysis suggested that these undetermined variations could possibly be used to trace the origins of co-infection events. For instance, in a representative co-infected sample (SRR14812179) with two lineages (Fig. 2C), four undetermined variations—NSP2_T434I, NSP12_M601I, NSP14_A435V, and NS3_K67N—had similar frequencies with the feature variations of the identified Alpha lineage (Fig. 2C). Accordingly, of all the global 4,858,598 viral genomes, only six other viral strains in B.1.526 lineage were detected to possess the above five variations as well. Regarding the source of the six viral strains, all were isolated in Maine, suggesting that the B.1.526 lineage in the co-infected sample might be a regional one. Similarly, four undetermined variations in this co-infected sample were detected to have similar frequencies with feature variations of Alpha (Fig. 2C). After scrutinizing all 4,858,598 viral genomes with the

Table 1
The metadata of all the SARS-CoV-2 co-infected samples.

ID	Collection Date	Location	Lineage	Host Age	Host Sex	CT Value	Lineage Detected
SRR15628265	2021/8/9	USA: California	B.1.617.2	71	male	\	Delta(AY.44)/Delta(AY.21)
SRR15656474	2021/8/2	USA: Connecticut	B.1.617.2	3	male	\	Delta(AY.37)/Delta(AY.40)
SRR15748551	2021/6/24	USA: Nevada	B.1.617.2	54	male	\	Delta(AY.2)/Delta(AY.44)
SRR15741952	2021/8/14	USA: Texas	AY.4	17	male	\	Delta(AY.24)/Delta(AY.8)
SRR16025135	2021/8/28	USA: Florida	B.1.617.2	32	female	24.91	Delta(AY.26)/Delta(AY.4)
SRR15747646	2021/6/29	USA: Florida	B.1.621	37	female	\	Mu(B.1.621)/Delta(AY.44)
SRR15822319	2021/7/10	USA: California	B.1.617.2	18	male	\	Delta(AY.30)/Delta(AY.19)
SRR15822746	2021/7/10	USA: Texas	B.1.617.2	1	female	\	Delta(AY.21)/Delta(AY.35)
SRR15753015	2021/6/19	USA: California	B.1.1.7	56	male	\	Alpha(B.1.1.7)/Delta(B.1.617.2)
SRR15752147	2021/6/10	USA: California	P.1	7	male	\	Gamma(P.1)/Alpha(Q.7)
SRR14452198	2021/4/13	USA: Illinois	B.1.1.7	26	female	16.62	Alpha(B.1.1.7)/Gamma(P.1)
SRR15822209	2021/7/12	USA: Nevada	B.1.620	34	male	\	Delta(AY.15)/Alpha(Q.8)
SRR14402893	2021/2/19	USA: Michigan	B.1.1.7	41	female	18.02	Alpha(B.1.1.7)/B.1.2
SRR15745419	2021/6/30	USA: Texas	B.1.1.7	32	female	19.3	Delta(AY.14)/Alpha(Q.4)/Delta(AY.25)
SRR15656298	2021/8/1	USA: Wisconsin	B.1.617.2	60	female	23.93	Delta(AY.46.3)/Delta(AY.9)
SRR15774852	2021/7/31	USA: Louisiana	AY.4	42	female	16.87	Delta(AY.21)/Delta(AY.14)
SRR15617395	2021/8/7	USA: Michigan	B.1.621	38	female	\	Mu(B.1.621)/Delta(B.1.617.2)
SRR15748802	2021/6/22	USA: California	B.1.621	59	female	\	Mu(B.1.621)/Delta(AY.44)

(*continued on next page*)

Table 1 (*continued*)

ID	Collection Date	Location	Lineage	Host Age	Host Sex	CT Value	Lineage Detected
SRR15748546	2021/6/23	USA: Texas	B.1.617.2	20	female	23.63	Delta(AY.2)/Delta(B.1.617.2)
SRR15749359	2021/6/23	USA: Florida	B.1.1.7	26	female	22.56	Alpha(B.1.1.7)/Delta(AY.4.1)
SRR14812112	2021/5/17	USA: Maine	B.1.526.1	9	male	20.06	B.1.637/Alpha(Q.8)
SRR15746526	2021/6/10	USA: Florida	B.1.623	58	male	\	Mu(B.1.621.1)/Alpha(Q.8)
SRR15822113	2021/7/12	USA: California	B.1.1.7	30	female	\	Alpha(Q.2)/Delta(AY.43)
SRR15753150	2021/6/19	USA: Hawaii	B.1.1.7	17	female	\	Alpha(Q.3)/Delta(B.1.617.2)
SRR14388832	2021/3/11	USA: Illinois	B.1.1.7	21	male	19.32	B.1.1.519/Alpha(Q.8)
SRR15628163	2021/8/8	USA: California	AY.4	22	female	\	Delta(AY.44)/Delta(AY.10)
SRR15747718	2021/6/28	USA: California	B.1.1.7	29	female	\	Alpha(Q.4)/Delta(AY.44)
SRR16024429	2021/9/1	USA: Pennsylvania	B.1.617.2	9	male	\	Delta(AY.26)/Delta(AY.4)
SRR15747695	2021/6/28	USA: California	B.1.526	39	female	\	B.1.637/B.1.526
SRR14390386	2021/4/2	USA: Pennsylvania	B.1.2	21	female	23.61	Alpha(Q.8)/B.1.2
SRR15748275	2021/6/28	USA: Maine	B.1.617.2	16	female	\	B.1.637/Delta(AY.46.2)
SRR15752457	2021/6/9	USA: Washington	B.1.1.7	58	female	\	Alpha(B.1.1.7)/Gamma(P.1)
SRR15745498	2021/7/1	USA: Texas	B.1.617.2	5	female	22.68	Delta(AY.46)/Alpha(Q.4)
SRR14812102	2021/5/17	USA: Massachusetts	B.1.1.7	31	female	20.58	B.1.526/Alpha(B.1.1.7)
SRR15752416	2021/6/8	USA: Ohio	B.1.1.7	58	male	17.27	Alpha(Q.4)/Delta(AY.4)
SRR15775300	2021/7/31	USA: Florida	B.1.617.2	50	female	24.4	Delta(AY.46.3)/Delta(AY.14)
SRR15747965	2021/6/29	USA: California	B.1.617.2	29	female	\	Delta(AY.2)/Delta(B.1.617.2)
SRR14452322	2021/4/13	USA: California	B.1	21	female	19.48	B.1.526/Alpha(B.1.1.7)
SRR16026869	2021/9/1	USA: Georgia	B.1.617.2	32	male	\	Delta(AY.40)/Delta(AY.14)
SRR15748808	2021/6/21	USA: California	AY.1	47	female	\	Delta(AY.3)/Delta(AY.2)
SRR15432225	2021/7/26	USA: Wisconsin	B.1.617.2	29	female	19.11	Delta(AY.37)/Delta(B.1.617.2)
SRR15433533	2021/7/25	USA: Illinois	B.1.617.2	15	female	\	Delta(AY.21)/Delta(AY.14)
SRR15822210	2021/7/12	USA: Nevada	B.1	47	male	\	Delta(B.1.617.2)/Alpha(B.1.1.7)
SRR15431950	2021/7/24	USA: Pennsylvania	B.1.617.2	36	female	\	Delta(B.1.617.2)/Alpha(Q.4)
SRR16026382	2021/9/1	USA: Pennsylvania	AY.3	62	male	\	Delta(AY.23)/Delta(AY.26)
SRR15753790	2021/6/16	USA: California	B.1.617.2	15	male	\	Delta(AY.44)/Delta(AY.2)
SRR15753511	2021/6/17	USA: Texas	B.1	42	female	\	Delta(AY.44)/Gamma(P.1.17)
SRR14389013	2021/3/10	USA: Michigan	B.1.2	40	male	15.72	Alpha(B.1.1.7)/B.1.2
SRR15749823	2021/6/22	USA: Colorado	B.1.617.2	40	female	\	Delta(AY.44)/Alpha(Q.8)
SRR15748424	2021/6/22	USA: Missouri	B.1	20	female	20.02	B.1.628/Delta(AY.9)
SRR14812095	2021/5/17	USA: Massachusetts	B.1.1.7	27	female	18.87	Alpha(Q.8)/Delta(B.1.617.2)
SRR15747790	2021/6/30	USA: South Carolina	B.1.617.2	25	female	\	Delta(AY.43)/Alpha(Q.4)
SRR15752297	2021/6/9	USA: Florida	B.1.1.7	13	female	17.62	Alpha(B.1.1.7)/Gamma(P.1)
SRR15749954	2021/6/21	USA: Illinois	B.1.1.7	25	female	21.1	Delta(AY.21)/Alpha(B.1.1.7)
SRR15822792	2021/7/12	USA: Colorado	B.1.617.2	44	female	\	Delta(AY.35)/Delta(AY.44)
SRR15433445	2021/7/24	USA: Florida	B.1.617.2	13	male	24.69	Delta(AY.14)/Delta(B.1.617.2)
SRR14453225	2021/4/10	USA: Texas	B.1	46	female	16.33	B.1.628/Gamma(P.1)
SRR15656671	2021/8/2	USA: Georgia	B.1.617.2	53	male	16.44	Delta(AY.47)/Delta(AY.14)
SRR16026426	2021/8/31	USA: Illinois	B.1.617.2	34	female	24.83	Delta(AY.44)/Delta(AY.14)
SRR16026805	2021/9/3	USA: California	B.1.617.2	8	female	\	Delta(AY.46.3)/Delta(AY.15)
SRR14452465	2021/4/10	USA: Arizona	B.1.526	60	female	23	B.1.526/B.1.1.519
SRR15749402	2021/6/25	USA: Nevada	B.1.1.7	31	male	\	Alpha(Q.3)/Delta(AY.14)
SRR14398742	2021/3/17	USA: Florida	B.1.526	21	male	21.27	B.1.526/B.1.429
SRR16024421	2021/8/30	USA: Florida	B.1.617.2	6	male	24.68	Delta(AY.35)/Delta(AY.15)
SRR14401586	2021/2/25	USA: Texas	B.1.2	32	male	19.86	B.1.576/B.1.2
SRR15617242	2021/8/8	USA: Nevada	AY.4	26	male	\	Delta(B.1.617.2)/Gamma(P.1)
SRR14398873	2021/3/16	USA: Ohio	B.1.2	24	female	18.86	Gamma(P.1.6)/B.1.2
SRR15746184	2021/6/11	USA: Michigan	B.1.526	21	male	17.24	B.1.637/Gamma(P.1)
SRR15752552	2021/6/10	USA: Arkansas	B.1.1.7	54	female	18.14	Alpha(Q.4)/Delta(B.1.617.2)
SRR14390840	2021/4/2	USA: Pennsylvania	B.1.1.7	80	female	24.22	B.1.243/Alpha(B.1.1.7)
SRR15747961	2021/6/29	USA: California	B.1.617.2	36	female	24.8	Delta(B.1.617.2)/Alpha(B.1.1.7)
SRR15749686	2021/6/26	USA: Nevada	B.1.617.2	74	female	\	Delta(AY.2)/Delta(AY.44)
SRR14462551	2021/2/23	USA: Florida	B.1.526	34	male	22.26	B.1.526/Alpha(B.1.1.7)
SRR15616913	2021/8/7	USA: Nevada	B.1.617.2	51	female	\	Delta(AY.10)/Delta(AY.8)
SRR16025838	2021/8/31	USA: Illinois	AY.9	22	female	\	Delta(AY.46.3)/Delta(AY.9)
SRR15752556	2021/6/10	USA: Missouri	B.1.617.2	77	female	20	Alpha(Q.3)/Delta(B.1.617.2)
SRR16024695	2021/9/1	USA: California	B.1.617.2	44	male	\	Delta(B.1.617.2)/Delta(AY.14)
SRR15746210	2021/6/11	USA: New York	B.1	42	female	\	Delta(B.1.617.2)/Gamma(P.1)
SRR14450785	2021/3/4	USA: Michigan	B.1.429	68	male	22.72	B.1.637/B.1.429
SRR15752591	2021/6/9	USA: Missouri	B.1.617.2	16	male	17.21	B.1.526/Delta(B.1.617.2)
SRR15752249	2021/6/11	USA: Nevada	B.1.617.2	24	male	0	Mu(B.1.621)/Delta(AY.5.2)/Alpha(B.1.1.7)
SRR14392570	2021/4/5	USA: Pennsylvania	B.1.427	56	male	24.18	Gamma(P.1.6)/B.1.427
SRR15822184	2021/7/12	USA: Florida	B.1.621	43	male	\	Delta(AY.46.3)/Mu(B.1.621)
SRR15628087	2021/8/8	USA: New York	AY.4	17	male	\	Delta(AY.44)/Delta(AY.19)
SRR14390248	2021/3/31	USA: New Jersey	B.1	7	male	17.69	B.1.637/B.1.526
SRR15752802	2021/6/21	USA: Utah	AY.2	65	female	\	Delta(AY.2)/Delta(AY.4)
SRR16025191	2021/8/28	USA: North Carolina	B.1.617.2	66	female	19.06	Delta(AY.44)/Delta(AY.26)
SRR16024629	2021/8/30	USA: Massachusetts	B.1.617.2	35	male	21.57	Delta(AY.37)/Delta(B.1.617.2)
SRR15748036	2021/6/29	USA: Nevada	B.1.617.2	55	male	\	Delta(AY.44)/Alpha(B.1.1.7)
SRR15749841	2021/6/22	USA: Illinois	B.1.1.7	31	female	20.03	Alpha(B.1.1.7)/Delta(AY.4.1)
SRR14390511	2021/3/29	USA: Arizona	B.1.596	12	male	24	Alpha(Q.1)/B.1.596
SRR15382970	2021/7/17	USA: California	B.1.617.2	28	male	\	Delta(AY.44)/Delta(AY.16)
SRR15749767	2021/6/25	USA: Nevada	P.1	25	female	\	Gamma(P.1.1)/Alpha(B.1.1.7)

Table 1 (*continued*)

ID	Collection Date	Location	Lineage	Host Age	Host Sex	CT Value	Lineage Detected
SRR15752445	2021/6/10	USA: Oregon	B.1.617.2	32	male	\	Mu(B.1.621)/Delta(B.1.617.2)
SRR15494083	2021/7/31	USA: Florida	B.1.617.2	14	male	22.58	Delta(AY.35)/Delta(B.1.617.2)
SRR15747898	2021/6/28	USA: Florida	B.1.1.7	15	male	\	B.1.628/Alpha(B.1.1.7)
SRR15748001	2021/6/30	USA: Missouri	AY.3	39	female	21.14	Delta(AY.40)/Delta(AY.35)
SRR15748265	2021/6/30	USA: Georgia	B.1.617.2	23	male	\	B.1.637/Delta(AY.44)
SRR15628168	2021/8/8	USA: California	B.1.617.2	24	female	\	Delta(AY.14)/Delta(AY.4.2)
SRR15806459	2021/8/10	USA: Massachusetts	B.1.617.2	54	male	20.11	Delta(AY.4.2)/Delta(AY.26)
SRR15656299	2021/8/1	USA: Illinois	B.1.617.2	30	male	\	Delta(AY.21)/Delta(AY.14)
SRR15628150	2021/8/9	USA: New York	AY.4	14	male	\	Delta(AY.46.1)/Delta(AY.35)
SRR15806872	2021/8/15	USA: Florida	B.1.617.2	56	female	15.2	Delta(AY.44)/Delta(AY.39)
SRR15745147	2021/7/4	USA: Tennessee	B.1.1.7	58	male	\	Alpha(B.1.1.7)/Delta(AY.46.3)
SRR14452997	2021/4/12	USA: Massachusetts	P.1	25	male	24.87	B.1.526/Gamma(P.1.10)
SRR15432304	2021/7/27	USA: Nevada	B.1.617.2	15	male	\	Delta(AY.14)/Delta(AY.44)
SRR15747544	2021/6/29	USA: California	B.1.1.7	29	female	\	Alpha(Q.3)/Delta(B.1.617.2)
SRR15432308	2021/7/27	USA: Nevada	B.1.617.2	78	male	\	Delta(B.1.617.2)/Delta(AY.14)
SRR15822230	2021/7/11	USA: Nevada	B.1.1.7	23	female	\	Alpha(B.1.1.7)/Delta(B.1.617.2)
SRR15748858	2021/6/24	USA: Nevada	B.1.617.2	33	female	\	Delta(AY.44)/Alpha(Q.3)
SRR15822152	2021/7/12	USA: Florida	B.1.617.2	72	male	\	Delta(B.1.617.2)/Mu(B.1.621.1)
SRR15752349	2021/6/9	USA: Texas	B.1.617.2	31	male	23.4	Delta(B.1.617.2)/Gamma(P.1)
SRR15433608	2021/7/20	USA: Texas	B.1.617.2	63	female	21.69	Delta(AY.35)/Delta(AY.25)
SRR16025203	2021/8/28	USA: Tennessee	B.1.617.2	24	female	13.56	Delta(AY.14)/Delta(AY.47)
SRR14812107	2021/5/17	USA: Massachusetts	B.1.1	40	male	17.2	B.1.526/Alpha(B.1.1.7)
SRR15749687	2021/6/26	USA: Nevada	B.1.617.2	39	female	\	Delta(AY.30)/Delta(AY.44)
SRR15752528	2021/6/11	USA: Missouri	B.1.617.2	60	male	19.01	Delta(B.1.617.2)/Alpha(B.1.1.7)
SRR15750502	2021/6/28	USA: California	B.1.1.7	57	female	\	B.1.628/Alpha(B.1.1.7)
SRR15494227	2021/7/31	USA: Georgia	AY.12	18	male	\	Delta(AY.26)/Delta(AY.4)
SRR15774966	2021/7/31	USA: Washington	B.1.617.2	78	male	\	Delta(AY.21)/Delta(AY.15)
SRR15783924	2021/8/23	USA: Nevada	B.1.617.2	85	female	\	Delta(AY.2)/Delta(B.1.617.2)
SRR15741958	2021/8/14	USA: Texas	AY.12	36	male	\	Delta(AY.21)/Delta(AY.14)
SRR16026646	2021/8/31	USA: Nevada	B.1.617.2	6	male	\	Delta(AY.26)/Delta(AY.4.1)
SRR15749444	2021/6/22	USA: California	B.1.1.7	23	female	\	Delta(AY.44)/Alpha(Q.3)
SRR14393680	2021/4/4	USA: Michigan	B.1.1.7	21	male	19.97	Alpha(B.1.1.7)/B.1.429
SRR15432914	2021/7/24	USA: Texas	B.1.617.2	43	male	\	Delta(AY.21)/Delta(AY.14)
SRR16024661	2021/8/30	USA: Connecticut	AY.3	32	female	\	Delta(AY.4)/Delta(AY.26)
SRR15383414	2021/7/19	USA: Missouri	B.1.617.2	28	male	15.45	Delta(AY.44)/Alpha(Q.8)
SRR15383173	2021/7/18	USA: New York	B.1.617.2	29	male	22.69	Delta(AY.46.3)/Delta(AY.26)
SRR15433019	2021/7/26	USA: Michigan	B.1.617.2	11	male	11.87	Delta(AY.21)/Delta(AY.26)
SRR15748054	2021/6/29	USA: Nevada	B.1.617.2	40	male	\	Delta(B.1.617.2)/Alpha(Q.1)
SRR15432222	2021/7/26	USA: Wisconsin	B.1.617.2	27	female	16.64	Delta(AY.37)/Delta(B.1.617.2)
SRR15753017	2021/6/19	USA: California	B.1.1.7	28	male	\	Alpha(B.1.1.7)/Delta(B.1.617.2)
SRR15747878	2021/6/29	USA: Missouri	B.1.620	34	male	24.97	Delta(B.1.617.2)/Alpha(B.1.1.7)
SRR15746381	2021/6/10	USA: Texas	B.1.617.2	30	female	\	Delta(AY.21)/Delta(AY.15)
SRR14451452	2021/3/5	USA: Massachusetts	B.1.361	62	female	21.67	B.1.637/B.1.568
SRR15383119	2021/7/18	USA: Georgia	B.1.617.2	6	female	\	Delta(AY.46.3)/Delta(AY.26)
SRR15748611	2021/6/22	USA: California	B.1.617.2	23	male	\	Delta(B.1.617.2)/Alpha(B.1.1.7)
SRR14392657	2021/4/4	USA: Arizona	B.1.1.7	46	female	23	Alpha(B.1.1.7)/B.1.429
SRR15775298	2021/7/31	USA: Florida	B.1.617.2	41	male	19.71	Delta(AY.25)/Delta(AY.14)
SRR14812183	2021/5/16	USA: Massachusetts	B.1.1.7	40	female	21.66	Alpha(Q.8)/Delta(B.1.617.2)
SRR15907338	2021/8/23	USA: Connecticut	B.1.617.2	17	female	17.49	Delta(AY.37)/Delta(AY.25)
SRR15656870	2021/8/3	USA: Hawaii	B.1.617.2	47	female	22.3	Delta(AY.46)/Delta(AY.14)
SRR15749399	2021/6/25	USA: Nevada	B.1.617.2	47	female	\	Delta(AY.21)/Delta(AY.8)
SRR15752558	2021/6/10	USA: Missouri	B.1.617.2	42	male	17.47	Delta(B.1.617.2)/Alpha(B.1.1.7)
SRR15628034	2021/8/11	USA: Washington	AY.4	80	female	\	Delta(AY.44)/Delta(AY.26)
SRR15750467	2021/6/8	USA: Arkansas	B.1.617.2	34	male	20.76	Alpha(Q.3)/Delta(AY.26)
SRR15822162	2021/7/12	USA: Florida	B.1.621.1	36	female	\	Mu(B.1.621.1)/Delta(B.1.617.2)
SRR15749330	2021/6/26	USA: California	P.1	57	female	\	Delta(AY.4.3)/Gamma(P.1)
SRR15493850	2021/7/25	USA: Michigan	B.1.617.2	32	male	21.45	Delta(AY.44)/Delta(AY.19)
SRR14450921	2021/2/27	USA: Pennsylvania	B.1.429	63	male	24.73	B.1.575/B.1.429
SRR15432981	2021/7/20	USA: Minnesota	B.1.617.2	29	female	23.15	Delta(AY.14)/Delta(B.1.617.2)
SRR14399607	2021/3/15	USA: Illinois	B.1.427	47	female	21.83	Alpha(Q.4)/B.1.427
SRR14448650	2021/2/25	USA: Pennsylvania	B.1.1.7	23	male	13.7	Alpha(B.1.1.7)/B.1.429
SRR15745144	2021/7/1	USA: South Carolina	B.1.1.7	36	male	\	Alpha(B.1.1.7)/Delta(B.1.617.2)
SRR15775075	2021/8/1	USA: Illinois	B.1.617.2	31	female	20.84	Delta(AY.25)/Delta(AY.26)
SRR14812179	2021/5/16	USA: Maine	B.1.526.2	43	male	21.08	B.1.526/Alpha(B.1.1.7)
SRR16024870	2021/8/30	USA: California	B.1.617.2	28	female	\	Delta(AY.21)/Delta(AY.26)
SRR15822915	2021/7/11	USA: Florida	B.1.617.2	29	female	\	Delta(AY.44)/Alpha(B.1.1.7)
SRR14391243	2021/4/9	USA: Ohio	B.1.1.7	34	female	19.27	Alpha(B.1.1.7)/B.1.2
SRR15742766	2021/8/16	USA: Oregon	B.1.617.2	29	female	\	Delta(AY.44)/Delta(AY.15)
SRR15822505	2021/7/10	USA: Texas	B.1.1.7	26	male	22.01	Alpha(Q.4)/Delta(B.1.617.2)
SRR15749161	2021/6/25	USA: Nevada	B.1.1.7	31	male	\	Alpha(Q.3)/Delta(AY.44)
SRR14395855	2021/3/24	USA: Texas	B.1	17	male	24.47	B.1.627/A.2.5.2
SRR15783519	2021/8/19	USA: Hawaii	B.1.617.2	6	male	\	Delta(AY.14)/Delta(AY.7.2)
SRR15753604	2021/6/21	USA: Texas	P.1	19	female	\	Delta(AY.44)/Gamma(P.1)
SRR14811846	2021/5/15	USA: Ohio	B.1.1.7	66	female	20.74	B.1.637/Alpha(B.1.1.7)
SRR14451894	2021/4/14	USA: Florida	B.1.526	34	male	15.03	B.1.526/Alpha(B.1.1.7)

(*continued on next page*)

Table 1 (*continued*)

ID	Collection Date	Location	Lineage	Host Age	Host Sex	CT Value	Lineage Detected
SRR14812093	2021/5/17	USA: Connecticut	B.1.1.7	18	female	23.75	B.1.526/Alpha(B.1.1.7)/Gamma(P.1)
SRR15746093	2021/6/11	USA: Florida	B.1.1.7	35	male	15.63	Alpha(B.1.1.7)/Gamma(P.1)
SRR15752434	2021/6/11	USA: Oregon	P.1.1	66	female	\	Gamma(P.1)/Alpha(Q.7)
SRR15628426	2021/8/8	USA: New Jersey	B.1.617.2	41	male	19	Delta(AY.14)/Delta(AY.19)
SRR15742679	2021/8/17	USA: Pennsylvania	B.1.617.2	44	female	\	Delta(AY.46.3)/Delta(AY.15)
SRR15432236	2021/7/27	USA: Nevada	B.1.617.2	58	female	\	Delta(B.1.617.2)/Mu(B.1.621.1)
SRR15742511	2021/8/17	USA: Massachusetts	B.1.617.2	21	female	20.09	Delta(AY.21)/Delta(AY.26)
SRR14152504	2021/3/14	USA: Michigan	B.1.1.7	16	female	23.48	Alpha(B.1.1.7)/B.1.2
SRR14152550	2021/3/16	USA: Massachusetts	B.1.1.7	47	female	25.85	B.1.396/Alpha(B.1.1.7)
SRR14152575	2021/3/15	USA: Georgia	B.1.526	29	male	22.55	B.1.526/Beta(B.1.351)
SRR14152615	2021/3/16	USA: Massachusetts	B.1.1.7	18	female	26.79	Alpha(B.1.1.7)/B.1.2
SRR14152622	2021/3/16	USA: Michigan	B.1.526.1	60	male	21.46	B.1.637/Alpha(B.1.1.7)
SRR14153082	2021/3/15	USA: Georgia	B.1.1.7	65	female	23.7	B.1.526/Alpha(B.1.1.7)
SRR14153096	2021/3/16	USA: Pennsylvania	B.1.526	58	female	22.18	B.1.526/B.1.2
SRR14154656	2021/3/13	USA: Pennsylvania	B.1.429	64	female	24.13	Alpha(Q.4)/B.1.429
SRR14154687	2021/3/14	USA: Florida	B.1.1.7	41	male	22.26	Alpha(B.1.1.7)/B.1.2
SRR14154713	2021/3/14	USA: Florida	B.1.1.7	41	female	20.75	B.1.526/Alpha(B.1.1.7)
SRR14154901	2021/3/14	USA: Texas	B.1.1.7	24	male	16.99	B.1.1.519/Alpha(B.1.1.7)
SRR14156532	2021/2/12	USA: California	B.1.404	42	male	18.74	B.1.561/B.1.2
SRR14157283	2021/2/26	USA: Georgia	B.1.1.7	33	male	25.29	B.1.637/Alpha(Q.3)
SRR14157409	2021/2/23	USA: Florida	B.1.429	19	female	17.38	Alpha(Q.4)/B.1.429
SRR14157800	2021/3/1	USA: Florida	B.1.1.7	26	male	26.42	Alpha(B.1.1.7)/B.1.526
SRR14157810	2021/3/3	USA: Minnesota	B.1.1.7	56	male	19.13	B.1.526/Alpha(B.1.1.7)
SRR14157910	2021/3/1	USA: Georgia	B.1.2	30	female	21.83	B.1.2/B.1.429
SRR14158337	2021/3/2	USA: Pennsylvania	B.1.526	35	female	17.68	B.1.526/B.1.427
SRR14158374	2021/3/3	USA: Michigan	B.1	18	male	23.26	B.1.637/Alpha(B.1.1.7)
SRR14158401	2021/3/3	USA: Georgia	B.1.2	47	female	20.51	B.1.526/B.1.2

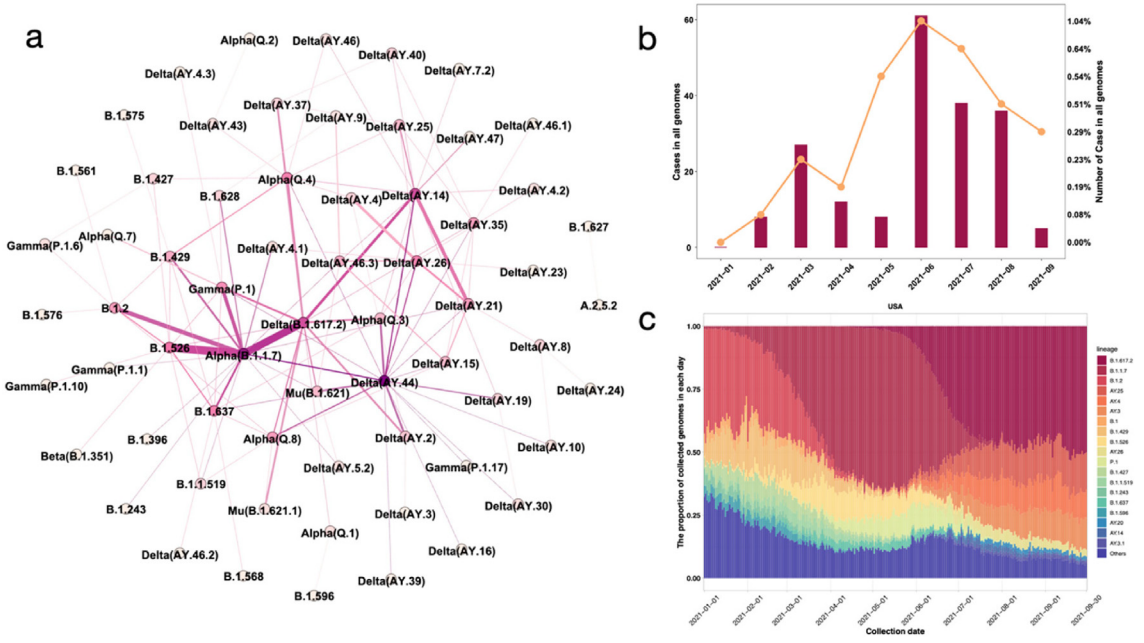

Fig. 4. Distribution of co-infection events according to lineage and collection date. a. Co-infected lineage network for all 175 identified co-infection samples. Every dot represents a lineage; the color depth of each lineage is associated with the occurrence number of this lineage in co-infection events. The thickness of the line between dots represents the co-occurrence degree of the linked lineages. b. The number and ratio of co-infection samples varied with time and dominant lineage. c. The co-circulation pattern of SARS-CoV-2 lineages in the United States from January 1 to September 30, 2021, when the B.1.2 lineage was outcompeted by Alpha (B.1.1.7), and that from April 2021 to June 2021, when the Alpha and Gamma lineages were the two major co-existing lineages. Later, from June 2021 onward, the Delta lineage began to outcompete all other lineages. *Data were collected until September 30, 2021.

four variations mentioned above, another 395 viral strains could be found. Apparently, different from the situation of the B.1.526 lineage, 28 of all 396 strains with the four undetermined variations were isolated in Maine, while most of the strains with these mutations were isolated in California and Texas, demonstrating a complex introduction of the co-infected Alpha lineage into Maine.

The distribution of co-infection events is both region-dependent and time-dependent, indicating that the occurrence of co-infection results from the interaction between at least two co-circulating SARS-CoV-2 lineages at that specific time and specific location. For instance, we found that co-infection events have lineage-bias (Fig. 4A) and increased with time in the United States

(Fig. 4B). One possible explanation for this phenomenon is the quick switch of the dominant lineages in the country during the first nine months of 2021. To be specific, with the change in dominant lineage in the United States from Alpha to Delta (B.1.617.2), the center co-infection lineage also changed from Alpha to Delta (B.1.617.2). However, from June 2021 onward, the co-infection situation changed from having one center lineage co-infected with other co-circulating lineages to multicenter lineages. In the previous variation of co-infection center, B.1.2, Alpha (B.1.1.7), and Delta (B.1.617.2) had different infection abilities. After Delta outcompeted all other lineages beginning around June 2021, Delta descendant lineages formed in different regions. The similar biological properties of Delta descendant lineages might prolong the co-infection time of two different lineages in the same patient. This might be why more co-infection cases were observed after the Delta lineage became dominant. Although the present situation of dominant variation is stable, this significantly improved co-infection rate might contribute to a new recombined variant.

Until early 2022, three large waves of SARS-CoV-2 pandemics have occurred with Alpha, Delta, and Omicron as the dominant variants in turn. It is worth noting that relatively higher co-infection rate was observed in the transition period of dominant variants, which indicates the urgent need to monitor the co-infected events for the recent transition from Delta variant to Omicron variant in global. In addition, we must point out that huge genetic diversity will quickly occur within the dominant SARS-CoV-2 variant with its evolution and divergence, such as the Delta variant and Omicron variant. Therefore, co-infection is still a critical problem with the co-circulation of multiple sub-lineages of the dominant variant. Recent studies have provided robust evidence of potential recombination events of different SARS-CoV-2 variants (https://github.com/cov-lineages/pango-designation/issues), that occurs due to the SARS-CoV-2 co-infections. Furthermore, SARS-CoV-2 has been reported to spill over to many wild animals and has evolved to new lineages [40]. The co-infection of these animal derived SARS-CoV-2 lineages might cause new recombinants with high genetic diversity with the dominant SARS-CoV-2 and pose a new threat to public health. In our opinion, strict epidemic prevention and control measures are important for reducing the number of co-infected patients, which is also better for reducing the possibility of SARS-CoV-2 recombination.

5. Author statement

All authors have seen and approved the final version of the manuscript being submitted. They warrant that the article is the authors' original work, hasn't received prior publication and isn't under consideration for publication elsewhere.

6. Data availability

The workflow for calling iSNVs can be retrieved from (https://dockstore.org/workflows/github.com/iwc-workflows/sars-cov-2-pe-illumina-artic-variant-calling/COVID-19-PE-ARTIC-ILLUMINA:main?tab=info). The identification numbers of screened samples and the homemade Python script for identifying potential co-infection events are available online (https://github.com/wuaipinglab/SARS-CoV-2_co-infection). All the detected co-infection samples could be found in the above link as well.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Dakai Liu from Department of Pathology and Clinical Laboratories, NewYork-Presbyterian Queens hospital for his kindly help in proofreading the original manuscript. We also thank Feng Qian, General Manager of Suzhou Supercomputing Center (SISCC), for the HPC resource support. In addition, we would like to thank the SISCC team for their great contribution to accelerating the scientific research process. We gratefully acknowledge all data contributors, i.e. the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

Funding

We appreciate the support we received from the National key research and development program (2021YFC2301300), the CAMS Innovation Fund for Medical Sciences (2021-I2M-1-061), the National Natural Science Foundation of China (92169106), the special research fund for central universities, Peking Union Medical College (2021-PT180-001), China postdoctoral science foundation grants (2019M660548 and 2020T130007ZX), the Suzhou science and technology development plan (szs2020311), the Youthful Teacher Project of Peking Union Medical College (3332019114) and the Beijing Municipal Commission of Health (shoufa-1G-1131).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.07.042.

References

[1] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 2020;395:1054–62.
[2] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020;20:533–4.
[3] Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 2020;5:1403–7.
[4] Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, Pearson CA, Russell TW, Tully DC, Washburne AD. 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. Science 372.
[5] Hoffmann M, Arora P, Groß R, Seidel A, Hörnich BF, Hahn AS, et al. SARS-CoV-2 variants B. 1.351 and P. 1 escape from neutralizing antibodies. Cell 2021;184(2384–2393):e12.
[6] Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature:1–7 2021.
[7] Liu C, Ginn HM, Dejnirattisai W, Supasa P, Wang B, Tuekprakhon A, et al. Reduced neutralization of SARS-CoV-2 B. 1.617 by vaccine and convalescent serum. Cell 2021.
[8] Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, Anyaneji UJ, Bester PA, Boni MF, Chand M. 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature:1-10.
[9] Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature 2022;602:657–63.
[10] Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. The Lancet 2021;398:2126–8.
[11] Tillett RL, Sevinsky JR, Hartley PD, Kerwin H, Crawford N, Gorzalski A, et al. Genomic evidence for reinfection with SARS-CoV-2: a case study. Lancet Infect Dis 2021;21:52–8.
[12] Sabino EC, Buss LF, Carvalho MP, Prete CA, Crispim MA, Fraiji NA, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. The Lancet 2021;397:452–5.
[13] He Y, Ma W, Dang S, Chen L, Zhang R, Mei S, et al. Possible recombination between two variants of concern in a COVID-19 patient. Emerging Microbes & Infections 2022:1–26.
[14] Jackson B, Boni MF, Bull MJ, Colleran A, Colquhoun RM, Darby AC, et al. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. Cell 2021;184(5179–5188):e8.

[15] Varabyou A, Pockrandt C, Salzberg SL, Pertea M. Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. Genetics 2021;218(iyab074).
[16] da Silva Francisco Jr R, Benites LF, Lamarca AP, de Almeida LG, Hansen AW, Gularte JS, Demoliner M, Gerber AL, de C Guimarães AP, Antunes AKE. Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. Virus Res 2021;296:198345.
[17] Tonkin-Hill G, Martincorena I, Amato R, Lawson AR, Gerstung M, Johnston I, Jackson DK, Park NR, Lensing SV, Quail MA. 2020. Patterns of within-host genetic diversity in SARS-CoV-2. BioRxiv.
[18] Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, et al. SARS-CoV-2 within-host diversity and transmission. Science 2021;372.
[19] Hashim HO, Mohammed MK, Mousa MJ, Abdulameer HH, Alhassnawi AT, Hassan SA, et al. Infection with different strains of SARS-CoV-2 in patients with COVID-19. Arch Biol Sci 2020;72:575–85.
[20] Samoilov A, Kaptelova V, Bukharina A, Shipulina O, Korneenko E, Lukyanov A, Grishaeva A, Ploskireva A, Speranskaya AS, Akimkin V. 2020. Change of dominant strain during dual SARS-CoV-2 infection. medRxiv.
[21] Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F, et al. Dual HIV-1 infection associated with rapid disease progression. The Lancet 2004;363:619–22.
[22] Van der Kuyl AC, Cornelissen M. Identifying HIV-1 dual infections. Retrovirology 2007;4:1–12.
[23] Ekouevi DK, Eholie SP. 2021. Update on HIV-1 and HIV-2 dual infection, p 1-10. In Hope TJ, Stevenson M, Richman D (ed), Encyclopedia of AIDS doi: 10.1007/978-1-4614-9610-6_49-1. Springer New York, New York, NY.
[24] Su S, Wong G, Shi W, Liu J, Lai AC, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol 2016;24:490–502.
[25] Zhu Z, Meng K, Meng G. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. Sci Rep 2020;10:1–10.
[26] Neches RY, McGee MD, Kyrpides NC. 2020. Recombination should not be an afterthought. Nature Reviews Microbiology 18:606-606.
[27] Terada Y, Matsui N, Noguchi K, Kuwata R, Shimoda H, Soma T, et al. Emergence of pathogenic coronaviruses in cats by homologous recombination between feline and canine coronaviruses. PLoS ONE 2014;9(9):e106534.
[28] Xiao Y, Rouzine IM, Bianco S, Acevedo A, Goldstein EF, Farkov M, et al. RNA recombination enhances adaptability and is required for virus spread and virulence. Cell Host Microbe 2016;19(4):493–503.
[29] Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J Virol 2010;84(7):3134–46.
[30] Jackwood MW, Boynton TO, Hilt DA, McKinley ET, Kissinger JC, Paterson AH, et al. Emergence of a group 3 coronavirus through recombination. Virology 2010;398(1):98–108.
[31] Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J Virol 2007;81(14):7620–8.
[32] Elie B, Lecorche E, Sofonea MT, Trombert-Paolantoni S, Foulongne V, Guedj J, Haim-Boukobza S, Roquebert B, Alizon S. 2021. Inferring SARS-CoV-2 variant within-host kinetics. medRxiv.
[33] Ruan Y, Hou M, Li J, Song Y, Wang H-Y, Zeng H, Lu J, Wen H, Chen C, Wu C-I. 2021. One viral sequence for each host?–The neglected within-host diversity as the main stage of SARS-CoV-2 evolution. bioRxiv.
[34] Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges 2017;1:33–46.
[35] Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance 2017;22:30494.
[36] Maier W, Bray S, van den Beek M, Bouvier D, Coraor N, Miladi M, et al. Ready-to-use public infrastructure for global SARS-CoV-2 monitoring. Nat Biotechnol 2021;39:1178–9.
[37] Kemp SA, Collier DA, Datir RP, Ferreira IA, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature 2021;592:277–82.
[38] Kim KW, Deveson IW, Pang CNI, Yeang M, Naing Z, Adikari T, et al. Respiratory viral co-infections among SARS-CoV-2 cases confirmed by virome capture sequencing. Sci Rep 2021;11:1–9.
[39] Bull RA, Adikari TN, Ferguson JM, Hammond JM, Stevanovski I, Beukers AG, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. Nat Commun 2020;11:1–8.
[40] Hale VL, Dennis PM, McBride DS, Nolting JM, Madden C, Huey D, et al. SARS-CoV-2 infection in free-ranging white-tailed deer. Nature 2022;602(7897):481–6.