Supplementary Materials

SUPPLEMENTARY MATERIALS AND METHODS

Coronavirus GenBrowser

The data used in this study were obtained from the Coronavirus GenBrowser (CGB) (Yu et al., 2022), which offers a panoramic vision of the transmission and evolution of SARS-CoV-2. The CGB global data file was "data.2021-09-02.tar.xz", which can be freely downloaded from https://ngdc.cncb.ac.cn/ncov/apis/archives/. The data file (21.8 Mb) was obtained by analyzing 1 610 125 high-quality SARS-CoV-2 genomic sequences. The data contained the tip-dated tree, pre-analyzed genomic mutations, and epidemiologically associated meta-data. The latest CGB global dataset (data.2022-06-03.tar.xz, n=4 039 521 genomic sequences) was used to confirm our conclusions.

Note, CGB only analyzes the sequence corresponding to the reference sequence (GenBank accession number: NC_045512) (Wu et al., 2020) between nucleotide positions 100 and 29 800. Thus, the considered genome sequence length was 29 701 bases in this study.

CGB lineage tracing

Here, CGB (Yu et al., 2022) lineage tracing was used in the Dalian outbreak to analyze strains isolated from the environment (i.e., outer packages). The query sequences were aligned with the reference genome (NC_045512) (Wu et al., 2020). The aligned sequences were used as input for lineage tracing, and the closest nodes to each query sequence (measured as the similarity of two genomic sequences) were identified on the tree.

Identification of mutation-dormant variants

To search for mutation-dormant variants in the SARS-CoV-2 evolutionary tree, we started from the root and traversed the tree. For each node, we recorded its direct and indirect non-mutated descendants. For a mutation-dormant variant, its repetitive cold-chain related transmissions may cause a long non-mutated path (LNMP) in the tip-dated evolutionary tree. In this study, we used the following criteria to identify mutation-dormant variants. First, there must be at least two non-mutated strains with no ambiguous nucleotides. Second, the minimum difference in collection date between the earliest and latest non-mutated strains should be greater than 100 days. Third, the LNMP should not depend on the existence of a single non-mutated strain. That is, if one non-mutated strain is removed, the LNMP should remain detectable. Fourth, neither side of the LNMP should rely on non-mutated strains submitted by a single organization. If the longest branch in the LNMP is larger than 30 days, strains on either side of the longest branch should be collected by two or more submitting organizations. After applying the above requirements, mutation-dormant variants were identified.

Detection of significantly reduced evolutionary rate in a lineage

The genome-wide evolutionary rate of SARS-CoV-2 was estimated as $\mu = 9.69 \times 10^{-4}$ per nucleotide per year (Yu et al., 2022) or 7.88×10^{-2} per genome per day. With the Poisson probability, we can determine whether the evolutionary rate is significantly reduced (Wang et al., 2020; Yu et al., 2022), that is:

$$P(x \leq \xi_{obs}) = \sum_{x=0}^{\xi_{obs}} e^{-\xi_{exp}} \, \xi_{exp}^{x} \, / x!,$$

where ξ_{obs} is the observed number of mutations on the branches, $\xi_{exp} = l\mu$, l is the summation length of branches related to the mutation-dormant variants, and μ is the evolutionary rate per genome per day. The test is one-tailed and lineage-specific and is therefore independent of SARS-CoV-2 demography.

The test can be easily extended to consider evolutionary rate heterogeneity. When different genes evolve at different rates, the *P*-value can be calculated using $\prod_i P(x \leq \xi_{obs,i})$, where $\xi_{obs,i}$ is the observed number of mutations on the *i*-th gene. The evolutionary rate for each gene is given in Supplementary Table S1.

The test was performed to detect a mutation-dormant variant in a clade. As each strain is examined, the number of strains in the clade is the number of tests performed. If the clade has *n* strains, the Bonferroni correction (Dunn, 1961) is performed to adjust the threshold for multiple tests (desired overall Type I error rate, i.e., family-wise error rate, is 0.05). The *P*-value threshold is $0.05/n$.

Let us consider the example shown in Supplementary Figure S8, and test whether the evolutionary rate is reduced between node A and strain 5. In this case, $\xi_{obs} = 2$ and l is the duration of time between node A and strain 5, i.e., 14 days. Then, $\xi_{exp} = l\mu = 14 \times \mu$ and P can be obtained according to the equation. There are seven strains in the clade, so seven tests are performed. The minimum *P*-value is observed on strain 1 because there are only two strains (strains 1 and 3) lacking mutations since the clade emerged, and strain 3 was collected earlier than strain 1. If the minimum *P*-value is less than the adjusted threshold for multiple tests, the variant is a mutation-dormant variant. As those strains share one evolutionary tree, tests among different strains are not independent, requiring Bonferroni correction.

Evolutionary tree of a mutation-dormant variant

To obtain the evolutionary tree of a mutation-dormant variant, each node from the rooted CGB evolutionary tree (*n*=1 610 125) was compared with the mutation-dormant variant. All nodes with genetic differences were filtered out, finally obtaining a rooted evolutionary tree for the mutation-dormant variant (Figure 1E, Supplementary Figures S2, S5, and S6).

In practice, the search process is extremely fast. When starting from the node of the variant, there is no need to traverse the entire tree to find all strains with identical genomic sequences. Therefore, the non-mutated path can be identified easily. An example is provided in the Supplementary Movie, which shows how to obtain an

evolutionary tree using the CGB.

Distribution of evolutionary rate among SARS-CoV-2 strains

To obtain the observed distribution of the evolutionary rate of SARS-CoV-2 strains, the evolutionary rate was calculated for each strain, i.e., number of mutations in each strain divided by its lineage length, measured as the number of days between the collection date and time of MRCA of analyzed strains. Genome sequence length was considered in the calculation. In total, 1 610 125 SARS-CoV-2 genomic sequences were analyzed.

To obtain the simulated distribution of the evolutionary rate of SARS-CoV-2 strains without cold-chain related transmission, the annotated tip-dated tree was obtained from the CGB (n=1 610 125) (Yu et al., 2022). To assign mutations for each branch randomly, the Poisson probability was applied. That is:

$$P(X = k_i) = e^{-\xi_{exp,i}} \xi_{exp,i}^{k_i} / k_i!$$

where k_i is the simulated number of mutations on the i-th branch, $\xi_{exp,i} = l_i \mu$, l_i is the length of the i-th branch, and μ is the evolutionary rate per genome per day. Here, the genome-wide evolutionary rate was 9.69×10^{-4} per site per year or 7.88×10^{-2} per genome per day in this study, unless mentioned otherwise. There were 2 236 398 branches in the CGB evolutionary tree. Finally, the simulated number of mutations in each strain was calculated, and the distribution of the evolutionary rate was obtained from the simulated data, in which there was no cold-chain related transmission.

SUPPLEMENTARY RESULTS

Probability of a mutation-dormant variant

When SARS-CoV-2 is transmitted through social contact, the virus duplicates in human hosts with a certain error rate (i.e., mutation rate). In this process, mutations are inevitable if the considered duration is long. Therefore, under the null hypothesis that SARS-CoV-2 is transmitted through social contact, the probability that SARS-CoV-2 does not mutate within 100 days in one lineage is very small ($P = 3.76 \times 10^{-4}$). This holds even when using the lower bound of the confidence interval of the estimated evolutionary rate ($P = 2.70 \times 10^{-3}$). The test is lineage-specific, rather than considering the whole viral population. If the null hypothesis is rejected, a mutation-dormant variant is detected.

As different regions of SARS-CoV-2 may have different evolutionary rates (Yu et al., 2022), we examined whether evolutionary rate heterogeneity affects P-value calculation. Let us consider two viral genes, A and B. Given the duration of time t, the probability that the virus does not mutate in one lineage is $e^{-t\mu_A}$ and $e^{-t\mu_B}$, where μ_A and μ_B are the evolutionary rates for gene A and B, respectively. The probability that the virus does not mutate in both genes is the product of the two probabilities, i.e., $e^{-t\mu_A} . e^{-t\mu_B} = e^{-t(\mu_A+\mu_B)} = e^{-t\mu_{A+B}}$, indicating that the evolutionary rate of two or

multiple genes can be combined. Therefore, evolutionary rate heterogeneity does not affect *P*-value calculation when no mutations occur in the lineage. If mutations occur, evolutionary rate heterogeneity may influence *P*-value calculation. In this case, evolutionary rate heterogeneity can be considered by taking different evolutionary rates for different genes (see Methods).

Cold-chain related outbreak in Xinfadi (Beijing, China)

In June 2020, there was an outbreak of COVID-19 in Xinfadi (Beijing, China) (Pang et al., 2020; Zhang et al., 2020). Epidemiological evidence indicated that a booth selling imported salmon may be the origin of infection (Pang et al., 2020). We found that the sequences of two isolates (Beijing/IVDC-02-06 and Beijing/BJ0617-01-Y) collected from two Xinfadi cases on 11 and 14 June 2020 were identical to the ancestral variant sequence (CGB4268.5142) (Supplementary Figure S1) dated 13 March 2020 (95% CI: 22 February–14 March 2020). The descendant of this variant was first isolated from a patient in Taoyuan who had recently traveled to several European countries (Gong et al., 2020), and was subsequently found in the Czech Republic, Denmark, India, Colombia, Ecuador, Slovakia, Beijing, and the United States. The two Xinfadi strains showed significantly slow evolution ($P = 8.28 \times 10^{-4}$and 6.53×10^{-4}, respectively) after Bonferroni correction (Dunn, 1961) for multiple tests ($n = 22$), with no mutations detected between March and June 2020, consistent with the cold-chain related transmission hypothesis.

Cold-chain related outbreak in Auckland (New Zealand)

We further validated our evolutionary genomic analysis method using an outbreak outside of China, for which partial epidemiological information is available. In August 2020, there was an outbreak of COVID-19 in Auckland after 102 days of no new locally transmitted cases in New Zealand (Supplementary News 3; Supplementary Figure S2). The first person that self-reported illness was an employee at a cool storage facility in Auckland. The employee had taken sick leave from August 3 and, together with three family members, tested positive on August 11. The viral lineage of the outbreak can be clearly observed in the evolutionary tree (Supplementary Figure S2C). All patients carried the Auckland characteristic mutation T15867G, together with 11 other mutations. Thus, we searched whole viral samples (n=1 610 125) to identify strains with the 11 exact mutations. We found 120 strains collected from March to August 2020 with identical genomic sequences (Supplementary Table S2). These strains contained the 11 mutations and were found in 23 countries (Supplementary Figure S2A), indicating that the variant did not mutate in this lineage in more than six months. The evolutionary rate was significantly low ($P = 3.65 \times 10^{-7}$) after Bonferroni correction (Dunn, 1961) for multiple tests (n = 8 437), indicating that the outbreak may be due to a spillover event from a cold-chain. Global transmission of the mutation-dormant variant was summarized in Supplementary Figure S2B, but the origin of the contaminated cold-chain products remains unknown. Moreover, the time of the MRCA of the Auckland strain was inferred to be 1 August 2020, consistent with the known epidemiological information.

Thus, we suggest that the Auckland characteristic mutation T15867G may have occurred in the employee before the virus was transmitted in the local population. Overall, these results indicate that evolutionary genomic analysis can be used to reliably detect cold-chain related SARS-CoV-2 transmission outside of China.

Putative cold-chain related spread of D614G variant and Delta variant of concern (VOC)

The first beneficial mutation was an amino-acid change (D614G) in the spike protein (Zhou et al., 2021). The variant carries three other mutations (C3037T, C14408T, and C241T). Interestingly, this variant may have emerged in Europe as early as September 2019 (Amendola et al., 2021; Ruan et al., 2022). We found that, among the 1 610 125 high-quality genomic sequences, this variant has 783 identical descendants collected on six continents (49 countries/regions) between February 2020 and May 2021 (Supplementary Figure S5A; Supplementary Table S2). Most of the descendants (675/783=86.2%) were found in Europe and North America, but the highest frequency was observed in Africa (Supplementary Figures S5B, C). As the variant did not mutate in the lineage in more than a year, its evolutionary rate was significantly low in the lineage ($P = 1.54 \times 10^{-17}$) after Bonferroni correction (Dunn, 1961) for multiple tests (n = 1 593 245), suggesting cold-chain related transmission. The results remained significant even when using the lower bound of the confidence interval of the estimated evolutionary rate ($P = 2.43 \times 10^{-13}$). As expected, this was a lineage-specific effect. After the spillover events from the cold-chain, the mutation-dormant variant re-started mutating normally. Three spillover clades were identified through genome sampling (Supplementary Figure S5A). Thus, we found evidence of cold-chain related transmission in the spread of the spike D614G variant.

We then examined the latest Coronavirus GenBrowser (CGB) global dataset (data.2022-06-03, n=4 0395 21 genomic sequences) (Yu et al., 2022) and found two recent spillover events in Turkey and the USA in August 2021. No other spillover events were identified within the next nine months (between September 2021 and May 2022). The long-term absence of spillover events in the mutation-dormant D614G variant indicates chronergy of cold-chain related transmission because spillover relies on exposure to contaminated and infectious cold-chain products.

The Delta VOC (B.1.617.2 lineage) spread rapidly around the world in 2021. We searched Delta variants that lacked mutations for more than 100 days. We identified 15 mutation-dormant variants (Supplementary Table S4). One variant (CGB531065.591827) appeared on 5 March 2021, with 34 identical descendant strains found within the next five months in eight countries without accumulating any new mutations. Another variant (CGB692105.692163) appeared on 9 April 2021, and it has a large number of non-mutated descendants (4 582), most of which were found in the United Kingdom (99%, or 4 539/4 582). In addition, we found three linearly connected mutation-dormant variants separated by one mutation (Supplementary Figure S6; Supplementary Table S2), suggesting that patients responsible for recurrent cold-chain contaminations were epidemiologically connected.

Detection of mutation-dormant variants using millions of genomic sequences

The method used in this study has been implemented in the CGB (Yu et al., 2022). Thus, newly emerged cold-chain related transmissions can be detected using the CGB. The CGB is available as a plug-in module of eGPS (http://www.egps-software.net/egpscloud/eGPS_Desktop.html) (Yu et al., 2019). It not only enables users who have no programming skills to analyze millions of genomic sequences, but also offers an effective surveillance tool for cold-chain related transmission. An example is shown in the Supplementary Movie.

SUPPLEMENTARY DISCUSSION

Robustness of pandemic-scale phylogenomics method

In this study, we developed a new method to detect cold-chain related SARS-CoV-2 transmissions, suitable for non-epidemic and epidemic areas of COVID-19. We provided a rough estimate for the frequency of cold-chain related transmissions (i.e., spillover events from cold-chains), although precise estimates are extremely difficult. Due to the random nature of the mutation process, the virus may not mutate after several human-to-human transmissions, which could result in an overestimation of cold-chain related transmissions. On the other hand, a viral mutation can occur in the first person infected after a spillover event from a cold-chain, such as observed in the Auckland outbreak (Supplementary Figure S2). Those mutated cases were directly linked to a cold-chain but were not considered when estimating the frequency of cold-chain related transmission. Moreover, our method can detect cold-chain related transmissions only if virus-contaminated products have been frozen for a relatively long period of time before the spillover event. If duration is short, no mutation-dormant variants can be detected. Mutation-dormant variants were also sensitive to the search criteria, for example, the mutation-dormant variant that caused the Xinfadi-Beijing outbreak (Supplementary Figure S1) was not identified in the section above because it did not mutate in 93 days (i.e., less than 100 days). Therefore, considering these limitations, our results indicate that the frequency of cold-chain related transmissions may be in the order of 0.1%–10%.

To detect mutation-dormant variants, it is important to precisely estimate the genome-wide evolutionary rate of SARS-CoV-2. Our estimated evolutionary rate is consistent with previous analyses (Duchene et al., 2020; Nie et al., 2020; van Dorp et al., 2020). Different viral genes exhibit different evolutionary rates (Tang et al., 2020; Yu et al., 2022), thus evolutionary rate heterogeneity was considered in this study. Moreover, as the effect of cold-chain related transmission may cause an underestimated evolutionary rate of SARS-CoV-2, the detection of mutation-dormant variants is conservative.

We further verified the identification of mutation-dormant variants. Incorrectly submitted collection dates may result in the misidentification of a mutation-dormant variant because a long non-mutated branch may appear in the tree (Supplementary

Figure S7). However, the CGB quality control procedures remove such samples, so the error rate of submitted collection date should be low (Yu et al., 2022). We also required that each mutation-dormant variant must be associated with at least two non-mutated strains. In other words, by removing one of the non-mutated strains, the variant should remain mutation-dormant. Therefore, it is very unlikely that an incorrectly submitted collection date will affect the results. Ambiguous nucleotides of a genomic sequence could also be mis-imputed in the CGB (Yu et al., 2022). Although this probability is small, each mutation-dormant variant must be determined by at least two sequences without any ambiguous nucleotides on positions that do not define the mutation-dormant variant. As the CGB (Yu et al., 2022) analyzes sequences corresponding to the reference sequence between nucleotides 100 and 29 800 of each genome, unseen mutations can occur on two sides of the genome. To explore this issue, we examined the corresponding alignments of mutation-dormant variants and found that our conclusions remained unchanged (Supplementary Table S2). Therefore, the above confounding factors were considered in this study to ensure the reliability of the identified mutation-dormant variants.

We also examined alternative hypotheses that may cause mutation-dormant variants. Poor sequencing quality may result in this observation; however, the percentage of ambiguous nucleotides was very small (<0.069%) after CGB quality control (Yu et al., 2022). Furthermore, to identify a single mutation-dormant variant, we required a certain number of genomic sequences with no ambiguous nucleotides, and repeated spillover events needed to be identified by multiple independent institutes, indicating observations across different nations or even continents. Thus, poor sequencing quality is very unlikely to explain our observations. It could be argued that purifying selection may lead to a mutation-dormant variant (Tang et al., 2020; Yu et al., 2022). However, the effect of purifying selection was considered in our analysis and could not explain the fast accumulation of mutations after a spillover event from a cold-chain. Two false COVID-19 cases have been documented due to contamination of an inactivated SARS-CoV-2 vaccine strain (To et al., 2022). However, inactivated SARS-CoV-2 vaccine strains are not infectious and will not cause spillover clades. Thus, contamination does not affect our conclusions. Overall, after considering alternative hypotheses, a spillover event from a cold-chain is a plausible explanation for mutation-dormant variants.

It can be difficult to reveal mutation-dormant variants (Figure 1E, Supplementary Figures S1, S2, S5, and S6) if an evolutionary tree is constructed by traditional methods (Hartigan, 1973; Kozlov et al., 2019; Rannala & Yang, 1996; Saitou & Nei, 1987). It is expected that cold-chain related strains will be clustered together given their extreme similarity. Therefore, it is recommended that the internal nodes of a tip-dated tree should be dated using an effective maximum-likelihood method (Sagulenko et al., 2018; Yu et al., 2022). Our results also demonstrate that it is important to have a panoramic vision of the transmission and evolution of SARS-CoV-2 (Yu et al., 2022) by properly analyzing all high-quality genomic sequences and their associated meta-data.

SUPPLEMENTARY NEWS

Supplementary News 1 - News resources for outbreaks
Dalian outbreak
Posted date: 30 July 2020
Link (in Chinese):
https://baijiahao.baidu.com/s?id=1673644074775655832&wfr=spider&for=pc

Supplementary News 2 - News resources for outbreaks
Confirmed case in Shenyang
Posted date: 16 May 2021
Link (in Chinese):
https://baijiahao.baidu.com/s?id=1699875627551244425&wfr=spider&for=pc

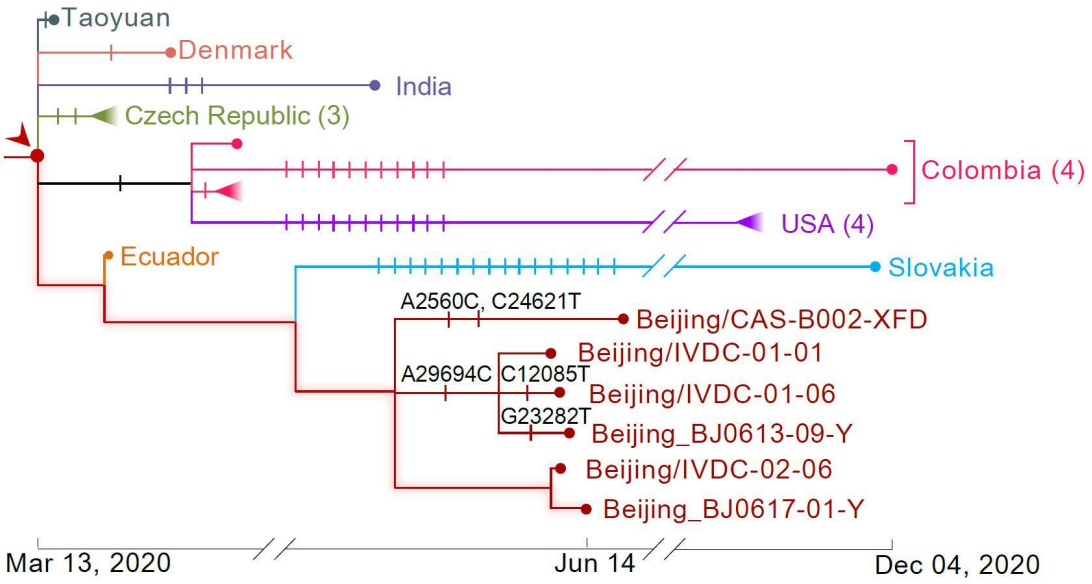
Two confirmed cases in An'hui
Posted date: 16 May 2021
Link (in Chinese):
https://baijiahao.baidu.com/s?id=1699919048786611091&wfr=spider&for=pc

Supplementary News 3 - News resources for outbreaks
Covid 19 Coronavirus: Finance Now and Americold cool store at centre of outbreak
By: Elizabeth Binning, Chief of staff, NZ Herald
Posted date: 12 August 2020
Link (in English):
https://www.nzherald.co.nz/nz/covid-19-coronavirus-finance-now-and-americold-cool-store-at-centre-of-outbreak/XKXFJOPMT3ZW7FJW4BAUX5SUPU/

Coronavirus: First outbreak in New Zealand after 102 days sees Auckland put into lockdown
By: Greg Heffer
Posted date: 12 August 2020
Link (in English):
https://news.sky.com/story/coronavirus-breaks-out-again-in-new-zealand-after-102-days-12047011

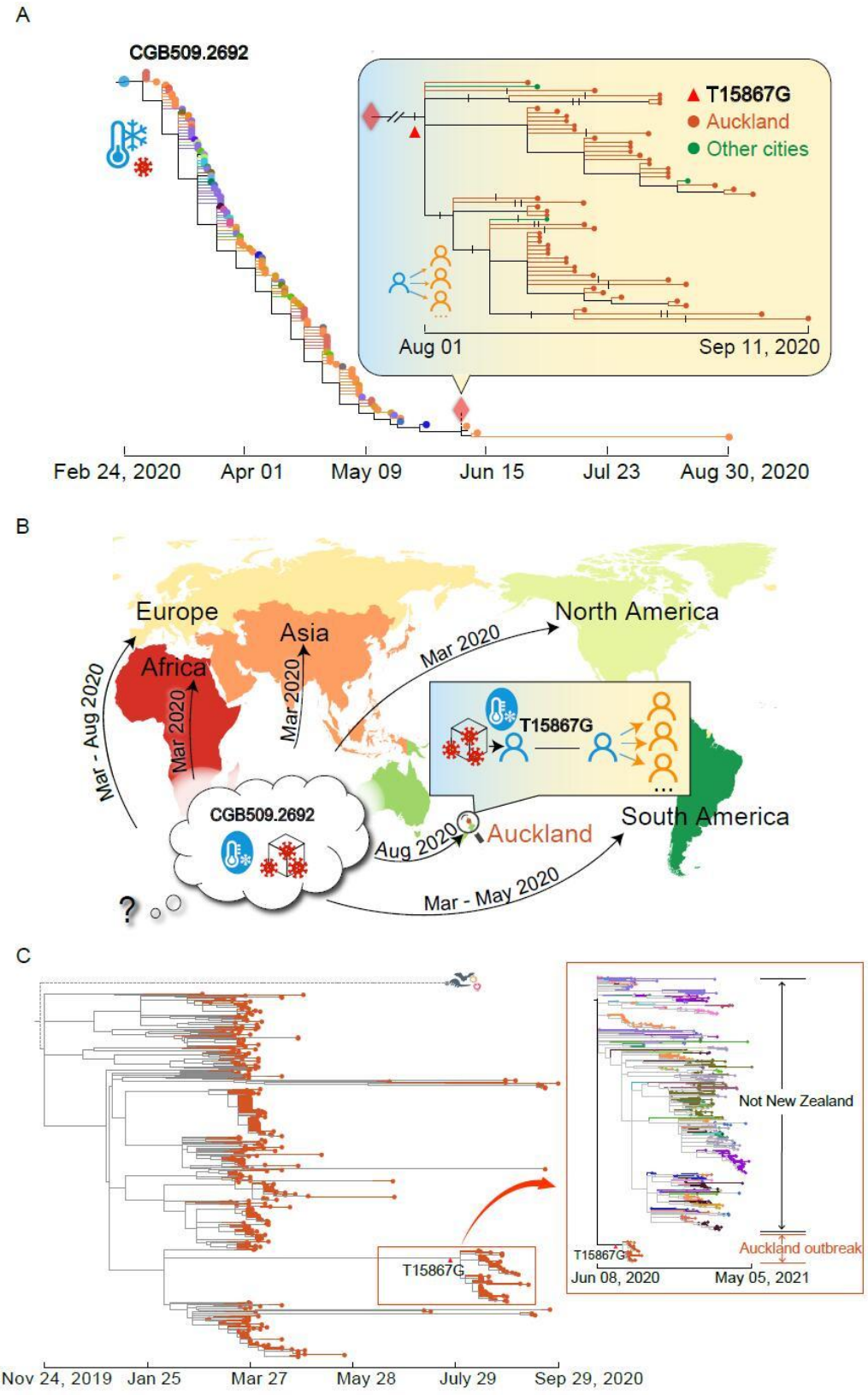
SUPPLEMENTARY MOVIE

The supplementary movie (30 s) shows how to use the CGB to visualize a tree of strains non-mutated between January 2019 and May 2021 (Figure 1E).



Supplementary Figure S1 Cold-chain related outbreaks in Xinfadi, Beijing, China

View of lineages associated with Beijing outbreak. Ancestral viral strain found in early March 2020 is marked with a dark-red solid circle and an arrowhead. This strain is identical to two strains (Beijing/IVDC-02-06 and Beijing/BJ0617-01-Y) collected from two Xinfadi cases on 11 and 14 June 2020, respectively. Each branch notch represents a mutation. Branches with no mutations are highlighted. CGB ID of subtree root is CGB4268.5142 and number of strains is 22.

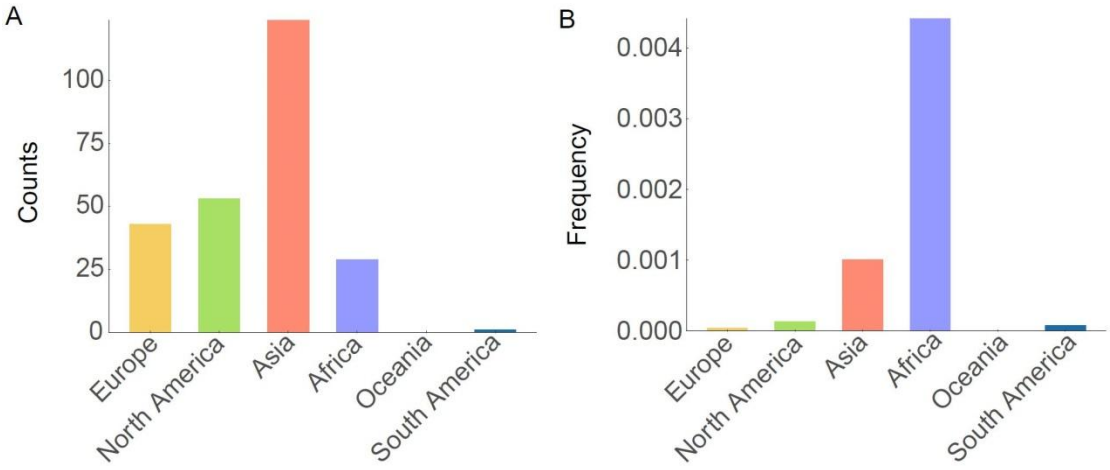


Supplementary Figure S2 Cold-chain related outbreak in Auckland, New Zealand

A: The mutation-dormant variant and Auckland spillover clade shown in sub-panel. Genomic sequences of 120 strains in evolutionary tree are identical to ancestral variant (CGB509.2692), indicated as a blue solid circle at the root of the subtree. Variant carries 11 mutations (C241T, C3037T, C4002T, G10097A, C13536T, C14408T, A23403G, C23731T, G28881A, G28882A, and G28883C). Spillover event from cold-chain is indicated by a thermometer icon. Auckland clade is derived, as indicated by a rhombus. Auckland spillover clade (*n*=56) is shown in sub-panel. Characteristic mutation of Auckland outbreak (T15867G) is marked with a red triangle in the clade. Each branch notch represents a mutation. Illness transmission through social contacts in Auckland outbreak is indicated by four linked people icons in sub-panel.

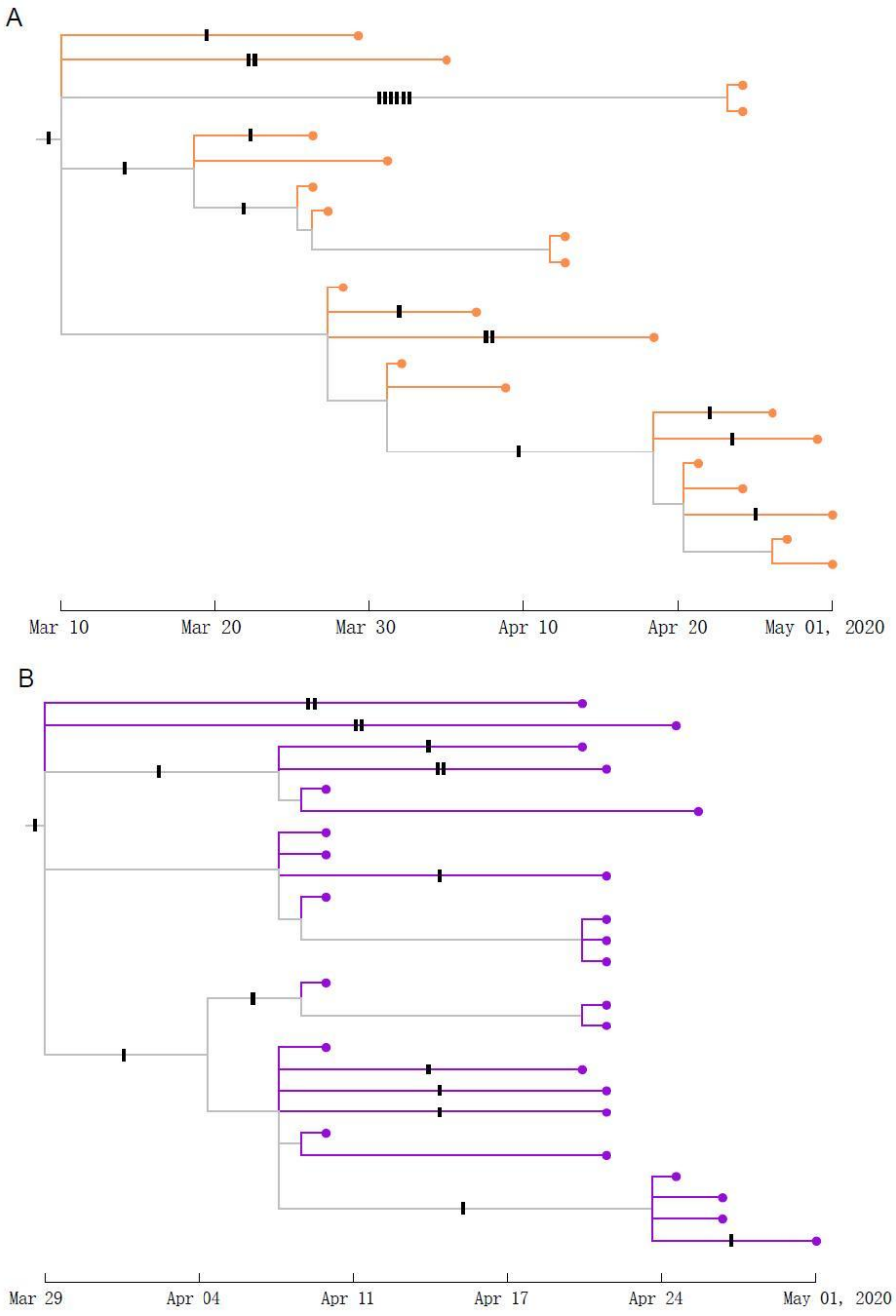
B: Schematic of Auckland outbreak. Origin of cold-chain related infection source remains unknown. Ancestral variant (CGB509.2692), carrying 11 mutations, spread to multiple continents during March to August 2020. Contaminated goods infected the patient zero (in blue) in Auckland in August 2020. Virus in patient mutated once (T15867G) before being transmitted to other human hosts.

C: Samples were first filtered by New Zealand and then only samples before 30 September 2020 were retained. CGB phylogenetic analysis revealed that highlighted Auckland clade (CGB55430.55433) had no epidemiological connection with other New Zealand strains because of its deep split lineage. Subpanel shows enlarged view of strains found in Auckland outbreak together with closely related sibling lineages found in other countries.



Supplementary Figure S3 Geographic distribution of strains with same state as the root (i.e., reference genome sequence NC_045512) (Wu et al., 2020)

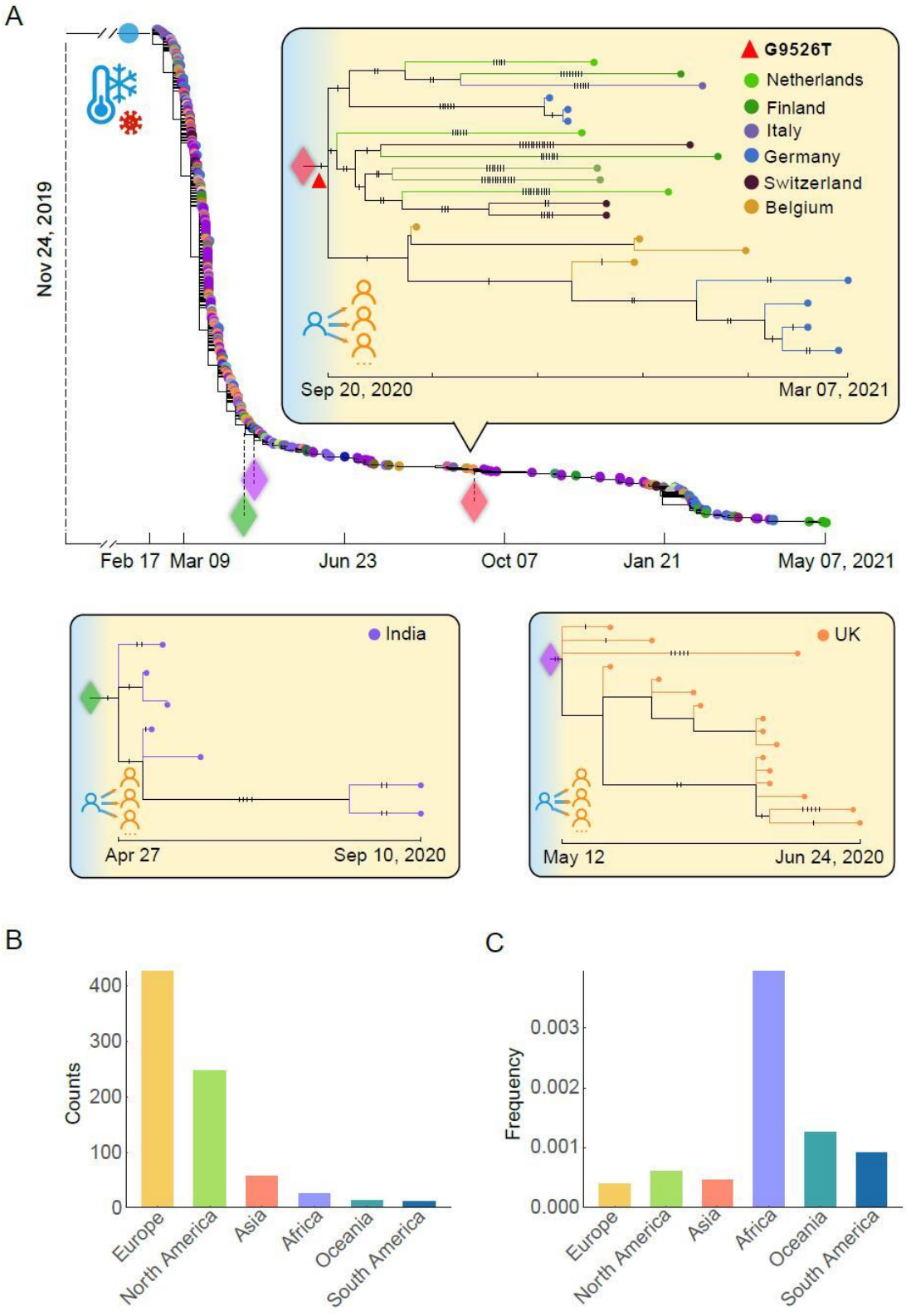
A: Presented as absolute counts.
B: Presented as relative frequencies.



Supplementary Figure S4 Local putative cold-chain related outbreaks

Inferred ancestral strains have same state as the root (i.e., reference genome sequence NC_045512) (Wu et al., 2020).

A: Case of spillover event from cold-chain in the United Kingdom (CGB9057.9058).
B: Case of spillover event from cold-chain in the USA (CGB4735.4736).



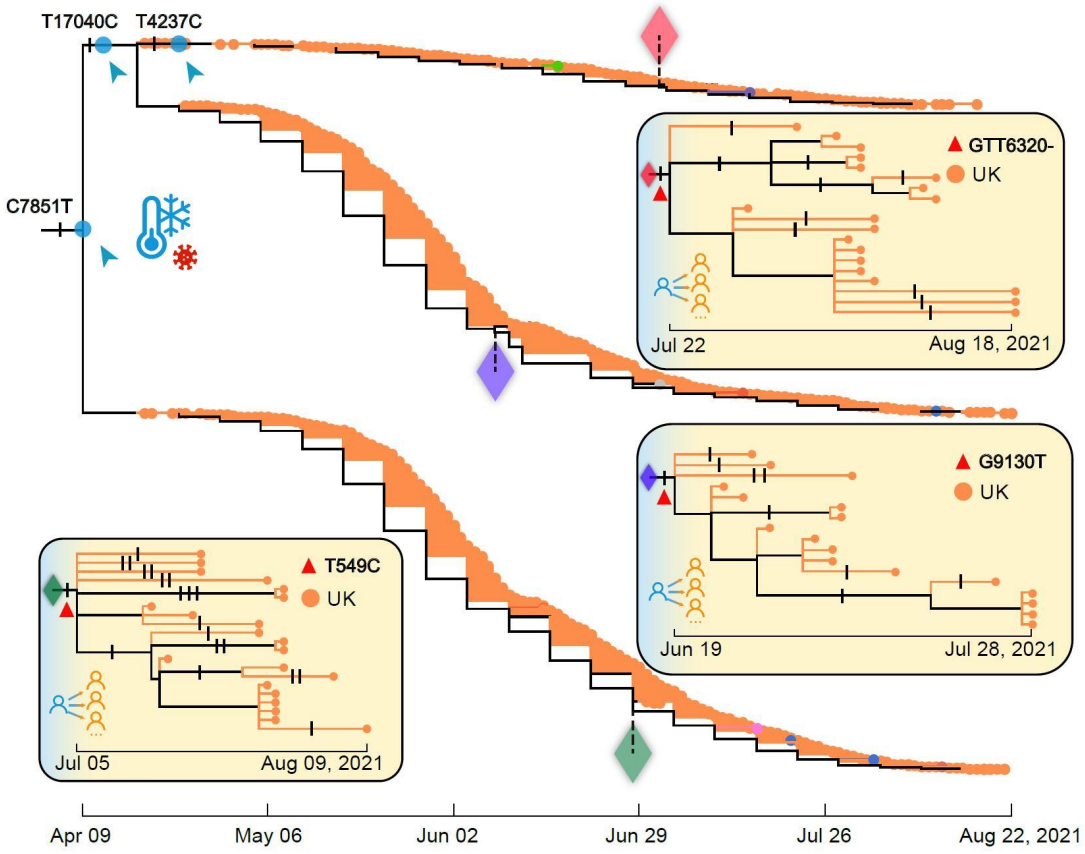
Supplementary Figure S5 Mutation-dormant D614G variant and geographic distribution

A: Tree visualization of 783 strains non-mutated between 2 January 2020 and 7 May

2021. These strains are descendants of a variant with four mutations (C3037T, A23403G, C14408T, and C241T) that emerged before 24 December 2019. Non-synonymous mutation A23403G is known as S:D614G. CGB ID of the tree root is CGB35.120. Three spillover clades are presented in sub-panels. CGB IDs of colored rhombi are CGB236904.257621 (red), CGB34228.34229 (green) and CGB21112.21124 (purple). Blue human icon represents human infected due to spillover event from a cold chain, and orange human icon represents human infected by virus from another human. Light-yellow background represents human-to-human transmission.

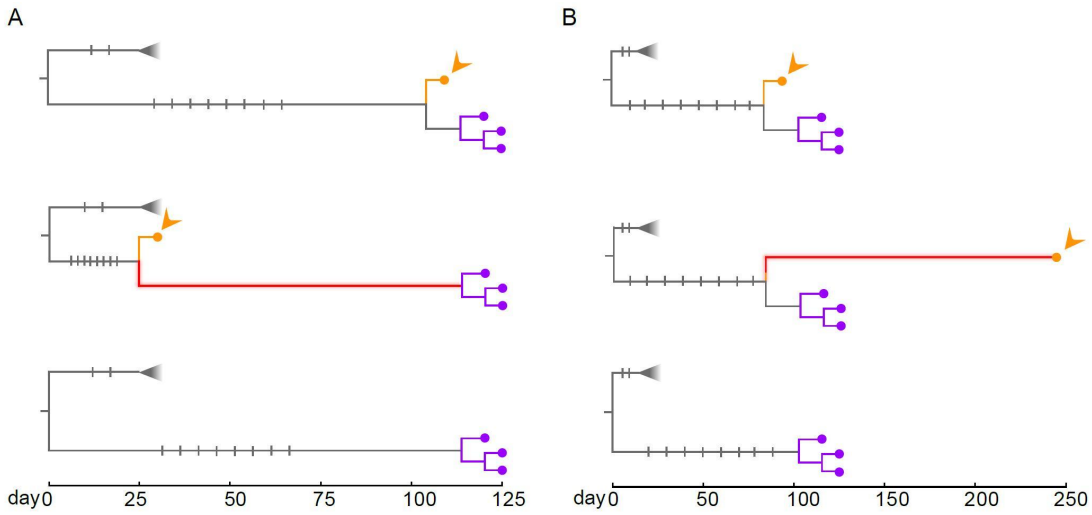
B: Geographic distribution of D614G mutation-dormant variant. Presented as absolute counts.

C: Geographic distribution of D614G mutation-dormant variant. Presented as relative frequencies.



Supplementary Figure S6 Three mutation-dormant Delta VOCs and three spillover UK clades presented in sub-panels

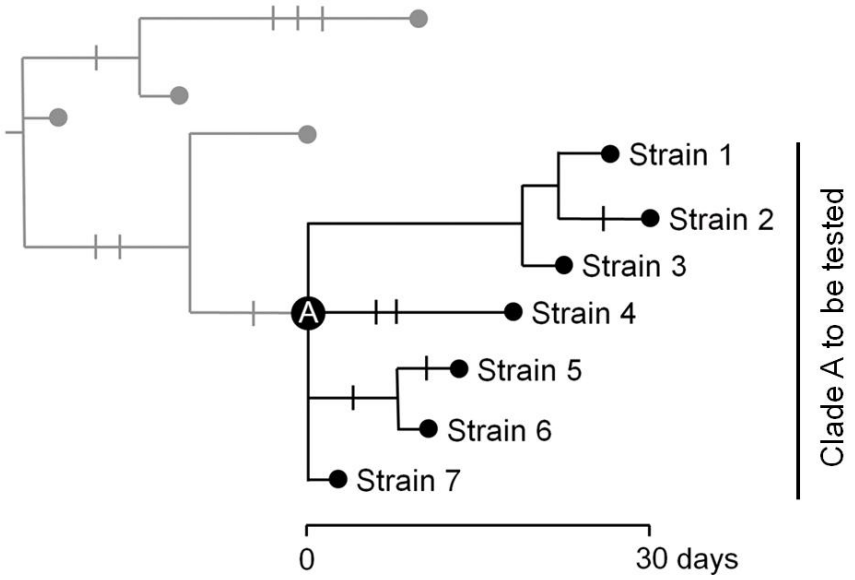
Three mutation-dormant variants are indicated as blue solid circles near root of the subtree (CGB531065.580055). Three characteristic mutations are marked to distinguish three variants. Clade derived from a cold-chain related spillover event is indicated by a thermometer icon. Three spillover clades are presented in sub-panels. Their characteristic mutations are labeled, and CGB IDs of colored rhombi are CGB1352928.1377873, CGB1224042.1265384, and CGB1488858.1503350, respectively. Blue human icon represents human infected due to spillover event from a cold chain, and orange human icon represents human infected by virus from another human. Light-yellow background represents human-to-human transmission.



Supplementary Figure S7 Schematic showing effect of mis-submitted collection date on identification of mutation-dormant variant

A: Misidentified mutation-dormant variant due to incorrectly submitted collection date (earlier than true collection date). Upper panel shows true condition with correct collection date, marked by arrowhead. Middle panel shows incorrectly submitted collection date of strain (marked by arrowhead). Incorrectly submitted collection date results in misidentification of long non-mutated path (marked in red). Lower panel shows disappear of misidentified long non-mutated path when strain with incorrectly submitted collection date is removed. Each branch notch represents a mutation.

B: Misidentified mutation-dormant variant due to incorrectly submitted collection date (later than true collection date).



Supplementary Figure S8 Schematic showing how the test is performed for clade A

Each branch notch represents a mutation. Node with black circled “A” represents clade to be tested.

Supplementary Table S1 Evolutionary rate of SARS-CoV-2.

Region	Length (sites)	Evolutionary rate (per site per year)
Genome-wide	29,701	9.69×10^{-4}
ORF1a	13,218	5.60×10^{-4}
ORF1b	8,088	3.58×10^{-4}
S	3,822	2.13×10^{-3}
ORF3a	828	3.10×10^{-4}
E	228	1.62×10^{-5}
M	669	4.80×10^{-4}
ORF6	186	9.01×10^{-5}
ORF7a	366	4.56×10^{-3}
ORF7b	132	4.22×10^{-3}
ORF8	366	4.52×10^{-3}
N	1,260	2.78×10^{-3}
ORF10	117	9.69×10^{-4}
noncoding	441	6.07×10^{-3}

Supplementary Table S2 List of non-mutated strains shown in Figure 1E and Supplementary Figures S2A, S5A.

Supplementary Table S3 The list of 362 mutation-dormant variants.

Supplementary Tables S2 and S3 are listed as separate excel files due to their large size.

Supplementary Table S4 Mutation-dormant variants in Delta variant of concern (VOC).

Index	CGB ID	Total length (days)	Num. of identical strains	Date of start	Date of end	Characteristic mutations	Other mutations[†]
1	CGB531065.591827	160	34	2021/3/5	2021/8/12	G21987A	Mut Set 01,02
2	CGB591827.645203	158	98	2021/3/8	2021/8/13	G21987A, C21846T	Mut Set 01,02
3	CGB736525.830096	144	55	2021/3/21	2021/8/12	A29700G, T14014G, C6040T, C7926T	Mut Set 01,02
4	CGB744625.824957	144	474	2021/3/30	2021/8/21	G21987A, C21846T, A25003G, C6408T	Mut Set 01,02
5	CGB736525.839242	141	283	2021/3/24	2021/8/12	A29700G, T14014G, C6040T, C7926T, G21987A	Mut Set 01,02
6	CGB843452.957320	134	126	2021/3/30	2021/8/11	G21987A, G26107C, A27507C, C21897T	Mut Set 01,02
7	CGB692105.692163	134	4,582	2021/4/9	2021/8/21	G21987A, C21846T, C7851T	Mut Set 01,02
8	CGB692105.758869	132	3,984	2021/4/12	2021/8/22	G21987A, C21846T, C7851T, T17040C	Mut Set 01,02
9	CGB759162.841282	131	89	2021/4/10	2021/8/19	G21987A, C27527T, G1048T, C27526T, C4455T	Mut Set 01,02
10	CGB763421.763514	130	929	2021/4/14	2021/8/22	G21987A, G2336A, G17331T	Mut Set 01,02
11	CGB685958.801178	118	577	2021/4/15	2021/8/11	G21987A, G24928T, C29253T	Mut Set 01,02
12	CGB801082.864797	116	779	2021/4/23	2021/8/17	G21987A, C21846T, C7851T, T17040C, T4237C	Mut Set 01,02
13	CGB844959.875198	115	112	2021/4/25	2021/8/18	G21987A, C21846T, C7851T, G22051C	Mut Set 01,02
14	CGB580144.642966	112	263	2021/4/1	2021/7/22	C9891T, C5184T, T11418C, C11514T, C22227T, C13019T, A5584G, G21987A, C22792T	Mut Set 01
15	CGB885562.885688	109	239	2021/5/4	2021/8/21	G21987A, G27566T, C27629A, G12718T, G11365T, G2944A, T15882G	Mut Set 01,02

[†]Mut Set 01 includes 23 mutations: C3037T, A23403G, C14408T, C241T, T22917G, G28881T, T27638C, C23604G, C27752T, C22995A, A28461G, T26767C, G29742T, C25469T, G24410A, C21618G, G29402T, C16466T, G15451A, G210T, GATTTC28248-, AGTTCA22029-, and A28271-.

Mut Set 02 includes 11 mutations: G28916T, C27874T, C7124T, G4181T, G9053T, C8986T, C10029T, C19220T, A11332G, A11201G, and C640.

REFERENCES

Amendola A, Canuti M, Bianchi S, Kumar S, Fappani C, Gori M, et al. 2021. Molecular evidence for SARS-CoV-2 in samples collected from patients with morbilliform eruptions since late summer 2019 in Lombardy, Northern Italy.

Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evolution*, **6**(2): veaa061.

Dunn OJ. 1961. Multiple comparisons among means. *Journal of the American Statistical Association*, **56**(293): 52–64.

Gong YN, Tsao KC, Hsiao MJ, Huang CG, Huang PN, Huang PW, et al. 2020. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. *Emerging Microbes & Infections*, **9**(1): 1457–1466.

Hartigan JA. 1973. Minimum mutation fits to a given tree. *Biometrics*, **29**(1): 53–65.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, **35**(21): 4453–4455.

Nie Q, Li XG, Chen W, Liu DH, Chen YY, Li HT, et al. 2020. Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Research*, **287**: 198098.

Pang XH, Ren LL, Wu SS, Ma WT, Yang J, Di L, et al. 2020. Cold-chain food contamination as the possible origin of Covid-19 resurgence in Beijing. *National Science Review*, **7**(12): 1861–1864.

Rannala B, Yang ZH. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution*, **43**(3): 304–311.

Ruan YS, Wen HJ, Hou M, He ZW, Lu XM, Xue YB, et al. 2022. The twin-beginnings of COVID-19 in Asia and Europe – one prevails quickly. *National Science Review*, **9**(4): nwab223.

Sagulenko P, Puller V, Neher RA. 2018. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evolution*, **4**(1): vex042.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**(4): 406–425.

Tang XL, Wu CC, Li X, Song YH, Yao XM, Wu XK, et al. 2020. On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, **7**(6): 1012–1023.

To KKW, Li X, Lung DC, Ip JD, Chan WM, Chu AWH, et al. 2022. False coronavirus disease 2019 cases due to contamination by inactivated virus vaccine. *Clinical Infectious Diseases*, **74**(8): 1485–1488.

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

Wang YT, Dai GY, Gu ZL, Liu GP, Tang K, Pan YH, et al. 2020. Accelerated

evolution of an *Lhx2* enhancer shapes mammalian social hierarchies. *Cell Research*, **30**(5): 408–420.

Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature*, **579**(7798): 265–269.

Yu DL, Dong LL, Yan FQ, Mu HL, Tang BX, Yang X, et al. 2019. eGPS 1.0: comprehensive software for multi-omic and evolutionary analyses. *National Science Review*, **6**(5): 867–869.

Yu DL, Yang X, Tang BX, Pan YH, Yang JN, Duan GY, et al. 2022. Coronavirus GenBrowser for monitoring the transmission and evolution of SARS-CoV-2. *Briefings in Bioinformatics*, **23**(2): bbab583.

Zhang Y, Pan Y, Zhao X, Shi WF, Chen ZX, Zhang S, et al. 2020. Genomic characterization of SARS-CoV-2 identified in a reemerging COVID-19 outbreak in Beijing's Xinfadi market in 2020. *Biosafety and Health*, **2**(4): 202–205.

Zhou B, Thao TTN, Hoffmann D, Taddeo A, Ebert N, Labroussaa F, et al. 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature*, **592**(7852): 122–127.