Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events

Junxian Ou[1,2], Wendong Lan[2,] Xiaowei Wu[2], Tie Zhao[1], Biyan Duan[1], Peipei Yang[1], Yi Ren[1], Lulu Quan[2], Wei Zhao[2], Donald Seto[3], James Chodosh[4], Jianguo Wu[1,5*] & Qiwei Zhang[1,2,5*]

[1]Guangdong Provincial Key Laboratory of Virology, Institute of Medical Microbiology, Jinan University, Guangzhou 510632, China

[2]BSL-3 Laboratory (Guangdong), Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou 510515, China

[3]Bioinformatics and Computational Biology Program, School of Systems Biology, George Mason University, Manassas, VA 20110, USA

[4]Department of Ophthalmology, Howe Laboratory Massachusetts Eye and Ear, Harvard Medical School, Boston, MA 02114, USA

[5]Foshan Institute of Medical Microbiology, Foshan 528315, China

* Correspondence: Q.Z., E-mail: zhangqw@jnu.edu.cn, ORCID: 0000-0002-2770-111X; J.W., E-mail: jwu898@jnu.edu.cn, ORCID: 0000-0002-8326-2895.

Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events

Abstract

The current pandemic of COVID-19 is fueled by more infectious emergent Omicron variants. Ongoing concerns of emergent variants include possible recombinants, as genome recombination is an important evolutionary mechanism for the emergence and re-emergence of human viral pathogens. Although recombination events among SARS-CoV-1 and MERS-CoV were well-documented, it has been difficult to detect the recombination signatures in SARS-CoV-2 variants due to their high degree of sequence similarity. In this study, we identified diverse recombination events between two Omicron major subvariants (BA.1 and BA.2) and other variants of concern (VOCs) and variants of interest (VOIs), suggesting that co-infection and subsequent genome recombination play important roles in the ongoing evolution of SARS-CoV-2. Through scanning high-quality completed Omicron spike gene sequences, eighteen core mutations of BA.1 variants (frequency >99%) were identified (eight in NTD, five near the S1/S2 cleavage site, and five in S2). BA.2 variants share three additional amino acid deletions with the Alpha variants. BA.1 subvariants share nine common amino acid mutations (three more than BA.2) in the spike protein with most VOCs, suggesting a possible recombination origin of Omicron from these VOCs. There are three more Alpha-related mutations (del69-70, del144) in BA.1 than BA.2, and therefore BA.1 may be phylogenetically closer to the Alpha variant. Revertant mutations are found in some dominant mutations (frequency >95%) in the BA.1 subvariant. Most notably, multiple additional amino acid mutations in the Delta spike protein were also identified in the recently emerged Omicron isolates, which implied possible recombination events occurred between the Omicron and Delta variants during the on-going pandemic. Monitoring the evolving SARS-CoV-2 genomes especially for recombination is critically important for recognition of abrupt changes to viral attributes including its epitopes which may call for vaccine modifications.

Introduction

The current COVID-19 pandemic is fueled by a more infectious emergent Omicron variant (B.1.1.529), which was first reported in South Africa and quickly spread worldwide[1]. A multitude of mutations (more than 30) in the spike gene of Omicron variant were detected, which when compared to the Alpha and Delta variants (typically less than 15)[2], raised concerns of enhanced infectivity and immune escape potential[3,4]. Omicron variants is divided into three lineages (BA.1, BA.2, and BA.3) and was classified as the fifth variant of concern (VOC) by the World Health Organization on November 26, 2021. It has been circulating in more than 170 countries/territories.

Mutations in the SARS-CoV-2 spike gene have altered protein binding efficiency and immunogenicity, and resulted in more invasive and adaptive variants[4-9]. Previous research on Alpha （B.1.1.7) and Delta (B.1.617.2 and AY.x) variants with spike gene mutations confirmed these effects on enhancing virus transmission[4-8]. Meanwhile, as a critical antigenic recognition site, the spike protein is also the principal vaccine design target, and these observed mutations have focused attention on this modified antigen and its putative immune escape potential and antibody resistance[3,10-12].

Ongoing concerns of emergent variants includes possible recombinants resulting from different variants replicating simultaneously in a host. Such variants, e.g., “Demicron” or “Deltacron” are controversial that if they are real recombinants or a possible sequencing error[13].

Genome recombination is an important evolutionary mechanism for the emergence and re-emergence of human pathogens and a major source of viral evolution, for example, the well-studied “model organism” adenovirus [14-20], and also in coronaviruses[21-23]. Recombination accelerates virus evolution through gene(s) and “function” transference and accumulation of selective and advantageous mutations, resulting in phenotype changes that include changes in pathogenicity profiles, host species virulence, zoonotic and anthroponotic transmission, and host adaptation [14-21,24,25].

Although recombination events among SARS-CoV-1 and MERS-CoV were well-documented[21-23], it has been difficult to detect the recombination signatures in SARS-CoV-2 variants due to the high degree of sequence similarity amongst SARS-CoV-2 isolates and the incomplete coverage of coronaviruses from other hosts, including pangolin[26,27].

Previous research distinguished active recombination events among the SARS-CoV-2 nucleoprotein and ORF1ab genes by using a phylogenetic network strategy based on single nucleotide substitution or SARS-CoV-2 lineage designation[27,28]. More than thirty amino acid mutations have been identified within Omicron spike protein, some of which are shared with other variants[1]. In this study, we demonstrate that the emerging and circulating Omicron subvariants originate in part through recombination with other variants. We first investigated the spike diversity of the Omicron variants along with the shared spike mutations between Omicron and other variants of concern (VOCs) and variants of interest (VOIs). The Omicron spike amino acid sequences archived during the early transmission phase, and released in the GISAID database (submitted before January 15th, 2022) were accessed, include 52,563 high quality Omicron spike sequences (representing 49,609 BA.1 and 2,954 BA.2 sequences). In this study, these were analyzed with Pymol 2.0, TBTools, BioEdit, BioAider, and jvenn[29-35]. The whole genome phylogenetic trees were constructed and annotated using NextClade[36].

Tracking the common mutations among Omicron (BA.1 and BA.2) and variants of concern (VOCs)

Circulating Omicron variant consists of two main subvariants, BA.1 and BA.2. BA.1 subvariant was more frequently detected than BA.2 during the early transmission phase. However, BA.2 is replacing BA.1 as the dominant epidemic subvariant in more and more countries over time[37].

Through scanning 52,563 high-quality completed Omicron spike gene sequences, most Omicron spike mutations appear stable (frequency >99%). Eighteen core mutations (frequency >99%) of BA.1 subvariant exist in NTD (A67V, del69-70, T95I, G142D, del143-145), SD (underpinning subdomain) near the S1/S2 cleavage site (T547K, D614G, H655Y, N679K, P681H), and S2 (D796Y, N856K, Q954H, N969K, L981F)(Table1).

BA.1 subvariant shares nine common amino acid mutations (del69-70, delY144, K417N, T478K, N501Y, D614G, H655Y, and P681H) in the spike protein with most VOCs, suggesting a possible origin of Omicron from these VOCs. Among these shared mutations, six common ones were found in Alpha variant (del69-70, delY144, N501Y, D614G, and P681H), to which the mutations of del69-70, delY144 and P681H are exclusive; three mutations were found in Beta variant (K417N, N501Y, and D614G), to which the mutation K417N is exclusive; three mutations found in Gamma (N501Y, D614G, and H655Y), to which the mutation H655Y is exclusive; two mutations found in Delta (T478K and D614G), to which the mutation T478K is exclusive (Fig.1A and Table 1). The seven Omicron mutations exclusive to other four VOCs suggested a possible recombination origin of Omicron.

Compared to BA.1 subvariant, BA.2 shares only six amino acid mutations (K417N, T478K, N501Y, D614G, H655Y, P681H) in the spike protein with most VOCs. Among these shared mutations, three mutations were found in Alpha variants (N501Y, D614G, P681H); there were no del69-70 and delY144 mutations. The other three mutations in Beta, three mutations in Gamma, and two mutations in Delta were identical in the BA.2 and BA.1 genomes (Fig.1 B and Table 1).

BA.1 and BA.2 subvariants share twenty-one spike amino acid mutations: One in the N-terminal domain (NTD) (G142D), twelve in the receptor binding domain (RBD) (G339D, S373P, S375F,

K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H), four in SD (D614G, H655Y, N679K, P681H), and four in S2 (N764K, D796Y, Q954H, N969K) (Fig.1 C and Table 1). In contrast to BA.2 subvariant, BA.1 share three additional amino acid deletions (del69-70, delY144) with the Alpha variants, suggesting a closer relationship between the BA.1 and Alpha variants (Fig.1A and 1B, and Table 1). As a whole, Omicron subvariants have a high number of amino acid mutations in the spike gene (40 in BA.1, and 31 in BA.2), of which some were found in other VOCs: Alpha (10x), Beta (11x), Gamma (12x), and Delta (9x). These mutations mainly occur in NTD and RBD (Fig.1D and Table 1).

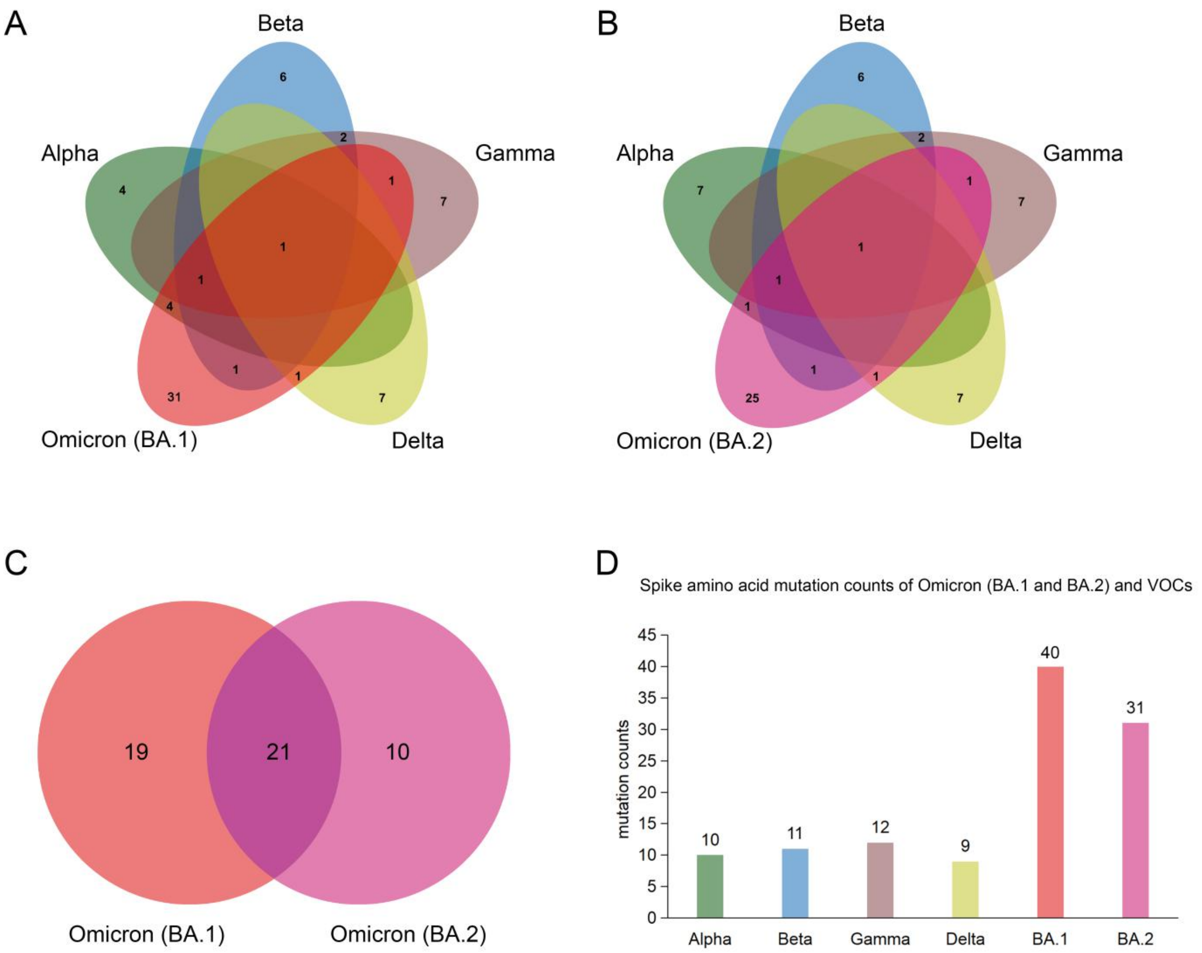


Figure 1. Spike protein amino acid mutations of the Omicron subvariants (BA.1 and BA.2) compared with mutations from the other four variants of concern (VOCs). (A) Venn diagram noting mutations of Omicron (BA.1) and those of VOCs. **(B)** Venn diagram of Omicron (BA.2) mutations compared to ones of VOCs. **(C)** Venn diagram of mutations between Omicron (BA.1) and Omicron (BA.2). **(D)** Spike protein amino acid mutation counts of Omicron (BA.1 and BA.2) subvariants compared with mutations of VOCs.

Table 1. Comparison of Spike protein amino acid mutations between the Omicron subvariants and other VOCs and VOIs. 52,563 high quality Omicron spike gene sequences (49,609 BA.1 sequences, and 2,954 BA.2 sequences) released before January 15, 2022 were analyzed. The mutations that have appeared in more than 800 sequences were used in this analysis. VOCs are variants of concern; VOIs are variants of interest.

Spike Region		Position	Mutation (BA.1)	Frequency (BA.1)	Percentage	Mutation (BA.2)	Frequency (BA.2)	Percentage	Mutation in VOCs/VOIs	New Mutation
		19				T19I	2953	100.00%		
		24				deletion	2513	85.10%		
		25				deletion	2513	85.10%		
		26				deletion	2513	85.10%		
		27				A27S	2513	85.10%		
		67	A67V	49475	99.73%					
		69	deletion	49385	99.55%				Alpha(del69)	
		70	deletion	49382	99.54%				Alpha(del70)	
		95	T95I	49513	99.81%				Mu(T95I)	
	NTD	142	G142D	49424	99.63%	G142D	2934	99.36%		
		143	deletion	49439	99.66%					
		144	deletion	49439	99.66%				Alpha(del144), Mu(Y144S)	
		145	deletion	49440	99.66%				Mu(Y145N)	
		211	deletion	47840	96.43%					
		212	N212I	47839	96.43%					
		213				V213G	2950	99.90%		
		214	insertE	44368	89.44%					
		214	insertP	44353	89.41%					
		214	insertE	44365	89.43%					
		339	G339D	48729	98.23%	G339D	2953	100.00%		
		346	R346K	16819	33.90%				Mu(R346K)	Yes
		371	S371L	48297	97.36%	S371F	2949	99.86%		
		373	S373P	48322	97.41%	S373P	2952	99.97%		
S1		375	S375F	48316	97.39%	S375F	2951	99.93%		
		376				T376A	2949	99.86%		
		405				D405N	2949	99.86%		
		408				R408S	2946	99.76%		
		417	K417N	44711	90.13%	K417N	2952	99.97%	Beta(K417N), Gamma(K417T)	
		440	N440K	46470	93.67%	N440K	2926	99.09%		
	RBD	446	G446S	46892	94.52%					
		452	L452R	899	1.81%				Delta(L452R)	Yes
		477	S477N	48185	97.13%	S477N	2952	99.97%		
		478	T478K	48320	97.40%	T478K	2952	99.97%	Delta(T478K)	
		484	E484A	48024	96.81%	E484A	2952	99.97%	Beta/Gamma/Mu (E484K)	
		493	Q493R	47999	96.75%	Q493R	2953	100.00%		
		496	G496S	47965	96.69%					
		498	Q498R	47917	96.59%	Q498R	2953	100.00%		
		501	N501Y	47933	96.62%	N501Y	2953	100.00%	Alpha/Beta/Gamma/Mu (N501Y)	
		505	Y505H	47888	96.53%	Y505H	2952	99.97%		
		547	T547K	49496	99.77%					
		614	D614G	49568	99.92%	D614G	2953	100.00%	Alpha/Beta/Gamma/Delta (N501Y)	
	SD	655	H655Y	49509	99.80%	H655Y	2953	100.00%	Gamma(H655Y)	
		679	N679K	49523	99.83%	N679K	2953	100.00%		
		681	P681H	49515	99.81%	P681H	2953	100.00%	Alpha/Mu(P681H), Delta(P681R)	
		701	A701V	2729	5.50%				Beta(A701V)	Yes
		764	N764K	49046	98.87%	N764K	2953	100.00%		
	FP	796	D796Y	49338	99.45%	D796Y	2952	99.97%		
S2		856	N856K	49488	99.76%					
		954	Q954H	49559	99.90%	Q954H	2953	100.00%		
	HR1	969	N969K	49537	99.85%	N969K	2953	100.00%		
		981	L981F	49373	99.52%					

Tracking novel mutations and mutations with decreased frequency in the spike gene of Omicron BA.1 and BA.2

We investigated additional mutations among recently emerged BA.1 isolates and identified eight novel mutations in Omicron variant which were also found in other VOCs and VOIs. For example, mutations R346K (33.90% of 49,609 BA.1 sequences) was found in Mu variants; A701V (5.50%) was found in Beta variants; L5F (0.37%) was found in Iota variants; and T76I (0.10%) was found in Lambda variants. Most notably, multiple representative amino acid mutations in the Delta spike protein were also identified in the recently emerged Omicron subvariants (del156-167, R158G, L452R, and P681R, at percentages of 0.14%, 0.14%, 1.81%, and 0.12%, respectively. This implied possible recombination events between the Omicron and Delta strains during the pandemic. The other newly noted mutations (L141F, F643L, I1081V, S1147L, and P1162S) may have originated independently (Table 2).

Several novel mutations were reported to be related to spike protein function, resulting in an enhancement of virus infectivity or in viral immune escape. Mutations that occurred in RBD, e.g., R346K, could result in a relatively weakened neutralizing antibody effect[3]. A L452R mutation may provide evasion from cellular immunity and increased infectivity[5,6]. The P681R as well as F643L and A701V mutations, near the S1/S2 cleavage site, may be associated with enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta variants[8]. Additionally, mutations T76I, L141F, G142Y, 156-167deletion, and R158G, located in the NTD region, were noted to affect antibody binding efficiencies and contribute to immune escape[38]. These mutations sites are mapped and shown in Fig. 2.

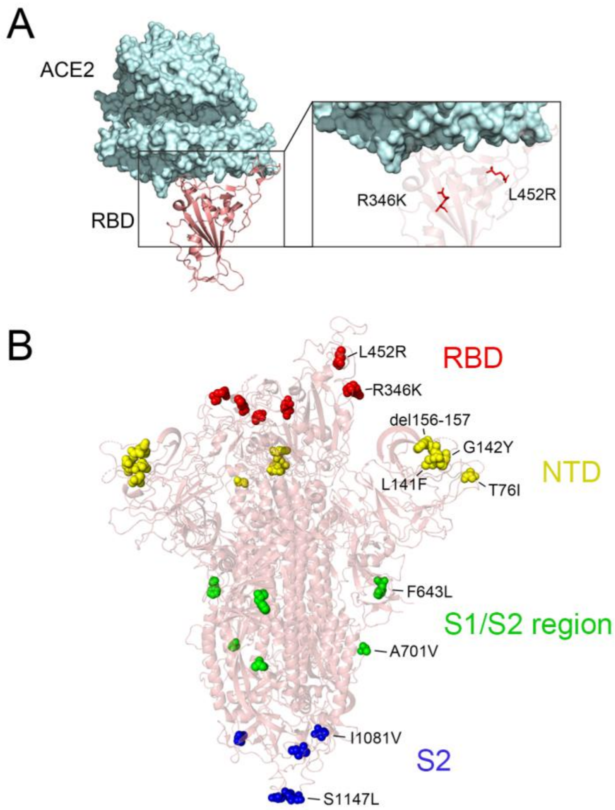


Figure 2. Structure of the Spike protein with amino acid mutations detected in Omicron BA.1 subvariant. (A) Structure of human ACE2 receptor complexed with SARS-CoV-2 Omicron RBD, mapped with the recent mutations. (**B)** Structure of SARS-CoV-2 Omicron spike protein mapped with the novel mutations. Mutated residues in each domain of the spike protein are annotated in color (red: RBD; yellow: NTD; green: S1/S2; blue: S2) using with Pymol 2.0 software through SARS-CoV-2 Omicron model PDB:7WBL and 7QO7 (Han, P. et al. Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. Cell, 2022, doi:10.1016/j.cell.2022.01.001).

Apparent revertant mutations are found in some dominant mutations (frequency >95%) in the BA.1 subvariant during the pandemic. Examples are the mutations in NTD (del211 and N212I) and RBD (G339D, S371L, S373P, S375F, K417N, 440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H). The frequency of insertions of the amino acids EPE at site 214 in BA.1 decreased during the pandemic from more than 95% on December 1st, 2021 to 89% on January 15th 2022. However, BA.2 spike protein remained constant (frequency >99%), with the exception of the three amino acid deletion (LPP) found at amino acids 24-26, which decreased from more than 95% frequency on December 1st, 2021 to 85% on January 15th 2022 (Table 1). This may possibly be due to selection pressure on the circulating Omicron strains.

Table 2. Novel mutations identified in the spike protein of the recently emerged Omicron subvariants (Released before January 15, 2022; frequency >50 sequences).

Spike Region		Position	New Mutation	Frequency	Percentage	Mutation in VOCs&VOIs	Early Event Occurrence Time
S1		5	L5F	184	0.37%	Iota（L5F)	2021.11.19
	NTD	77	T76I	51	0.10%	Lambda（T76I）	2021.11.26
		141	L141F	56	0.11%		2021.11.19
		142	G142Y	51	0.10%		2021.12.16
		156	deletion	68	0.14%	Delta(del156)	2021.12.13
		157	deletion	71	0.14%	Delta(del157)	2021.12.13
		158	R158G	69	0.14%	Delta(E158G)	2021.12.15
	RBD	346	R346K	16819	33.90%	Mu(R346K)	2021.11.4
		452	L452R	899	1.81%	Delta(L452R)	2021.11.11
	SD	643	F643L	138	0.28%		2021.11.29
		681	P681R	62	0.12%	Delta(P681R)	2021.11.23
S2	FP	701	A701V	2729	5.50%	Beta（A701V)	2021.11.10
	HR2	1081	I1081V	351	0.71%		2021.11.18
		1147	S1147L	120	0.24%		2021.12.13
		1162	P1162S	60	0.12%		2021.12.10

The phylogenetic network of Omicron spike genes shows novel recombination events during the pandemic

Further investigation of the Omicron subvariants examined spike protein haplotypes . These were identified, screened, and calculated with the R package tidyfst, aligned with MAFFT[39,40]. These haplotypes were analyzed with DnaSP 6.0[41], and a subsequent phylogenetic network was constructed using PopART (http://popart.otago.ac.nz/index.shtml), with mutations annotated with Nextclade (https://clades.nextstrain.org).

The spike gene of Omicron subvariants consists of 49 representative haplotypes (each occurring in more than 50 sequences). BA.1, BA.2, R346K, L452R, A701V, and a revertant type were identified in the phylogenetic network analysis (Fig. 3A). A large number of BA.1 spike mutations delineated haplotype 2, R346K, L452R, and A701V clusters and formed distinct subgroups (detailed mutations defining each haplotype are listed in Supplementary Table 1).

Multiple nucleotide mutations were detected in the haplotypes compared with BA.1, e.g., haplotype 19 and the revertant subgroup (Hap 30, 32, 43, *etc.*) and BA.2. The L452R subgroup consists of different haplotypes with multiple nucleotide substitutions, indicating a possibly separate origin of L452R haplotypes or prior recombination events. The revertant subgroup consisted of Omicron haplotypes in which several BA.1 representative mutations were lost and appeared to have reverted to the bases of the Wu-hu-1 strain. Multiple nucleotide differences in other haplotypes occurred, likely as multiple independent mutation events, or perhaps as recombination events among highly similar sequences. Haplotype 25 in L452R subgroup, with multiple nucleotide differences compared with BA.1, could have resulted through recombination between Omicron and Delta variants, gaining the mutation L452R from Delta and losing multiple mutations from Omicron (Fig. 3A). Some of these "Demicron" or "Deltacron" haplotypes are being tracked by the UK Health Security Agency (https://www.gov.uk/government/publications/sars-cov-2-variants-of-public-health-interest/sars-cov-2-variants-of-public-health-interest-25-february-2022) and underway to confirm by Santé publique

France (https://t.co/tVAKmHRYSy). Bootscan analysis of Omicron spike sequences also indicated that the reversion haplotypes (Hap_18, Hap_39, Hap_44) were more similar to Delta variants when compared to typical Omicron haplotypes (Fig. 3A and 3B).

Furthermore, single nucleotide differences could also originate from recombination events among highly similar strains. Loops detected in phylogenetic networks also indicate possible recombination events among highly similar Omicron variants or subvariants (Fig. 3A and 3C). Multiple newly detected or recent mutations in the Omicron spike gene make it possible to trace a putative mutation origin from representative mutations in VOIs or VOCs, especially the Delta variant, which suggests possible recombination events between Omicron and Delta variants (Table 2).

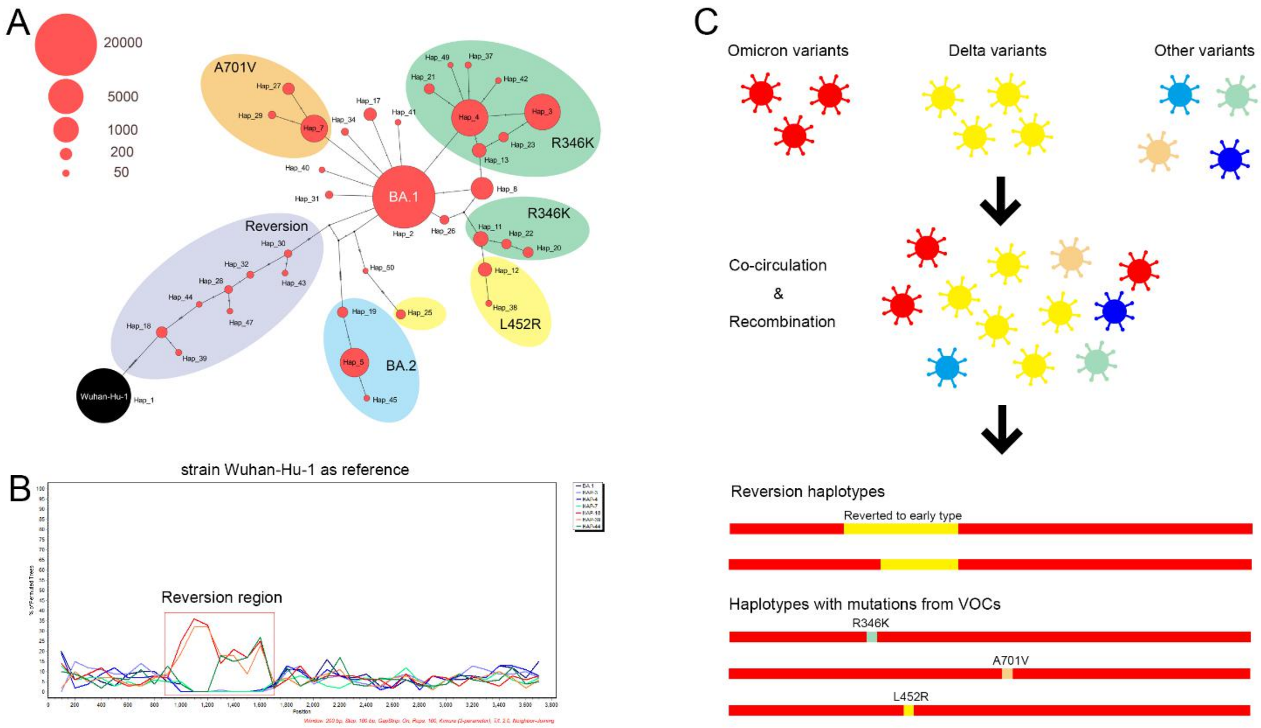


Figure 3. Phylogenetic network and scanning of the spike gene from representative Omicron subvariant sequences. (A) Representative Omicron spike protein haplotypes (each consisted of at least 50 sequences) were constructed with PopART using the median-joining method[42]. Nucleotide changes were notated with lines. The spike gene from Wuhan-Hu-1 strain was set as the root. The number of sequences in each haplotype were modified into different orders of magnitude, and subgroups based on the mutation types were delineated by color. **(B)** BootScan analysis of revertant and representative haplotypes of Omicron spike gene. Representative spike Omicron haplotypes (Hap_3, Hap_4, Hap_7) sequences and selected reversion haplotypes (Hap_18, Hap_39, Hap_44) sequences are included. Bootscan map was constructed by Simplot 3.5.1 (http://www.welch.jhu.edu/~ sray/download) using neighboring-joining method with 100 bootstrap replicates.Wuhan-Hu-1 spike sequences was set as reference, reversion region was annotated. **(C)** Overview of possible evolution mechanism of reversion haplotypes and haplotypes with mutations from Delta and other variants.

Co-infections of different SARS-CoV-2 variants in the population accelerates their evolution through recombination

Virus co-infection and recombination can amplify pathogenicity, for example, the well-studied "model organism" adenovirus [14-20], and also in coronaviruses[21-23]. SARS-CoV-2 has been shown to co-infect and recombine [26,43]. In host populations with disproportionate immunocompromised conditions, such as Africa[44], the possibility of long-term infections of SARS-CoV-2 variants may be higher than in populations otherwise healthy and/or vaccinated. A case report described prolonged infectious SARS-CoV-2 shedding up to 70 days from an asymptomatic immunocompromised individual with cancer[43]. A SARS-CoV-2 isolated from her presented with four new mutations within the spike protein and also eight in structural proteins and polymerase region. The marked within-host genomic evolution of SARS-CoV-2 with continuous turnover of dominant viral variants was observed[43]. Under reduced immune pressure or immune-suppression, long-term infections create conditions and increase the likelihood of simultaneous co-infections with multiple SARS-CoV-2 variants, and optimizing conditions for genome recombination. For example, on June 10, 2021, a passenger on a flight from Johannesburg, South Africa to Shenzhen, China tested positive for SARS-CoV-2[26]. The patient was found to be coinfected with two SARS-CoV-2 variants: Beta and Delta, with the ratio of the relative abundance between the two variants maintained at 1:9 (Beta: Delta) in a 14-day period. Furthermore, putative evidence of recombination in the Orf1ab and spike genes was shown[26]. Such recombination events may not be rare, especially considering that there are hundreds of variants circulating in the general population.

Among the Omicron subvariants and VOCs, many shared mutations were identified in this study. We speculate that some of the Omicron spike protein mutations resulted from co-infections of variants. Recombination among diverse variants may have contributed to the shared presence of different mutations between the VOCs. For example, the BA.1 subvariant has three more Alpha-related mutations (del69-70, delY144) than BA.2, and therefore may be phylogenetically closer to the Alpha variant, suggesting that Alpha or other unknown variants that carry these mutations may have contributed to the emergence of the BA.1 subvariant (Table 1). Multiple mutation differences causing reversion haplotypes may have originated from the recombination between the Omicron and other variants (Fig. 3 and Supplementary Table 1).

Except for shared mutations, many other mutations (30 in BA.1 and 25 in BA.2) could not be accounted for among previous dominant variants (Fig. 1). Because the Omicron variants are believed to have emerged in South Africa[1], we speculate that some of these spike mutations may have been produced by long-term virus infections in immunocompromised patients. It was previously reported that evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient led to immune escape variants[45,46]. Deletions in NTD, for example, delY144, were detected in multiple immunosuppressed COVID-19 patients, which resulted in immune escape[45-47].

This extended pandemic is likely yielding novel recombinant SARS-CoV-2 variants

A recently reported recombinant SARS-CoV-2, "Deltacron" or "Demicron", and its genome sequences, elicited controversy and concerns of sequencing errors and sample contamination[13]. Nevertheless, it was confirmed that co-infections by Omicron and Delta variants have already occurred in specific populations (https://www.gov.uk/government/publications/sars-cov-2-variants-of-public-health-interest/sars-cov-2-variants-of-public-health-interest-11-february-2022). Recombination among the extant variants may lead to the emergence of new variants. A total of 10 cases of "Deltacron" are underway to confirm by Santé publique France (SPF) (https://t.co/tVAKmHRYSy). In our study, multiple VOC and VOI mutations were detected in Omicron variants circulating before January 15, 2022 (Fig. 1 and Table 1). The integration of these mutations may lead to changes in phenotype. Five additional typical amino acid mutations in Delta variants were also identified in recently emergent Omicron isolates (before January 15, 2022) (Table 2). For example, 899 Omicron sequences of high quality contained L452R mutation reported for the Delta variant (Fig. 4A). Whole genome analysis also corroborated the diversity among these L452R containing Omicron genomes. The mutation profiles among whole genomes of BA.1 are diverse, and the sequences branched to diverse clades by phylogenetic analyses (Fig. 4B).

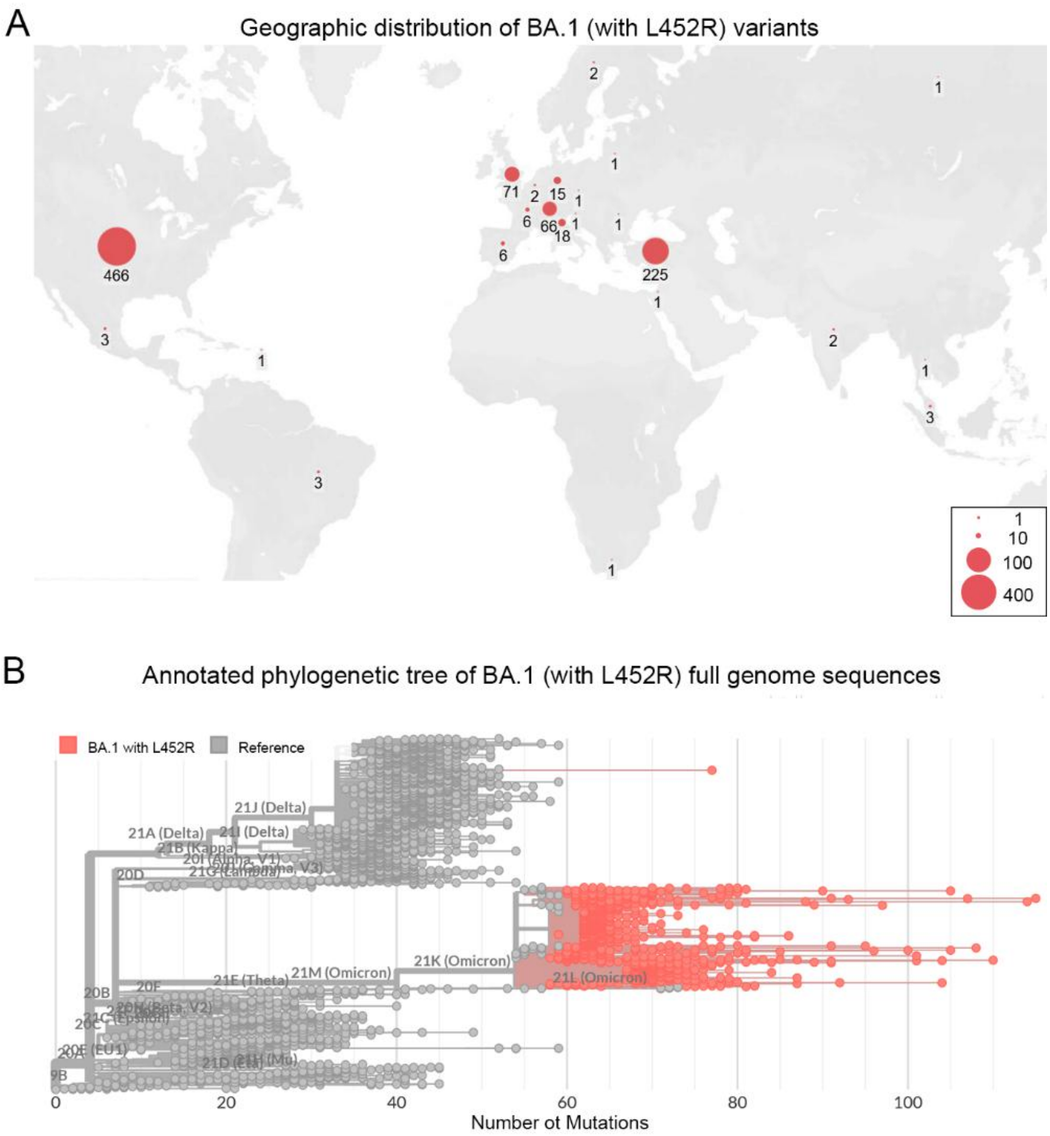


Figure 4. Geographic distribution and whole genome analyses of BA.1 (with L452R) variants. **(A)** Geographic distribution of BA.1 (with L452R) subvariants, with the number of genome sequences noted. (**B**) Whole genome phylogenetic tree highlighting the BA.1 (with L452R) subvariant, clades were annotated using NextClade (Hadfield et al., 2018). Low quality sequences were excluded. 891 SARS-CoV-2 BA.1 with L452R spike amino mutation full genome sequences submitted to the GISAID database before January 15th, 2022 and reference sequences from SARS-CoV-2 each clade were included. The red circles are BA.1 variants.

Conclusion

By analyzing sequences from a large number of Omicron subvariants, we identified diverse recombination events between two Omicron subvariants and several SARS-CoV-2 variants, suggesting that co-infection and subsequent genome recombination play important roles in the on-going evolution of SARS-CoV-2. Some of the recombination events may have led to modifications in protein function and viral fitness. Continued monitoring of SARS-CoV-2 genomes for mutations is critically important to our understanding of its evolution and impact on human health, and is also essential for the recognition of changes to viral epitopes that would require vaccine modifications.

Reference

1 Saxena, S. K. *et al.* Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. *J Med Virol*, doi:10.1002/jmv.27524 (2021).

2 Li, J., Lai, S., Gao, G. F. & Shi, W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* **600**, 408-418, doi:10.1038/s41586-021-04188-6 (2021).

3 Cui, Z. *et al.* Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. *Cell*, doi:10.1016/j.cell.2022.01.019 (2022).

4 Salleh, M. Z., Derrick, J. P. & Deris, Z. Z. Structural Evaluation of the Spike Glycoprotein Variants on SARS-CoV-2 Transmission and Immune Evasion. *Int J Mol Sci* **22**, doi:10.3390/ijms22147425 (2021).

5 Deng, X. *et al.* Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell* **184**, 3426-3437 e3428, doi:10.1016/j.cell.2021.04.025 (2021).

6 Li, Q. *et al.* The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **182**, 1284-1294 e1289, doi:10.1016/j.cell.2020.07.012 (2020).

7 Liu, Y. *et al.* The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*, doi:10.1038/s41586-021-04245-0 (2021).

8 Saito, A. *et al.* Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature*, doi:10.1038/s41586-021-04266-9 (2021).

9 Ou, J. *et al.* V367F mutation in SARS-CoV-2 spike RBD emerging during the early transmission phase enhances viral infectivity through increased human ACE2 receptor binding affinity. *J. Virol.* **95**, e0061721, doi:doi:10.1128/JVI.00617-21 (2021).

10 Bayarri-Olmos, R. *et al.* Functional Effects of Receptor-Binding Domain Mutations of SARS-CoV-2 B.1.351 and P.1 Variants. *Front Immunol* **12**, 757197, doi:10.3389/fimmu.2021.757197 (2021).

11 Collier, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136-141, doi:10.1038/s41586-021-03412-7 (2021).

12 Dejnirattisai, W. *et al.* Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **184**, 2939-2954 e2939, doi:10.1016/j.cell.2021.03.055 (2021).

13 Kreier, F. Deltacron: the story of the variant that wasn't. **602**, 19-19 (2022).

14 Dehghan, S. *et al.* A zoonotic adenoviral human pathogen emerged through genomic recombination amongst human and nonhuman simian hosts. *Journal of virology* **93**, e00564-00519, doi:10.1128/JVI.00564-19 (2019).

15 Walsh, M. P. *et al.* Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLoS ONE* **4**, e5635, doi:10.1371/journal.pone.0005635 (2009).

16 Walsh, M. P. *et al.* Computational analysis identifies human adenovirus type 55 as a re-emergent acute respiratory disease pathogen. *J. Clin. Microbiol.* **48**, 991-993, doi:10.1128/JCM.01694-09 (2010).

17 Purkayastha, A. *et al.* Genomic and bioinformatics analysis of HAdV-4, a human adenovirus causing acute respiratory disease: implications for gene therapy and vaccine vector development. *J. Virol.* **79**, 2559-2572, doi:10.1128/JVI.79.4.2559-2572.2005 (2005).

18 Dehghan, S. *et al.* Computational analysis of four human adenovirus type 4 genomes reveals molecular evolution through two interspecies recombination events. *Virology* **443**, 197-207, doi:10.1016/j.virol.2013.05.014 (2013).

19 Dehghan, S. *et al.* Simian adenovirus type 35 has a recombinant genome comprising human and simian adenovirus sequences, which predicts its potential emergence as a human respiratory pathogen. *Virology* **447**, 265-273, doi:10.1016/j.virol.2013.09.009 (2013).

20 Zhang, J. *et al.* A Survey of Recent Adenoviral Respiratory Pathogens in Hong Kong Reveals Emergent and Recombinant Human Adenovirus Type 4 (HAdV-E4) Circulating in Civilian Populations. *Viruses* **11**, 129, doi:10.3390/v11020129 (2019).

21 Chu, D. K. W. *et al.* MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. *Proc Natl Acad Sci U S A* **115**, 3144-3149, doi:10.1073/pnas.1718769115 (2018).

22 Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* **13**, e1006698, doi:10.1371/journal.ppat.1006698 (2017).

23 Hon, C. C. *et al.* Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J Virol* **82**, 1819-1826, doi:10.1128/JVI.01926-07 (2008).

24 Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* **5**, 1408-1417, doi:10.1038/s41564-020-0771-4 (2020).

25 Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276-278, doi:10.1126/science.1087139 (2003).

26 He, Y. *et al.* Possible recombination between two variants of concern in a COVID-19 patient. *Emerg Microbes Infect*, 1-26, doi:10.1080/22221751.2022.2032375 (2022).

27 Tang, X. *et al.* Evolutionary analysis and lineage designation of SARS-CoV-2 genomes. *Sci Bull (Beijing)* **66**, 2297-2311, doi:10.1016/j.scib.2021.02.012 (2021).

28 Yi, H. 2019 Novel Coronavirus Is Undergoing Active Recombination. *Clin Infect Dis* **71**, 884-887, doi:10.1093/cid/ciaa219 (2020).

29 Chen, C. *et al.* TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Molecular Plant* **13**, 1194-1202, doi:https://doi.org/10.1016/j.molp.2020.06.009 (2020).

30 Zhou, Z. J., Qiu, Y., Pu, Y., Huang, X. & Ge, X. Y. BioAider: An efficient tool for viral genome analysis and its application in tracing SARS-CoV-2 transmission. *Sustain Cities Soc* **63**, 102466, doi:10.1016/j.scs.2020.102466 (2020).

31 Hall, T. A. BioEdit : a user-friendly biological sequence alignment editor and analysis. *Nuclc Acids Symp* **41** (1999).

32 Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn diagram viewer. *BMC bioinformatics* **15**, 293, doi:10.1186/1471-2105-15-293 (2014).

33 Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293, doi:10.1186/1471-2105-15-293 (2014).

34 Delano, W. L. The PyMol Molecular Graphics System. *Proteins Structure Function and Bioinformatics* **30**, 442-454 (2002).

35 Han, P. *et al.* Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell*, doi:10.1016/j.cell.2022.01.001 (2022).

36 Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121-4123, doi:10.1093/bioinformatics/bty407 (2018).

37 Lyngse, F. P. *et al.* Transmission of SARS-CoV-2 Omicron VOC subvariants BA.1 and BA.2: Evidence from Danish Households. *medRxiv* (2022).

38 Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **19**, 409-424, doi:10.1038/s41579-021-00573-0 (2021).

39 Huang, T. Y. & Zhao, B. tidyfst: Tidy Verbs for Fast Data Manipulation. *The Journal of Open Source Software* **5**, 2388 (2020).

40 Katoh & K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066 (2002).

41 Librado & Rozas. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* **25**, 1451-1452 (2009).

42 Bandelt, H.-J., Forster, P. & R?hl, A. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48 (1999).

43 Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901-1912 e1909, doi:10.1016/j.cell.2020.10.049 (2020).

44 Ghosn, J., Taiwo, B., Seedat, S., Autran, B. & Katlama, C. HIV. *Lancet* **392**, 685-697, doi:10.1016/s0140-6736(18)31311-4 (2018).

45 Clark, S. A. *et al.* SARS-CoV-2 evolution in an immunocompromised host reveals shared neutralization escape mechanisms. *Cell* **184**, 2605-2617.e2618, doi:10.1016/j.cell.2021.03.027 (2021).

46 Weigang, S. *et al.* Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants. *Nat Commun* **12**, 6405, doi:10.1038/s41467-021-26602-3 (2021).

47 Chen, L. *et al.* Emergence of Multiple SARS-CoV-2 Antibody Escape Variants in an Immunocompromised Host Undergoing Convalescent Plasma Treatment. *mSphere* **6**, e0048021, doi:10.1128/mSphere.00480-21 (2021).

Author contribution statement

JO and QZ contribute to study design and manuscript writing. JO, WL, XW, TZ, BD, PY, YR, LQ, and QZ contribute to data analysis and data visualization. WZ, DS, JC, JW and QZ contribute to manuscript revision.

Acknowledgments

We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiFlu™ Database on which this research is based. All submitters of data may be contacted directly via www.gisaid.org.

Funding statement

This work was supported by grants from the National Key Research and Development Program of China (2018YFE0204503), National Natural Science Foundation of China (32170139 and 81730061), and Natural Science Foundation of Guangdong Province (2018B030312010 and 2021A1515010788).

Competing interests

The authors declare no competing interests.

Figure 1. Spike protein amino acid mutations of the Omicron subvariants (BA.1 and BA.2) compared with mutations from the other four variants of concern (VOCs). (A) Venn diagram noting mutations of Omicron (BA.1) and those of VOCs. **(B)** Venn diagram of Omicron (BA.2) mutations compared to ones of VOCs. **(C)** Venn diagram of mutations between Omicron (BA.1) and Omicron (BA.2). **(D)** Spike protein amino acid mutation counts of Omicron (BA.1 and BA.2) subvariants compared with mutations of VOCs.

Figure 2. Structure of the Spike protein with amino acid mutations detected in Omicron BA.1 subvariant. (A) Structure of human ACE2 receptor complexed with SARS-CoV-2 Omicron RBD, mapped with the recent mutations. (**B)** Structure of SARS-CoV-2 Omicron spike protein mapped with the novel mutations. Mutated residues in each domain of the spike protein are annotated in color (red: RBD; yellow: NTD; green: S1/S2; blue: S2) using with Pymol 2.0 software through SARS-CoV-2 Omicron model PDB:7WBL and 7QO7 (Han, P. et al. Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. Cell, 2022, doi:10.1016/j.cell.2022.01.001).

Figure 3. Phylogenetic network and scanning of the spike gene from representative Omicron subvariant sequences. (A) Representative Omicron spike protein haplotypes (each consisted of at least 50 sequences) were constructed with PopART using the median-joining method[42]. Nucleotide changes were notated with lines. The spike gene from Wuhan-Hu-1 strain was set as the root. The number of sequences in each haplotype were modified into different orders of magnitude, and subgroups based on the mutation types were delineated by color. **(B)** BootScan analysis of revertant and representative haplotypes of Omicron spike gene. Representative spike Omicron haplotypes (Hap_3, Hap_4, Hap_7) sequences and selected reversion haplotypes (Hap_18, Hap_39, Hap_44) sequences are included. Bootscan map was constructed by Simplot 3.5.1 (http://www.welch.jhu.edu/~sray/download) using neighboring-joining method with 100 bootstrap replicates.Wuhan-Hu-1 spike sequences was set as reference, reversion region was annotated. **(C)**

Overview of possible evolution mechanism of reversion haplotypes and haplotypes with mutations from Delta and other variants.

Figure 4. Geographic distribution and whole genome analyses of BA.1 (with L452R) variants. **(A)** Geographic distribution of BA.1 (with L452R) subvariants, with the number of genome sequences noted. (**B)** Whole genome phylogenetic tree highlighting the BA.1 (with L452R) subvariant, clades were annotated using NextClade (Hadfield et al., 2018). Low quality sequences were excluded. 891 SARS-CoV-2 BA.1 with L452R spike amino mutation full genome sequences submitted to the GISAID database before January 15th, 2022 and reference sequences from SARS-CoV-2 each clade were included. The red circles are BA.1 variants.

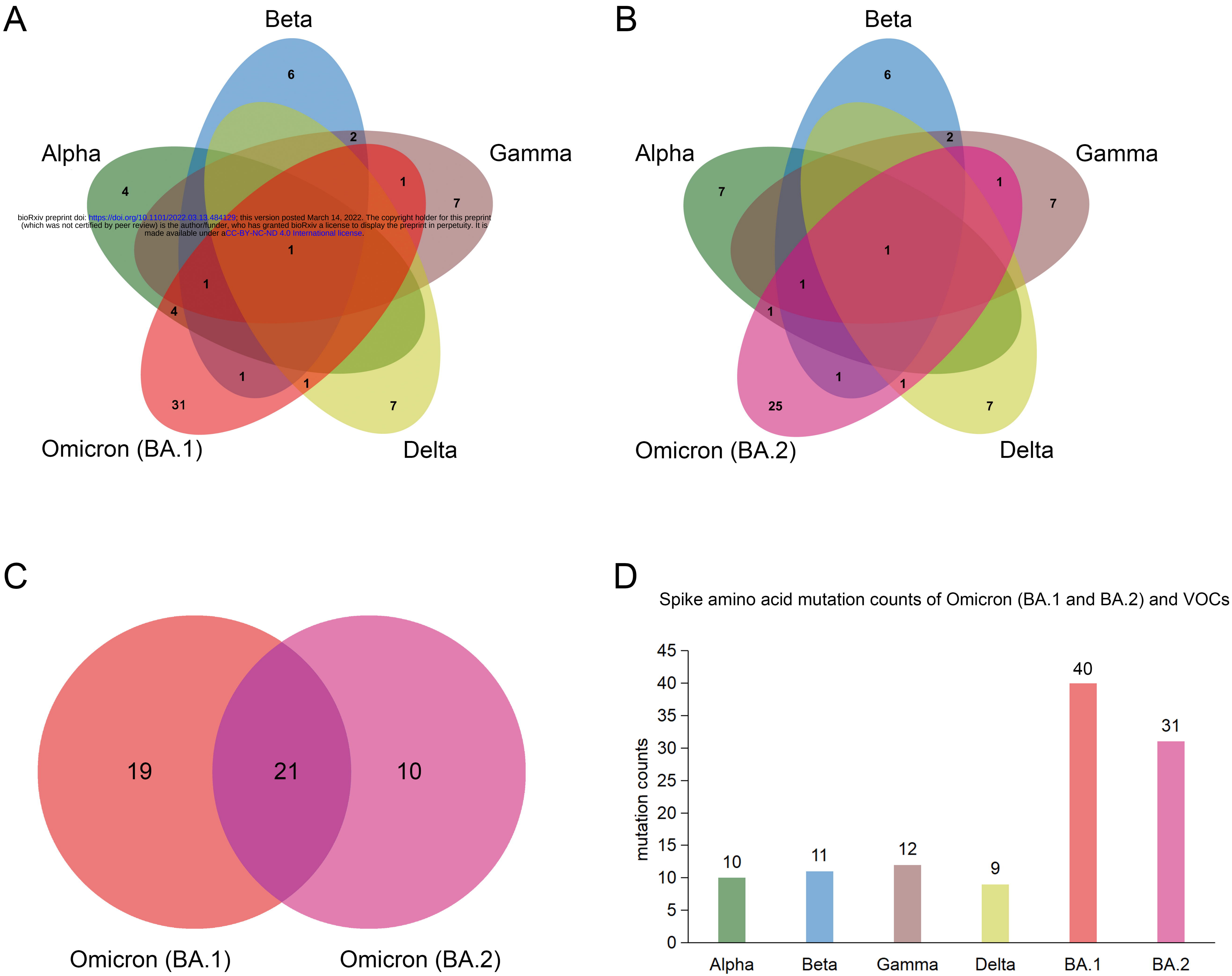
Table 1. Comparison of Spike protein amino acid mutations between the Omicron subvariants and other VOCs and VOIs. 52,563 high quality Omicron spike gene sequences (49,609 BA.1 sequences, and 2,954 BA.2 sequences) released before January 15, 2022 were analyzed. The mutations that have appeared in more than 800 sequences were used in this analysis. VOCs are variants of concern; VOIs are variants of interest.

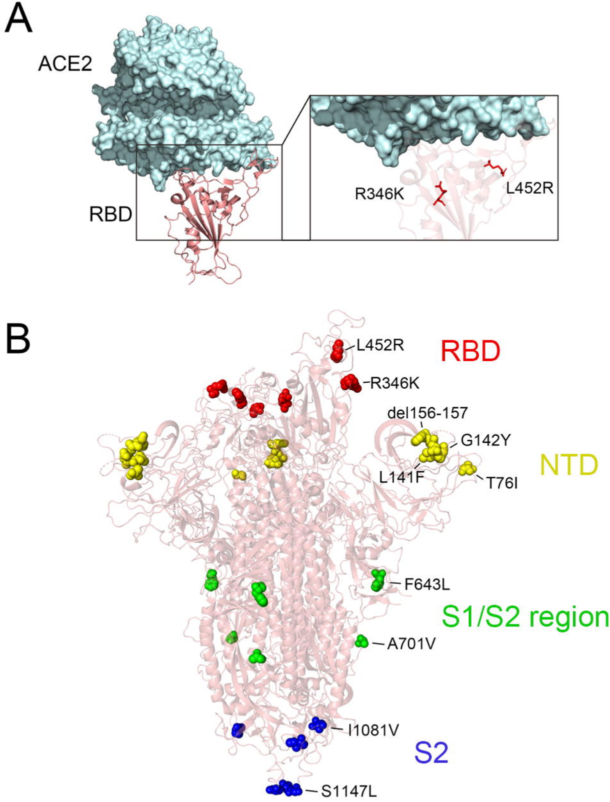
Table 2. Novel mutations identified in the spike protein of the recently emerged Omicron subvariants (Released before January 15, 2022; frequency >50 sequences).

Supplementary information

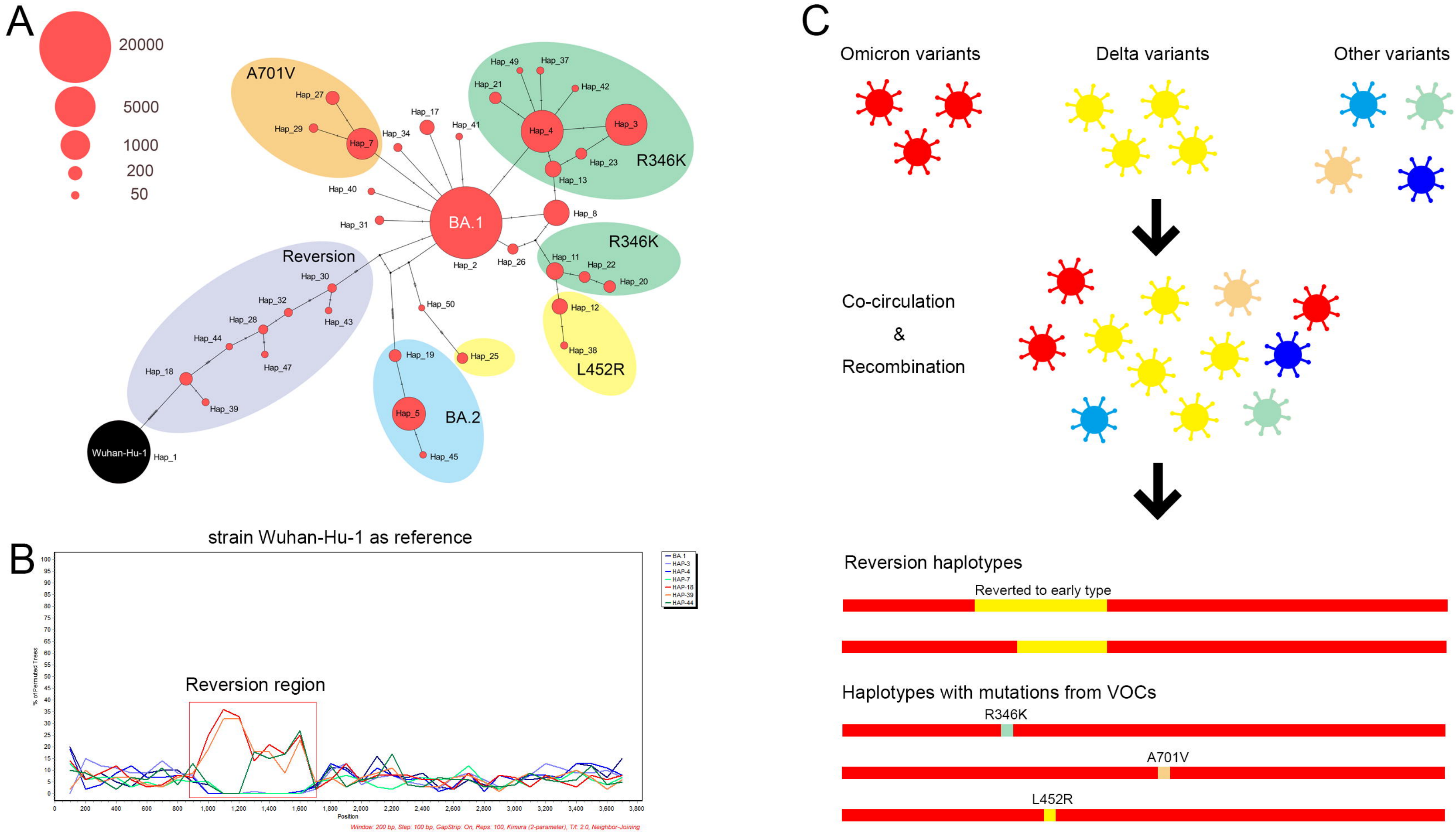
Supplementary Table 1. The annotation of haplotypes of Omicron spike protein.

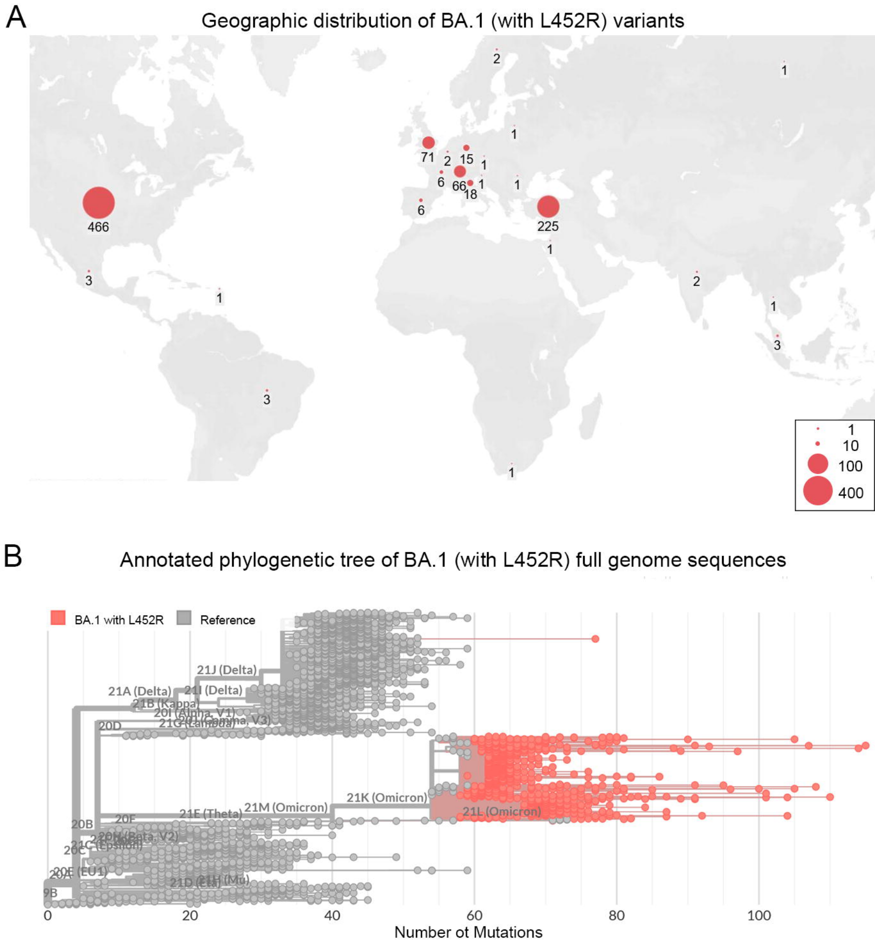
This version was submitted for peer review on March 3, 2022.


A
Beta
Alpha
Gamma
Omicron (BA.1)
Delta
6
2
4
1
7
1
1
4
1
1
31
7
B
Beta
Alpha
Gamma
Omicron (BA.2)
Delta
6
2
7
1
7
1
1
1
1
1
25
7
C
19
21
10
Omicron (BA.1)
Omicron (BA.2)
D
Spike amino acid mutation counts of Omicron (BA.1 and BA.2) and VOCs
mutation counts
0
5
10
15
20
25
30
35
40
45
10
11
12
9
40
31
Alpha
Beta
Gamma
Delta
BA.1
BA.2



A
ACE2
RBD
R346K
L452R
B
L452R
R346K
RBD
del156-157
G142Y
L141F
T76I
NTD
F643L
S1/S2 region
A701V
I1081V
S2
S1147L


A
20000
5000
1000
200
50
A701V
Hap_27
Hap_29
Hap_7
Hap_34
Hap_17
Hap_41
Hap_49
Hap_37
Hap_21
Hap_42
Hap_4
Hap_3
Hap_23
R346K
Hap_13
Hap_40
BA.1
Hap_8
Hap_31
Hap_2
Hap_26
R346K
Hap_11
Hap_22
Hap_20
Reversion
Hap_30
Hap_32
Hap_43
Hap_28
Hap_44
Hap_47
Hap_50
Hap_12
Hap_38
L452R
Hap_19
Hap_25
Hap_18
Hap_39
Hap_5
BA.2
Hap_45
Wuhan-Hu-1
Hap_1
B
strain Wuhan-Hu-1 as reference
% of Permuted Trees
Position
BA.1
HAP-3
HAP-4
HAP-7
HAP-18
HAP-39
HAP-44
Reversion region
Window: 200 bp, Step: 100 bp, GapStrip: On, Reps: 100, Kimura (2-parameter), T/t: 2.0, Neighbor-Joining
C
Omicron variants
Delta variants
Other variants
Co-circulation
&
Recombination
Reversion haplotypes
Reverted to early type
Haplotypes with mutations from VOCs
R346K
A701V
L452R


A
Geographic distribution of BA.1 (with L452R) variants
2
1
1
71
2
15
1
6
66
1
18
1
466
6
225
1
3
1
2
1
3
3
1
1
10
100
400
B
Annotated phylogenetic tree of BA.1 (with L452R) full genome sequences
BA.1 with L452R
Reference
21J (Delta)
21A (Delta)
21I (Delta)
21B (Kappa)
20D
21K (Omicron)
21E (Theta)
21M (Omicron)
21L (Omicron)
20F
20B
20E (EU1)
9B
0
20
40
60
80
100
Number of Mutations