

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

3 Junxian Ou^{1,2}, Wendong Lan², Xiaowei Wu², Tie Zhao¹, Biyan Duan¹, Peipei Yang¹, Yi Ren¹, Lulu
4 Quan², Wei Zhao², Donald Seto³, James Chodosh⁴, Jianguo Wu^{1,5*} & Qiwei Zhang^{1,2,5*}

5

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

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

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

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

14 ⁵Foshan Institute of Medical Microbiology, Foshan 528315, China

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

17

18

19

20

21

22

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

3

4 **Abstract**

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

27

1 **Introduction**

2 The current COVID-19 pandemic is fueled by a more infectious emergent Omicron variant
3 (B.1.1.529), which was first reported in South Africa and quickly spread worldwide¹. A multitude of
4 mutations (more than 30) in the spike gene of Omicron variant were detected, which when compared
5 to the Alpha and Delta variants (typically less than 15)², raised concerns of enhanced infectivity and
6 immune escape potential^{3,4}. Omicron variants is divided into three lineages (BA.1, BA.2, and BA.3)
7 and was classified as the fifth variant of concern (VOC) by the World Health Organization on
8 November 26, 2021. It has been circulating in more than 170 countries/territories.

9
10 Mutations in the SARS-CoV-2 spike gene have altered protein binding efficiency and
11 immunogenicity, and resulted in more invasive and adaptive variants⁴⁻⁹. Previous research on Alpha
12 (B.1.1.7) and Delta (B.1.617.2 and AY.x) variants with spike gene mutations confirmed these
13 effects on enhancing virus transmission⁴⁻⁸. Meanwhile, as a critical antigenic recognition site, the
14 spike protein is also the principal vaccine design target, and these observed mutations have focused
15 attention on this modified antigen and its putative immune escape potential and antibody
16 resistance^{3,10-12}.

17
18 Ongoing concerns of emergent variants includes possible recombinants resulting from different
19 variants replicating simultaneously in a host. Such variants, e.g., “Demicron” or “Deltacron” are
20 controversial that if they are real recombinants or a possible sequencing error¹³.

21
22 Genome recombination is an important evolutionary mechanism for the emergence and
23 re-emergence of human pathogens and a major source of viral evolution, for example, the
24 well-studied “model organism” adenovirus¹⁴⁻²⁰, and also in coronaviruses²¹⁻²³. Recombination
25 accelerates virus evolution through gene(s) and “function” transference and accumulation of
26 selective and advantageous mutations, resulting in phenotype changes that include changes in

1 pathogenicity profiles, host species virulence, zoonotic and anthroponotic transmission, and host
2 adaptation^{14-21,24,25}.

3
4 Although recombination events among SARS-CoV-1 and MERS-CoV were well-documented²¹⁻²³, it
5 has been difficult to detect the recombination signatures in SARS-CoV-2 variants due to the high
6 degree of sequence similarity amongst SARS-CoV-2 isolates and the incomplete coverage of
7 coronaviruses from other hosts, including pangolin^{26,27}.

8
9 Previous research distinguished active recombination events among the SARS-CoV-2 nucleoprotein
10 and ORF1ab genes by using a phylogenetic network strategy based on single nucleotide substitution
11 or SARS-CoV-2 lineage designation^{27,28}. More than thirty amino acid mutations have been identified
12 within Omicron spike protein, some of which are shared with other variants¹. In this study, we
13 demonstrate that the emerging and circulating Omicron subvariants originate in part through
14 recombination with other variants. We first investigated the spike diversity of the Omicron variants
15 along with the shared spike mutations between Omicron and other variants of concern (VOCs) and
16 variants of interest (VOIs). The Omicron spike amino acid sequences archived during the early
17 transmission phase, and released in the GISAID database (submitted before January 15th, 2022) were
18 accessed, include 52,563 high quality Omicron spike sequences (representing 49,609 BA.1 and 2,954
19 BA.2 sequences). In this study, these were analyzed with Pymol 2.0, TBTools, BioEdit, BioAider,
20 and jvenn²⁹⁻³⁵. The whole genome phylogenetic trees were constructed and annotated using
21 NextClade³⁶.

22 23 **Tracking the common mutations among Omicron (BA.1 and BA.2) and variants of concern** 24 **(VOCs)**

25 Circulating Omicron variant consists of two main subvariants, BA.1 and BA.2. BA.1 subvariant was
26 more frequently detected than BA.2 during the early transmission phase. However, BA.2 is replacing
27 BA.1 as the dominant epidemic subvariant in more and more countries over time³⁷.

1

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

7

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

18

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

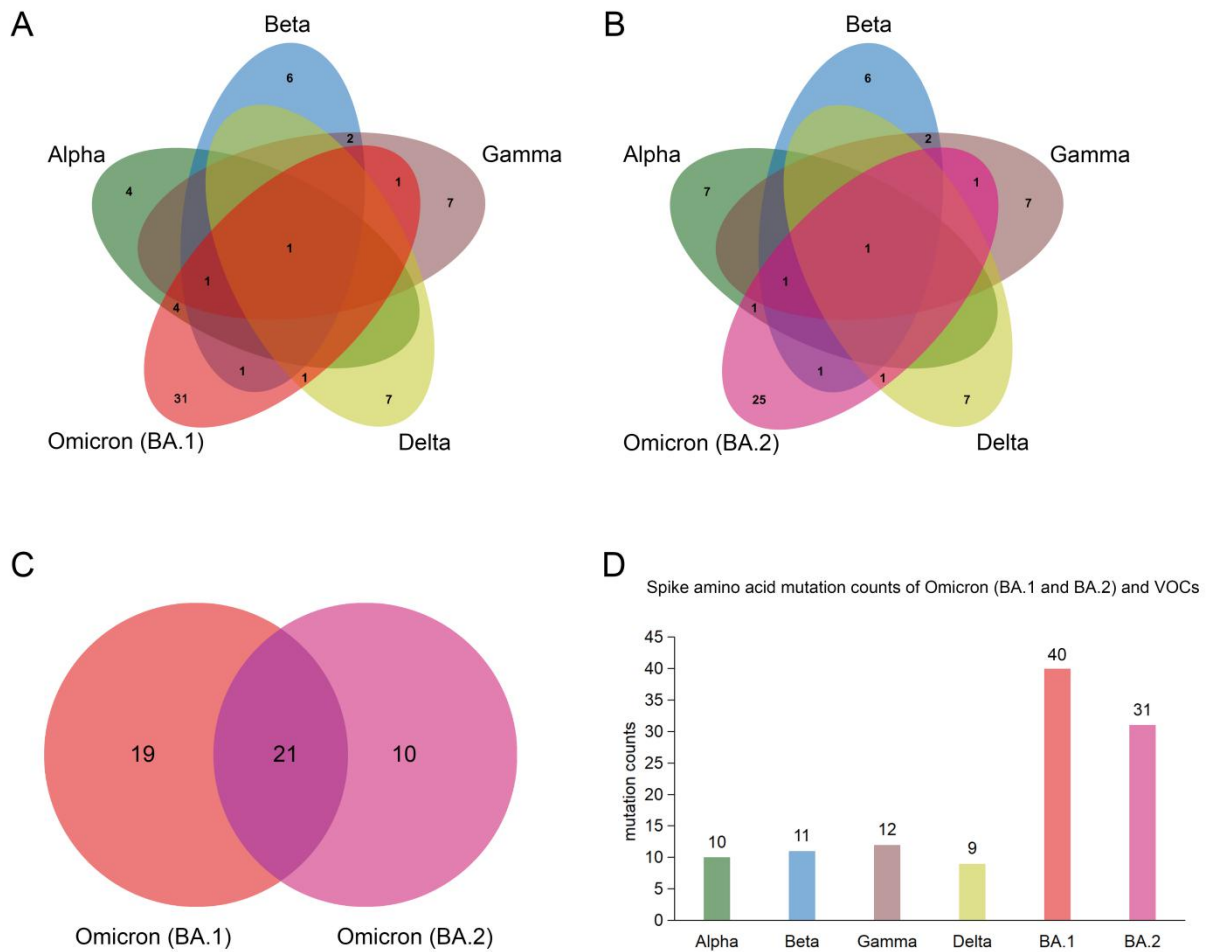
24

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

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

1

2



3

4 **Figure 1. Spike protein amino acid mutations of the Omicron subvariants (BA.1 and BA.2)**

5 **compared with mutations from the other four variants of concern (VOCs). (A) Venn diagram**

6 **noting mutations of Omicron (BA.1) and those of VOCs. (B) Venn diagram of Omicron (BA.2)**

7 **mutations compared to ones of VOCs. (C) Venn diagram of mutations between Omicron (BA.1) and**

8 **Omicron (BA.2). (D) Spike protein amino acid mutation counts of Omicron (BA.1 and BA.2)**

9 **subvariants compared with mutations of VOCs.**

10

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

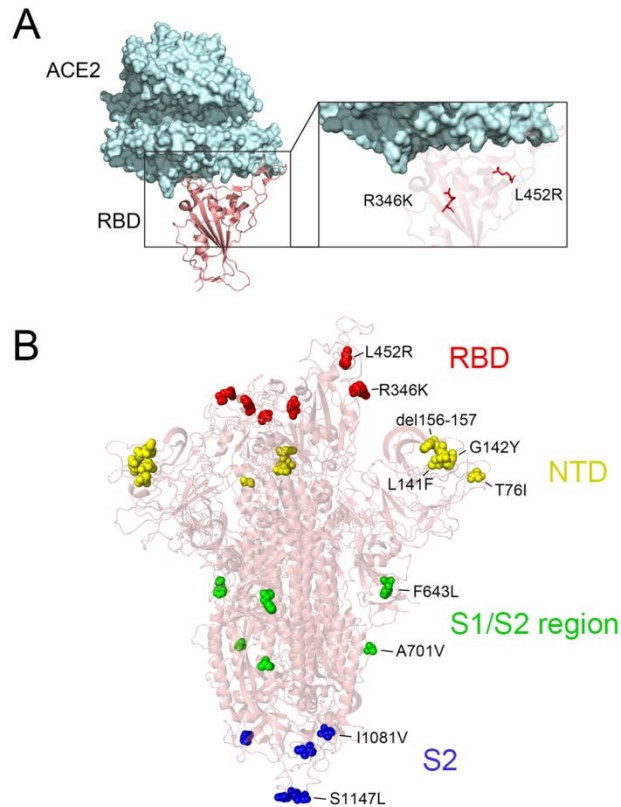
6

Spike Region	Position	Mutation (BA.1)	Frequency (BA.1)	Percentage	Mutation (BA.2)	Frequency (BA.2)	Percentage	Mutation in VOCs/VOIs	New Mutation	
NTD	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)		
	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%						
S1	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%			
	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)		
	RBD	440	N440K	46470	93.67%	N440K	2926	99.09%		
		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%			
SD	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)		
	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	
S2	FP	764	N764K	49046	98.87%	N764K	2953	100.00%		
		796	D796Y	49338	99.45%	D796Y	2952	99.97%		
		856	N856K	49488	99.76%					
	HR1	954	Q954H	49559	99.90%	Q954H	2953	100.00%		
		969	N969K	49537	99.85%	N969K	2953	100.00%		
	981	L981F	49373	99.52%						

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

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

13
14 Several novel mutations were reported to be related to spike protein function, resulting in an
15 enhancement of virus infectivity or in viral immune escape. Mutations that occurred in RBD, e.g.,
16 R346K, could result in a relatively weakened neutralizing antibody effect³. A L452R mutation may
17 provide evasion from cellular immunity and increased infectivity^{5,6}. The P681R as well as F643L and
18 A701V mutations, near the S1/S2 cleavage site, may be associated with enhanced fusogenicity and
19 pathogenicity of SARS-CoV-2 Delta variants⁸. Additionally, mutations T76I, L141F, G142Y,
20 156-167deletion, and R158G, located in the NTD region, were noted to affect antibody binding
21 efficiencies and contribute to immune escape³⁸. These mutations sites are mapped and shown in Fig.
22 2.



1

2 **Figure 2. Structure of the Spike protein with amino acid mutations detected in Omicron BA.1**

3 **subvariant. (A)** Structure of human ACE2 receptor complexed with SARS-CoV-2 Omicron RBD,

4 mapped with the recent mutations. **(B)** Structure of SARS-CoV-2 Omicron spike protein mapped

5 with the novel mutations. Mutated residues in each domain of the spike protein are annotated in color

6 (red: RBD; yellow: NTD; green: S1/S2; blue: S2) using with Pymol 2.0 software through

7 SARS-CoV-2 Omicron model PDB:7WBL and 7QO7 (Han, P. et al. Receptor binding and complex

8 structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. Cell, 2022,

9 doi:10.1016/j.cell.2022.01.001).

10

1 Apparent revertant mutations are found in some dominant mutations (frequency >95%) in the BA.1
2 subvariant during the pandemic. Examples are the mutations in NTD (del211 and N212I) and RBD
3 (G339D, S371L, S373P, S375F, K417N, 440K, G446S, S477N, T478K, E484A, Q493R, G496S,
4 Q498R, N501Y, and Y505H). The frequency of insertions of the amino acids EPE at site 214 in
5 BA.1 decreased during the pandemic from more than 95% on December 1st, 2021 to 89% on January
6 15th 2022. However, BA.2 spike protein remained constant (frequency >99%), with the exception of
7 the three amino acid deletion (LPP) found at amino acids 24-26, which decreased from more than
8 95% frequency on December 1st, 2021 to 85% on January 15th 2022 (Table 1). This may possibly be
9 due to selection pressure on the circulating Omicron strains.

10

1 **Table 2. Novel mutations identified in the spike protein of the recently emerged Omicron**
 2 **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	
	77	T76I	51	0.10%	Lambda (T76I)	2021.11.26	
	NTD	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
	RBD	158	R158G	69	0.14%	Delta(E158G)	2021.12.15
		346	R346K	16819	33.90%	Mu(R346K)	2021.11.4
	SD	452	L452R	899	1.81%	Delta(L452R)	2021.11.11
		643	F643L	138	0.28%		2021.11.29
S2	681	P681R	62	0.12%	Delta(P681R)	2021.11.23	
	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	

3

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

3 Further investigation of the Omicron subvariants examined spike protein haplotypes . These were
4 identified, screened, and calculated with the R package tidyfst, aligned with MAFFT^{39,40}. These
5 haplotypes were analyzed with DnaSP 6.0⁴¹, and a subsequent phylogenetic network was constructed
6 using PopART (<http://popart.otago.ac.nz/index.shtml>), with mutations annotated with Nextclade
7 (<https://clades.nextstrain.org>).

8

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

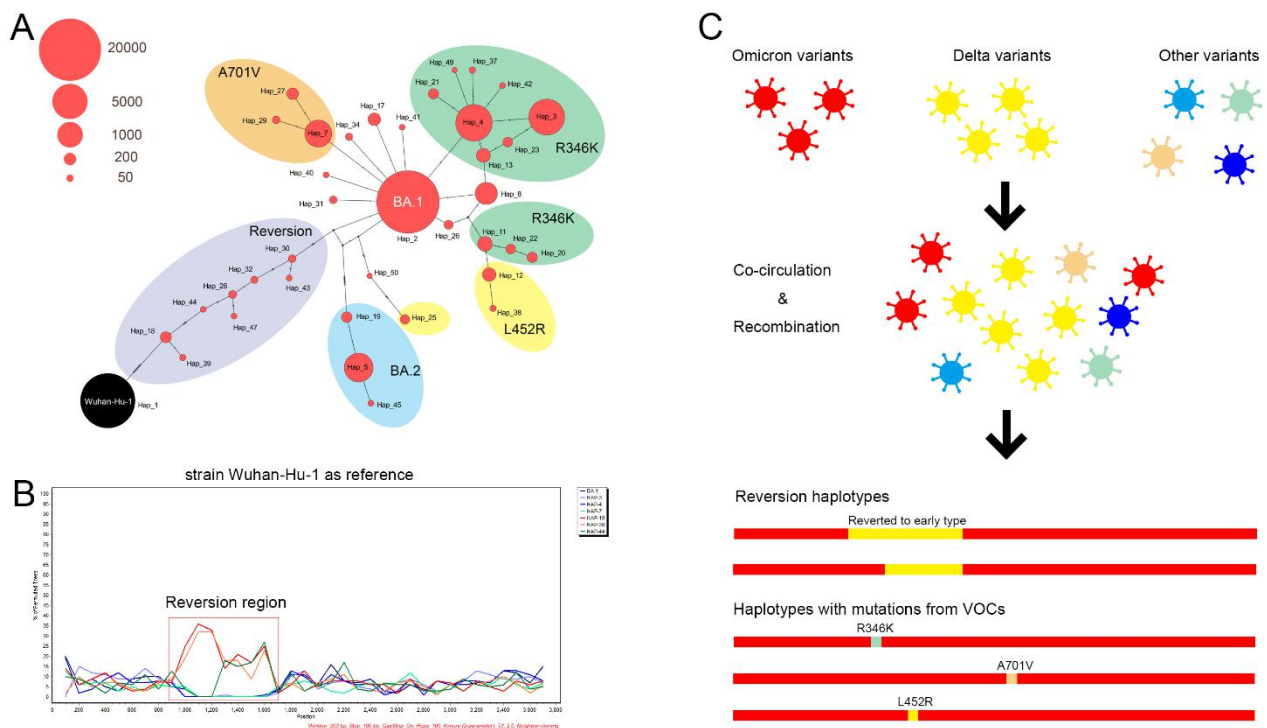
14

15 Multiple nucleotide mutations were detected in the haplotypes compared with BA.1, e.g., haplotype
16 19 and the revertant subgroup (Hap 30, 32, 43, *etc.*) and BA.2. The L452R subgroup consists of
17 different haplotypes with multiple nucleotide substitutions, indicating a possibly separate origin of
18 L452R haplotypes or prior recombination events. The revertant subgroup consisted of Omicron
19 haplotypes in which several BA.1 representative mutations were lost and appeared to have reverted
20 to the bases of the Wu-hu-1 strain. Multiple nucleotide differences in other haplotypes occurred,
21 likely as multiple independent mutation events, or perhaps as recombination events among highly
22 similar sequences. Haplotype 25 in L452R subgroup, with multiple nucleotide differences compared
23 with BA.1, could have resulted through recombination between Omicron and Delta variants, gaining
24 the mutation L452R from Delta and losing multiple mutations from Omicron (**Fig. 3A**). Some of
25 these “Demicron” or “Deltacron” haplotypes are being tracked by the UK Health Security Agency
26 ([https://www.gov.uk/government/publications/sars-cov-2-variants-of-public-health-interest/sars-cov-](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)
27 [2-variants-of-public-health-interest-25-february-2022](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

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

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

10



1

2 **Figure 3. Phylogenetic network and scanning of the spike gene from representative Omicron**
 3 **subvariant sequences. (A)** Representative Omicron spike protein haplotypes (each consisted of at
 4 least 50 sequences) were constructed with PopART using the median-joining method⁴². Nucleotide
 5 changes were notated with lines. The spike gene from Wuhan-Hu-1 strain was set as the root. The
 6 number of sequences in each haplotype were modified into different orders of magnitude, and
 7 subgroups based on the mutation types were delineated by color. **(B)** BootScan analysis of revertant
 8 and representative haplotypes of Omicron spike gene. Representative spike Omicron haplotypes
 9 (Hap_3, Hap_4, Hap_7) sequences and selected reversion haplotypes (Hap_18, Hap_39, Hap_44)
 10 sequences are included. Bootscan map was constructed by Simplot 3.5.1
 11 (<http://www.welch.jhu.edu/~sray/download>) using neighboring-joining method with 100 bootstrap
 12 replicates. Wuhan-Hu-1 spike sequences was set as reference, reversion region was annotated. **(C)**
 13 Overview of possible evolution mechanism of reversion haplotypes and haplotypes with mutations
 14 from Delta and other variants.

15

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

3 Virus co-infection and recombination can amplify pathogenicity, for example, the well-studied
4 “model organism” adenovirus¹⁴⁻²⁰, and also in coronaviruses²¹⁻²³. SARS-CoV-2 has been shown to
5 co-infect and recombine^{26,43}. In host populations with disproportionate immunocompromised
6 conditions, such as Africa⁴⁴, the possibility of long-term infections of SARS-CoV-2 variants may be
7 higher than in populations otherwise healthy and/or vaccinated. A case report described prolonged
8 infectious SARS-CoV-2 shedding up to 70 days from an asymptomatic immunocompromised
9 individual with cancer⁴³. A SARS-CoV-2 isolated from her presented with four new mutations
10 within the spike protein and also eight in structural proteins and polymerase region. The marked
11 within-host genomic evolution of SARS-CoV-2 with continuous turnover of dominant viral variants
12 was observed⁴³. Under reduced immune pressure or immune-suppression, long-term infections create
13 conditions and increase the likelihood of simultaneous co-infections with multiple SARS-CoV-2
14 variants, and optimizing conditions for genome recombination. For example, on June 10, 2021, a
15 passenger on a flight from Johannesburg, South Africa to Shenzhen, China tested positive for
16 SARS-CoV-2²⁶. The patient was found to be coinfecting with two SARS-CoV-2 variants: Beta and
17 Delta, with the ratio of the relative abundance between the two variants maintained at 1:9 (Beta:
18 Delta) in a 14-day period. Furthermore, putative evidence of recombination in the Orf1ab and spike
19 genes was shown²⁶. Such recombination events may not be rare, especially considering that there are
20 hundreds of variants circulating in the general population.

21
22 Among the Omicron subvariants and VOCs, many shared mutations were identified in this study. We
23 speculate that some of the Omicron spike protein mutations resulted from co-infections of variants.
24 Recombination among diverse variants may have contributed to the shared presence of different
25 mutations between the VOCs. For example, the BA.1 subvariant has three more Alpha-related
26 mutations (del69-70, delY144) than BA.2, and therefore may be phylogenetically closer to the Alpha
27 variant, suggesting that Alpha or other unknown variants that carry these mutations may have

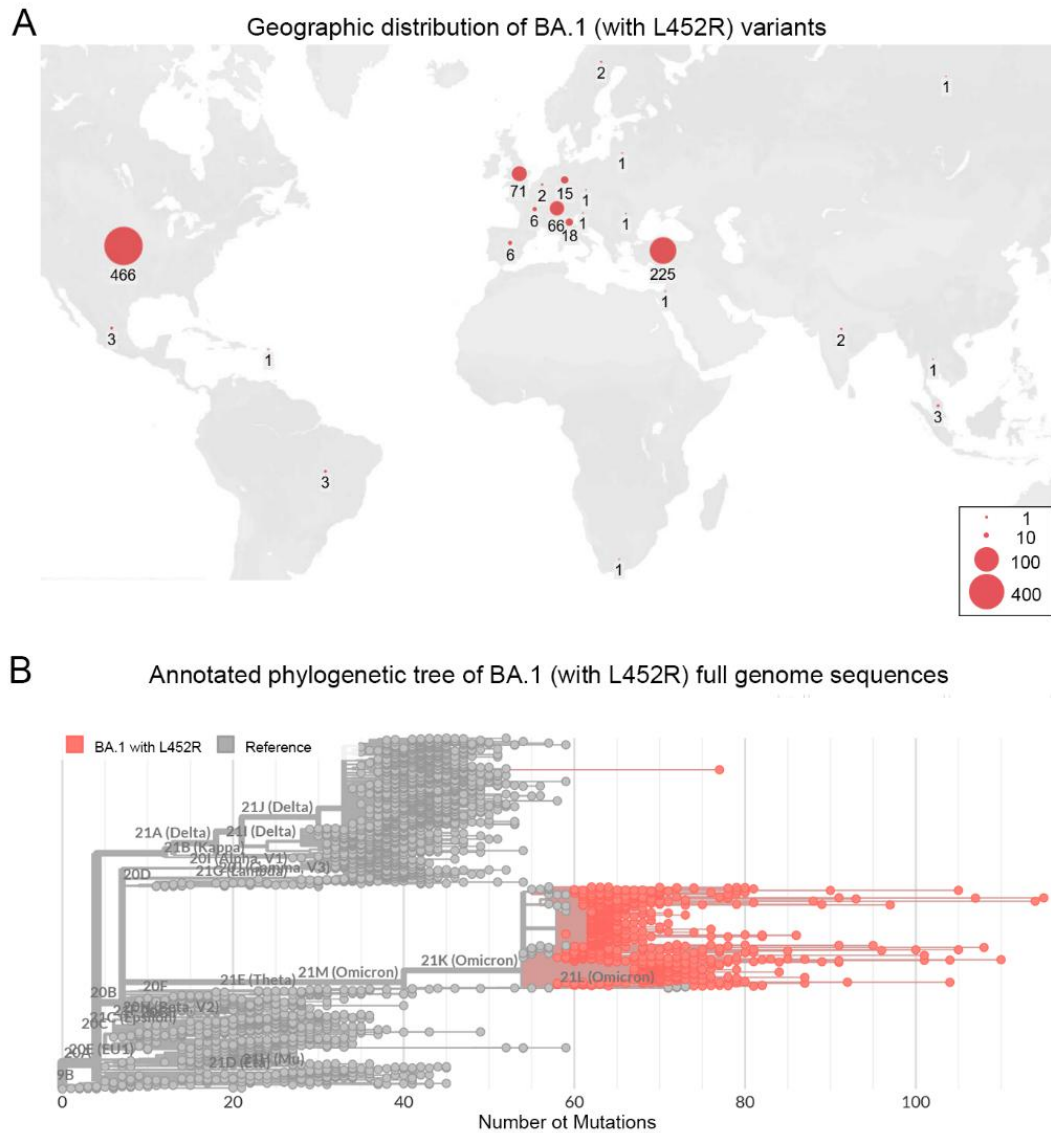
1 contributed to the emergence of the BA.1 subvariant (Table 1). Multiple mutation differences
2 causing reversion haplotypes may have originated from the recombination between the Omicron and
3 other variants (Fig. 3 and Supplementary Table 1).

4
5 Except for shared mutations, many other mutations (30 in BA.1 and 25 in BA.2) could not be
6 accounted for among previous dominant variants (Fig. 1). Because the Omicron variants are believed
7 to have emerged in South Africa¹, we speculate that some of these spike mutations may have been
8 produced by long-term virus infections in immunocompromised patients. It was previously reported
9 that evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient led to immune escape
10 variants^{45,46}. Deletions in NTD, for example, delY144, were detected in multiple immunosuppressed
11 COVID-19 patients, which resulted in immune escape⁴⁵⁻⁴⁷.

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

14 A recently reported recombinant SARS-CoV-2, “Deltacron” or “Demicron”, and its genome
15 sequences, elicited controversy and concerns of sequencing errors and sample contamination¹³.
16 Nevertheless, it was confirmed that co-infections by Omicron and Delta variants have already
17 occurred in specific populations
18 (<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
19 may lead to the emergence of new variants. A total of 10 cases of “Deltacron” are underway to
20 confirm by Santé publique France (SPF) (<https://t.co/tVAKmHRYSy>). In our study, multiple VOC
21 and VOI mutations were detected in Omicron variants circulating before January 15, 2022 (Fig. 1
22 and Table 1). The integration of these mutations may lead to changes in phenotype. Five additional
23 typical amino acid mutations in Delta variants were also identified in recently emergent Omicron
24 isolates (before January 15, 2022) (Table 2). For example, 899 Omicron sequences of high quality
25 contained L452R mutation reported for the Delta variant (Fig. 4A). Whole genome analysis also
26 corroborated the diversity among these L452R containing Omicron genomes. The mutation profiles
27

1 among whole genomes of BA.1 are diverse, and the sequences branched to diverse clades by
2 phylogenetic analyses (Fig. 4B).
3



1

2 **Figure 4. Geographic distribution and whole genome analyses of BA.1 (with L452R) variants.**

3 **(A)** Geographic distribution of BA.1 (with L452R) subvariants, with the number of genome

4 sequences noted. **(B)** Whole genome phylogenetic tree highlighting the BA.1 (with L452R)

5 subvariant, clades were annotated using NextClade (Hadfield et al., 2018). Low quality sequences

6 were excluded. 891 SARS-CoV-2 BA.1 with L452R spike amino mutation full genome sequences

7 submitted to the GISAID database before January 15th, 2022 and reference sequences from

8 SARS-CoV-2 each clade were included. The red circles are BA.1 variants.

9

1 **Conclusion**

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

10

1 Reference

- 2 1 Saxena, S. K. *et al.* Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant
3 of concern and its global perspective. *J Med Virol*, doi:10.1002/jmv.27524 (2021).
- 4 2 Li, J., Lai, S., Gao, G. F. & Shi, W. The emergence, genomic diversity and global spread of
5 SARS-CoV-2. *Nature* **600**, 408-418, doi:10.1038/s41586-021-04188-6 (2021).
- 6 3 Cui, Z. *et al.* Structural and functional characterizations of infectivity and immune evasion of
7 SARS-CoV-2 Omicron. *Cell*, doi:10.1016/j.cell.2022.01.019 (2022).
- 8 4 Salleh, M. Z., Derrick, J. P. & Deris, Z. Z. Structural Evaluation of the Spike Glycoprotein
9 Variants on SARS-CoV-2 Transmission and Immune Evasion. *Int J Mol Sci* **22**,
10 doi:10.3390/ijms22147425 (2021).
- 11 5 Deng, X. *et al.* Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2
12 variant. *Cell* **184**, 3426-3437 e3428, doi:10.1016/j.cell.2021.04.025 (2021).
- 13 6 Li, Q. *et al.* The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and
14 Antigenicity. *Cell* **182**, 1284-1294 e1289, doi:10.1016/j.cell.2020.07.012 (2020).
- 15 7 Liu, Y. *et al.* The N501Y spike substitution enhances SARS-CoV-2 infection and
16 transmission. *Nature*, doi:10.1038/s41586-021-04245-0 (2021).
- 17 8 Saito, A. *et al.* Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R
18 mutation. *Nature*, doi:10.1038/s41586-021-04266-9 (2021).
- 19 9 Ou, J. *et al.* V367F mutation in SARS-CoV-2 spike RBD emerging during the early
20 transmission phase enhances viral infectivity through increased human ACE2 receptor
21 binding affinity. *J. Virol.* **95**, e0061721, doi:doi:10.1128/JVI.00617-21 (2021).
- 22 10 Bayarri-Olmos, R. *et al.* Functional Effects of Receptor-Binding Domain Mutations of
23 SARS-CoV-2 B.1.351 and P.1 Variants. *Front Immunol* **12**, 757197,
24 doi:10.3389/fimmu.2021.757197 (2021).
- 25 11 Collier, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies.
26 *Nature* **593**, 136-141, doi:10.1038/s41586-021-03412-7 (2021).
- 27 12 Dejnirattisai, W. *et al.* Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **184**,
28 2939-2954 e2939, doi:10.1016/j.cell.2021.03.055 (2021).
- 29 13 Kreier, F. Deltacron: the story of the variant that wasn't. **602**, 19-19 (2022).
- 30 14 Dehghan, S. *et al.* A zoonotic adenoviral human pathogen emerged through genomic
31 recombination amongst human and nonhuman simian hosts. *Journal of virology* **93**,
32 e00564-00519, doi:10.1128/JVI.00564-19 (2019).
- 33 15 Walsh, M. P. *et al.* Evidence of molecular evolution driven by recombination events
34 influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis.
35 *PLoS ONE* **4**, e5635, doi:10.1371/journal.pone.0005635 (2009).

- 1 16 Walsh, M. P. *et al.* Computational analysis identifies human adenovirus type 55 as a
2 re-emergent acute respiratory disease pathogen. *J. Clin. Microbiol.* **48**, 991-993,
3 doi:10.1128/JCM.01694-09 (2010).
- 4 17 Purkayastha, A. *et al.* Genomic and bioinformatics analysis of HAdV-4, a human adenovirus
5 causing acute respiratory disease: implications for gene therapy and vaccine vector
6 development. *J. Virol.* **79**, 2559-2572, doi:10.1128/JVI.79.4.2559-2572.2005 (2005).
- 7 18 Dehghan, S. *et al.* Computational analysis of four human adenovirus type 4 genomes reveals
8 molecular evolution through two interspecies recombination events. *Virology* **443**, 197-207,
9 doi:10.1016/j.virol.2013.05.014 (2013).
- 10 19 Dehghan, S. *et al.* Simian adenovirus type 35 has a recombinant genome comprising human
11 and simian adenovirus sequences, which predicts its potential emergence as a human
12 respiratory pathogen. *Virology* **447**, 265-273, doi:10.1016/j.virol.2013.09.009 (2013).
- 13 20 Zhang, J. *et al.* A Survey of Recent Adenoviral Respiratory Pathogens in Hong Kong Reveals
14 Emergent and Recombinant Human Adenovirus Type 4 (HAdV-E4) Circulating in Civilian
15 Populations. *Viruses* **11**, 129, doi:10.3390/v11020129 (2019).
- 16 21 Chu, D. K. W. *et al.* MERS coronaviruses from camels in Africa exhibit region-dependent
17 genetic diversity. *Proc Natl Acad Sci U S A* **115**, 3144-3149, doi:10.1073/pnas.1718769115
18 (2018).
- 19 22 Hu, B. *et al.* Discovery of a rich gene pool of bat SARS-related coronaviruses provides new
20 insights into the origin of SARS coronavirus. *PLoS Pathog* **13**, e1006698,
21 doi:10.1371/journal.ppat.1006698 (2017).
- 22 23 Hon, C. C. *et al.* Evidence of the recombinant origin of a bat severe acute respiratory
23 syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS
24 coronavirus. *J Virol* **82**, 1819-1826, doi:10.1128/JVI.01926-07 (2008).
- 25 24 Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible
26 for the COVID-19 pandemic. *Nat Microbiol* **5**, 1408-1417, doi:10.1038/s41564-020-0771-4
27 (2020).
- 28 25 Guan, Y. *et al.* Isolation and characterization of viruses related to the SARS coronavirus from
29 animals in southern China. *Science* **302**, 276-278, doi:10.1126/science.1087139 (2003).
- 30 26 He, Y. *et al.* Possible recombination between two variants of concern in a COVID-19 patient.
31 *Emerg Microbes Infect*, 1-26, doi:10.1080/22221751.2022.2032375 (2022).
- 32 27 Tang, X. *et al.* Evolutionary analysis and lineage designation of SARS-CoV-2 genomes. *Sci*
33 *Bull (Beijing)* **66**, 2297-2311, doi:10.1016/j.scib.2021.02.012 (2021).
- 34 28 Yi, H. 2019 Novel Coronavirus Is Undergoing Active Recombination. *Clin Infect Dis* **71**,
35 884-887, doi:10.1093/cid/ciaa219 (2020).

- 1 29 Chen, C. *et al.* TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big
2 Biological Data. *Molecular Plant* **13**, 1194-1202,
3 doi:<https://doi.org/10.1016/j.molp.2020.06.009> (2020).
- 4 30 Zhou, Z. J., Qiu, Y., Pu, Y., Huang, X. & Ge, X. Y. BioAider: An efficient tool for viral
5 genome analysis and its application in tracing SARS-CoV-2 transmission. *Sustain Cities Soc*
6 **63**, 102466, doi:10.1016/j.scs.2020.102466 (2020).
- 7 31 Hall, T. A. BioEdit : a user-friendly biological sequence alignment editor and analysis. *Nucl*
8 *Acids Symp* **41** (1999).
- 9 32 Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn
10 diagram viewer. *BMC bioinformatics* **15**, 293, doi:10.1186/1471-2105-15-293 (2014).
- 11 33 Bardou, P., Mariette, J., Escudié, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn
12 diagram viewer. *BMC Bioinformatics* **15**, 293, doi:10.1186/1471-2105-15-293 (2014).
- 13 34 Delano, W. L. The PyMol Molecular Graphics System. *Proteins Structure Function and*
14 *Bioinformatics* **30**, 442-454 (2002).
- 15 35 Han, P. *et al.* Receptor binding and complex structures of human ACE2 to spike RBD from
16 omicron and delta SARS-CoV-2. *Cell*, doi:10.1016/j.cell.2022.01.001 (2022).
- 17 36 Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**,
18 4121-4123, doi:10.1093/bioinformatics/bty407 (2018).
- 19 37 Lyngse, F. P. *et al.* Transmission of SARS-CoV-2 Omicron VOC subvariants BA.1 and BA.2:
20 Evidence from Danish Households. *medRxiv* (2022).
- 21 38 Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev*
22 *Microbiol* **19**, 409-424, doi:10.1038/s41579-021-00573-0 (2021).
- 23 39 Huang, T. Y. & Zhao, B. tidyfst: Tidy Verbs for Fast Data Manipulation. *The Journal of*
24 *Open Source Software* **5**, 2388 (2020).
- 25 40 Katoh & K. MAFFT: a novel method for rapid multiple sequence alignment based on fast
26 Fourier transform. *Nucleic Acids Research* **30**, 3059-3066 (2002).
- 27 41 Librado & Rozas. DnaSP v5: a software for comprehensive analysis of DNA polymorphism
28 data. *Bioinformatics (Oxford, England)* **25**, 1451-1452 (2009).
- 29 42 Bandelt, H.-J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific
30 phylogenies. *Molecular Biology and Evolution* **16**, 37-48 (1999).
- 31 43 Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an
32 Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901-1912 e1909,
33 doi:10.1016/j.cell.2020.10.049 (2020).
- 34 44 Ghosn, J., Taiwo, B., Seedat, S., Autran, B. & Katlama, C. HIV. *Lancet* **392**, 685-697,
35 doi:10.1016/s0140-6736(18)31311-4 (2018).

- 1 45 Clark, S. A. *et al.* SARS-CoV-2 evolution in an immunocompromised host reveals shared
2 neutralization escape mechanisms. *Cell* **184**, 2605-2617.e2618,
3 doi:10.1016/j.cell.2021.03.027 (2021).
- 4 46 Weigang, S. *et al.* Within-host evolution of SARS-CoV-2 in an immunosuppressed
5 COVID-19 patient as a source of immune escape variants. *Nat Commun* **12**, 6405,
6 doi:10.1038/s41467-021-26602-3 (2021).
- 7 47 Chen, L. *et al.* Emergence of Multiple SARS-CoV-2 Antibody Escape Variants in an
8 Immunocompromised Host Undergoing Convalescent Plasma Treatment. *mSphere* **6**,
9 e0048021, doi:10.1128/mSphere.00480-21 (2021).

10

1 **Author contribution statement**

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

5
6 **Acknowledgments**

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

10

11 **Funding statement**

12 This work was supported by grants from the National Key Research and Development Program of
13 China (2018YFE0204503), National Natural Science Foundation of China (32170139 and
14 81730061), and Natural Science Foundation of Guangdong Province (2018B030312010 and
15 2021A1515010788).

16

17 **Competing interests**

18 The authors declare no competing interests.

19

20

21

22

23

24

25

26

27

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

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

16
17 **Figure 3. Phylogenetic network and scanning of the spike gene from representative Omicron**
18 **subvariant sequences. (A)** Representative Omicron spike protein haplotypes (each consisted of at
19 least 50 sequences) were constructed with PopART using the median-joining method⁴². Nucleotide
20 changes were notated with lines. The spike gene from Wuhan-Hu-1 strain was set as the root. The
21 number of sequences in each haplotype were modified into different orders of magnitude, and
22 subgroups based on the mutation types were delineated by color. **(B)** BootScan analysis of revertant
23 and representative haplotypes of Omicron spike gene. Representative spike Omicron haplotypes
24 (Hap_3, Hap_4, Hap_7) sequences and selected reversion haplotypes (Hap_18, Hap_39, Hap_44)
25 sequences are included. Bootscan map was constructed by Simplot 3.5.1
26 (<http://www.welch.jhu.edu/~sray/download>) using neighboring-joining method with 100 bootstrap
27 replicates. Wuhan-Hu-1 spike sequences was set as reference, reversion region was annotated. **(C)**

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

3

4 **Figure 4. Geographic distribution and whole genome analyses of BA.1 (with L452R) variants.**

5 (A) Geographic distribution of BA.1 (with L452R) subvariants, with the number of genome

6 sequences noted. (B) Whole genome phylogenetic tree highlighting the BA.1 (with L452R)

7 subvariant, clades were annotated using NextClade (Hadfield et al., 2018). Low quality sequences

8 were excluded. 891 SARS-CoV-2 BA.1 with L452R spike amino mutation full genome sequences

9 submitted to the GISAID database before January 15th, 2022 and reference sequences from

10 SARS-CoV-2 each clade were included. The red circles are BA.1 variants.

11

12

13 **Table 1. Comparison of Spike protein amino acid mutations between the Omicron subvariants**

14 **and other VOCs and VOIs.** 52,563 high quality Omicron spike gene sequences (49,609 BA.1

15 sequences, and 2,954 BA.2 sequences) released before January 15, 2022 were analyzed. The

16 mutations that have appeared in more than 800 sequences were used in this analysis. VOCs are

17 variants of concern; VOIs are variants of interest.

18

19 **Table 2. Novel mutations identified in the spike protein of the recently emerged Omicron**

20 **subvariants (Released before January 15, 2022; frequency >50 sequences).**

21

22 **Supplementary information**

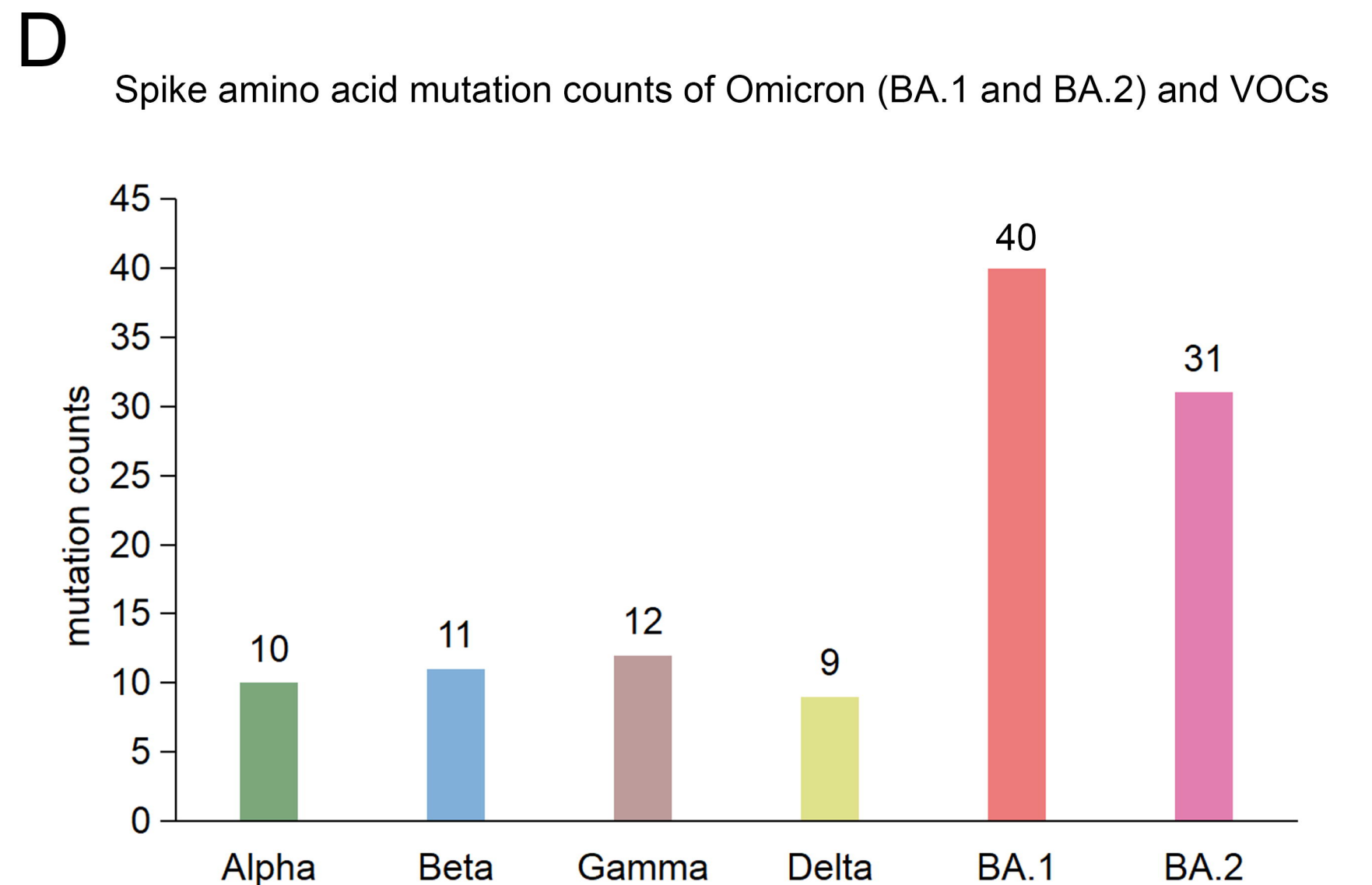
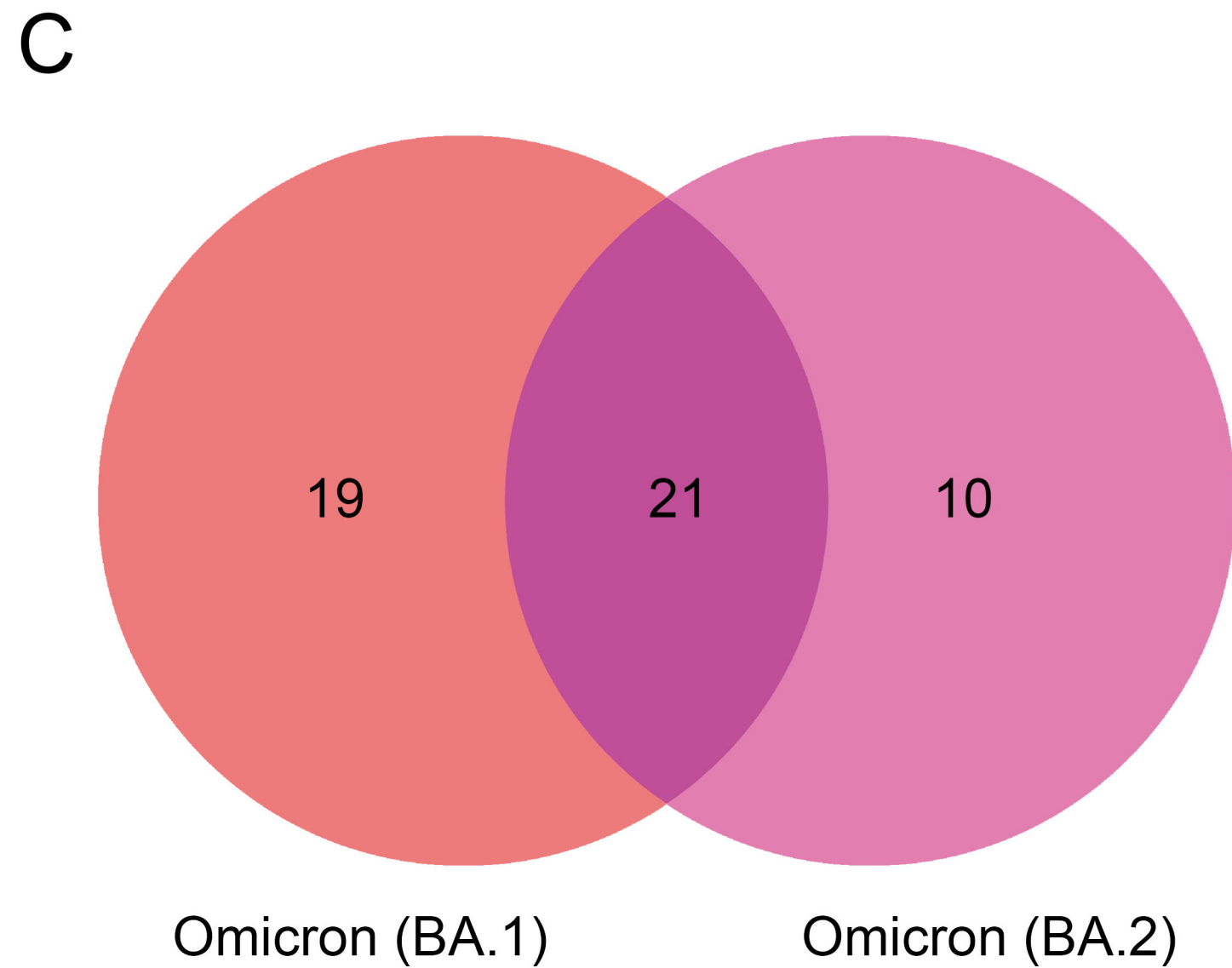
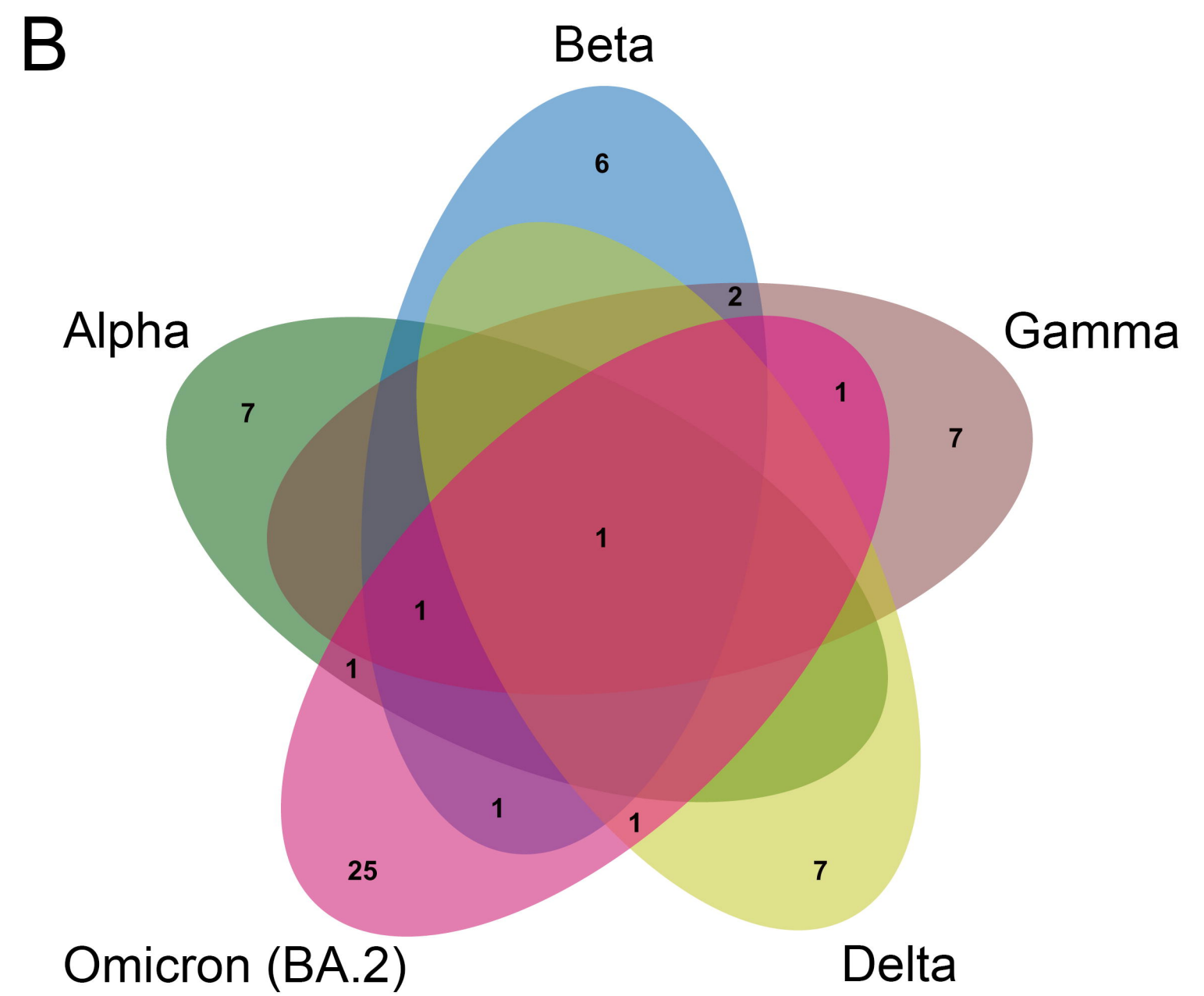
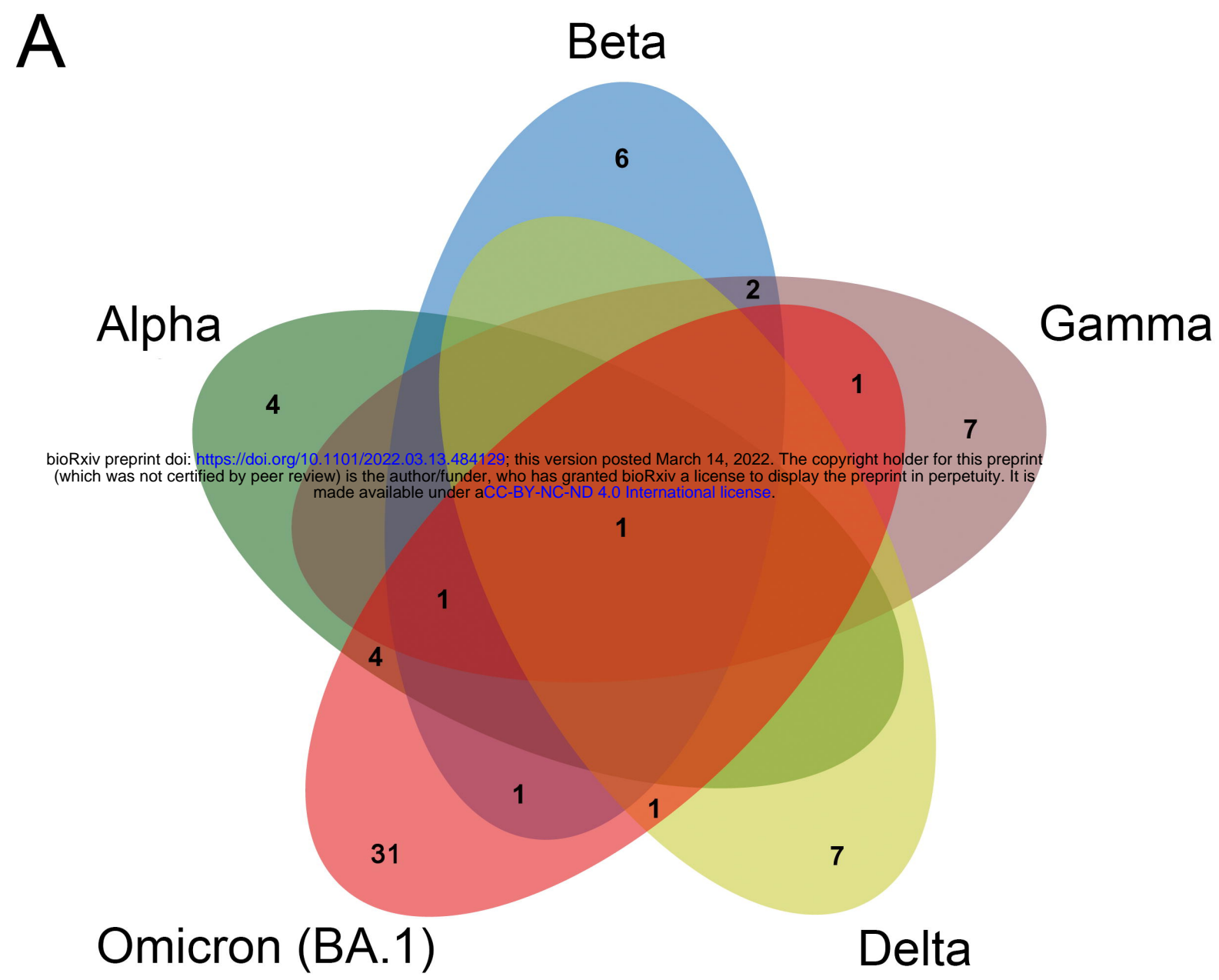
23 **Supplementary Table 1.** The annotation of haplotypes of Omicron spike protein.

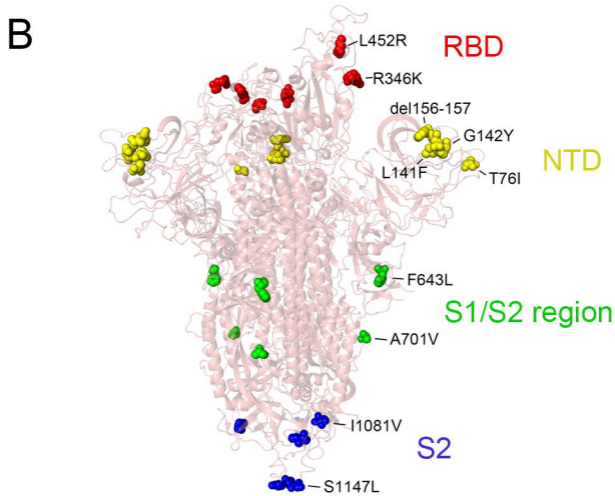
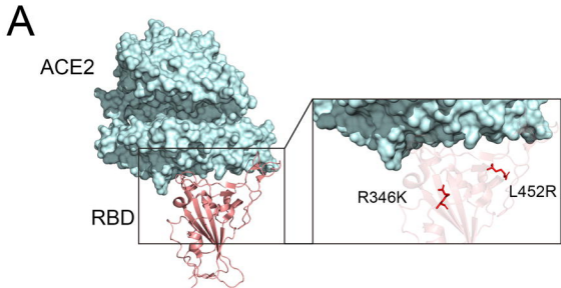
24

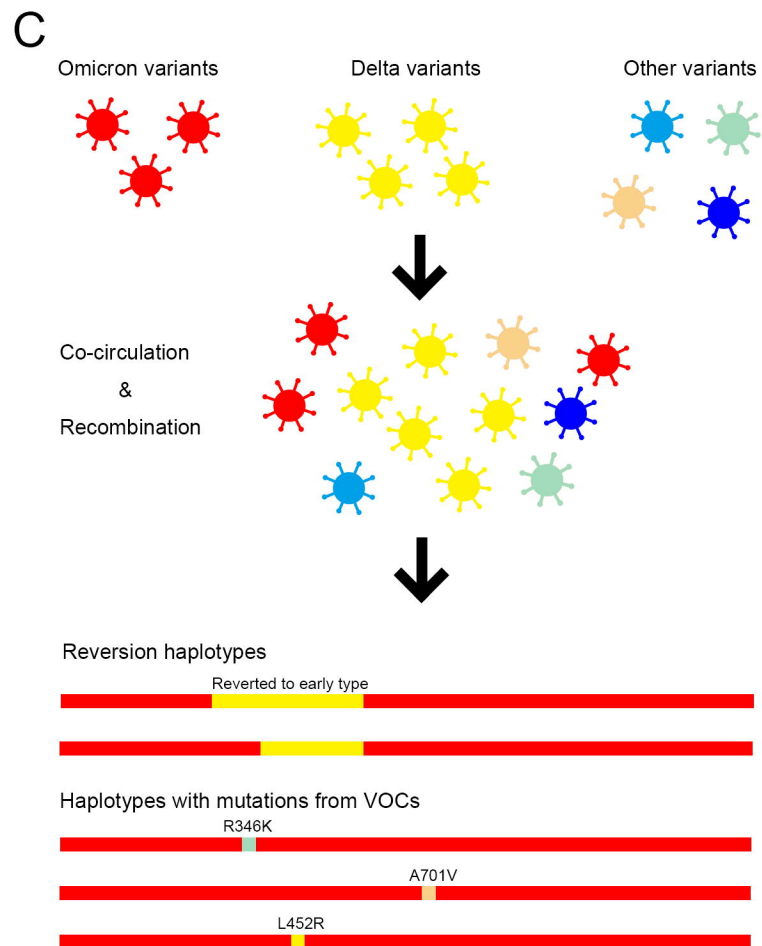
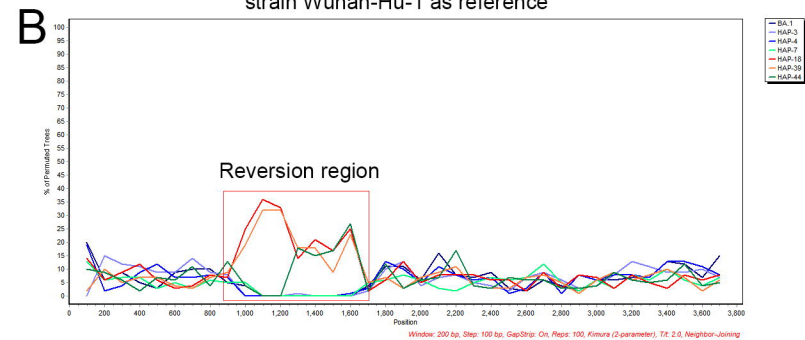
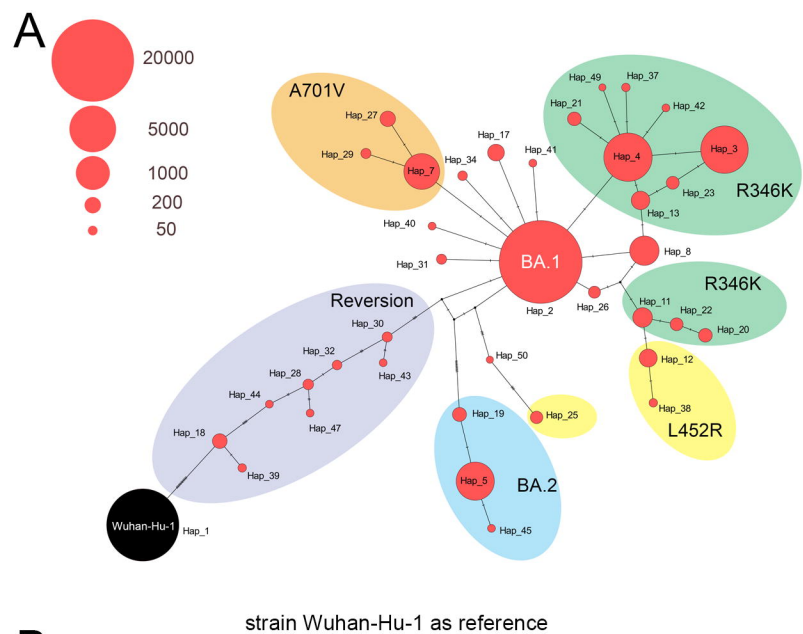
25

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

27







A Geographic distribution of BA.1 (with L452R) variants



B Annotated phylogenetic tree of BA.1 (with L452R) full genome sequences

