Anomalies in the regional distribution and mutation spectrum of the SARS-CoV-2 Omicron BA.2.86 lineage

Hideki Kakeya

Institute of Systems and Information Engineering, University of Tsukuba,

1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

ORCID: 0000-0003-3788-9133

Corresponding author

Hideki Kakeya

1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan

kake@iit.tsukuba.ac.jp

Keywords

SARS-CoV-2, Omicron variant, BA.2 lineage, epidemiology, mutation spectrum

Abstract

SARS-CoV-2 remains globally prevalent, despite reduced virulence, causing a range of mild to severe diseases. Since mid-2023, major variants such as JN.1, KP.3.1.1, and XEC have emerged as predominant strains, all of which are descendants of BA.2.86. BA.2.86 exhibited substantial divergence from earlier strains, with around 30 mutations in the spike protein alone compared to Omicron BA.2, a divergence as striking as the initial emergence of Omicron BA.1. In this study, we focus on BA.2.86 to understand how this striking variant emerged. Epidemiological data from early global collections of BA.2.86 and BA.2.86.1 variants show a worldwide, simultaneous emergence, without clear epicenters, suggesting a mechanism beyond natural human-to-human community transmission. Molecular analysis of mutation spectra reveals a divergence from human SARS-CoV-2, pointing toward evolution in a non-human host, possibly through experiments involving animal models.

Introduction

While SARS-CoV-2 attracts less public attention due to its decreased virulence after subsequent mutations, the virus itself is still prevalent around the world, causing both mild and sometimes severe diseases. Since mid-2023, the major variants have been close descendants of BA.2.86, with JN.1, KP.3.1.1, and XEC prevailing from mid-2023 to the end of 2024. These variants are one, four, and six mutations away from BA.2.86 in the spike protein, respectively. What was striking was the emergence of BA.2.86, which is approximately 30 mutations away from BA.2 in the spike protein, representing a significant deviation similar to the emergence of the first Omicron variant (BA.1).

Various hypotheses have been proposed to explain the massive mutations that appeared in the Omicron variant. Two primary theories suggest either long-term evolution within an immunocompromised individual or mutation in a non-human host before spilling over back into humans [1]. However, studies have shown that mutations in immunocompromised patients typically number around 10 or fewer [2-4], which is far fewer than the mutations observed in Omicron. As for the theory of evolution in a non-human host, some researchers, like Wei et al. [5] and Zhang et al. [6], have suggested that Omicron could have evolved in mice based on its mutation spectrum and the structure of receptor binding domain. However, the original SARS-CoV-2 strain does not infect mice [7]. Kakeya et al. proposed a possible lab origin, potentially due to a spillover from transgenic mice [8].

While there has been extensive research on the emergence of the first Omicron variant, fewer studies have focused on the origins of BA.2.86, which represents a mutation shift as significant as the emergence of BA.1.

In understanding the evolution of SARS-CoV-2 and its variants, epidemiological studies are informative for detecting their source. Regarding the original Wuhan strain, Pekar et al. [9] and Worobey et al. [10] claimed that SARS-CoV-2 had originated in the Huanan Seafood Market, based on early mutations and sampling locations at the onset of the outbreak. However, allegations of sampling bias were raised against them in subsequent publications [11, 12]. Additionally, it was found that the ACE2 receptors of raccoon dogs, which were considered intermediate hosts in the market origin theory, do not bind strongly to the spike protein of SARS-CoV-2 [13]. For these reasons, it is now almost impossible to assert that the Huanan Seafood Market was the origin of SARS-CoV-2.

The transmission dynamics of the early variant were studied extensively in the UK, revealing the presence of central transmission hubs [14]. The first major mutation in SARS-CoV-2 was D614G in the surface glycoprotein (spike protein), which was traced in the UK (United Kingdom) [15] and in other parts of the world [16]. After D614G, several other mutations followed, leading to the emergence of major variants like B.1.1.7 (Alpha, UK), and B.1.617.2 (Delta, India), which dominated from late 2020 to mid-2021. Early mutations in SARS-CoV-2 showed high diversity, with independent mutations appearing globally, particularly in the spike protein [17].

Among the major variants, Omicron, which emerged in November 2021, stood out with over 30 non-synonymous mutations in its spike protein, compared to just 10 or so in previous variants. Phylogenetic analysis indicated that Omicron did not evolve from earlier variants, suggesting a distinct evolutionary path

[18].

Numerous studies have tracked the spread of Omicron variants globally. For instance, studies from Chile [19] and Taiwan [20] focused on the chronological spread of the variant, while others, like those in Brazil, examined both geographical and chronological patterns. One study showed that the Omicron variant was primarily introduced through São Paulo, from where it spread to other states and regions across Brazil [21]. Another study investigated the spatiotemporal spread of emerging SARS-CoV-2 lineages, including Omicron variants, in Sergipe State in northeastern Brazil, revealing that densely populated cities served as epicenters for each lineage's circulation, regardless of vaccination status [22]. Additionally, a separate study conducted genomic surveillance of SARS-CoV-2 in the southeastern region of São Paulo State to track the regional emergence of the BA.2 lineage during its initial spread [23].

In Mexico, the Omicron BA.1 surge at the end of 2021 originated in the eastern region of the country and spread westward [24]. In the UK, it was confirmed that Omicron BA.1 spread from London [25], a trend also seen with Omicron BA.2 [26]. In Hong Kong, the spatiotemporal patterns of Omicron BA.1, BA.2, and Delta AY.127 were examined using phylogenetic trees and transmission modeling [27]. In South Africa, Omicron BA.4 and BA.5 were found to have spread from Johannesburg and Durban, respectively [28]. In Cyprus, transmission patterns of Omicron BA.1, BA.2, and BA.5 between Cyprus and other countries were tracked [29]. Together, these studies demonstrate that the spread of emerging variants is generally traceable.

In the United States, which has the most comprehensive SARS-CoV-2 genome data collection, one study analyzed the attack rates of different Omicron variants across states, although it did not detail the

chronological progression [30]. Recently, the author investigated the spread of the Omicron BA.1 and BA.1.1 lineages and discovered that reversion mutants were widespread from the beginning, with no clear epicenters identified for most of these mutants [31, 32]. These anomalous emergences of mutations cannot be explained by any biological theories of natural evolution known to date.

In this study, the author investigates how the BA.2.86 lineage, which is as distant from the preceding variants as the BA.1 lineage was, emerged from both epidemiological and molecular points of view. From the epidemiological perspective, the collection locations of early entries of BA.2.86 in GISAID are surveyed and compared with other variants. From the molecular perspective, the mutation spectrum of BA.2.86 is calculated and analyzed.

Methods

The locations and collection dates of BA.2.86 and BA.2.86.1 registered in GISAID were downloaded in August 2024. The locations of early sample collections of BA.2.86 and BA.2.86.1 were mapped globally to observe the chronological and geographical emergence patterns.

For comparison, 11 variants (B.1.351, B.1.525, B.1.526, B.1.617.1, C.37, B.1.621, B.2.75, XBB.1.16, XBB.1.16.16, JG.3, and XBB.2.3) registered in GISAID were downloaded in January 2024. Among the 11 variants, the first six are major variants of concern (Beta, Eta, Iota, Kappa, Lambda, Mu) with a moderate number of entries. The remaining five variants were chosen from those with similar numbers of entries that appeared from 2022 to 2023. The collection locations of the first 50 samples of BA.2.86.1 and the other 11 variants were compared.

The nucleotide sequences of the Omicron variants BA.1, BA.2, and BA.2.86.1 were downloaded from GenBank in August 2024. Wuhan-Hu-1 (NC_045512.2) was used as the consensus sequence of the original Wuhan strain. The consensus sequences of BA.1 and BA.2 were calculated based on the first 10,000 sequences registered in the NCBI (National Center for Biotechnology Information) GenBank, identifying the most frequent sequence. The consensus sequence of BA.2.86.1 was calculated based on the first 200 sequences without missing reads registered in the NCBI GenBank.

Mutation spectra of the whole genome and the spike genome from BA.2, the closest ancestor of BA.2.86.1, to BA.2.86.1 were calculated based on the consensus sequences. Mutation spectra from Wuhan-Hu-1 to BA.1 were also calculated for reference. To find the mutation spectrum, dynamic programming (DP) matching using Levenshtein distance with +1 for matches, -1 for mismatches, and -2 for deletions and insertions was applied to find the best matching alignment.

Results

The locations of early sample collections of BA.2.86 and BA.2.86.1 from July 2023 to August 2023 are mapped in Figure 1. As this figure shows, BA.2.86 and BA.2.86.1 infections were observed worldwide from the onset, and no epicenters for these variants are detectable.

The locations of the first 50 collected samples of 12 variants (B.1.351, B.1.525, B.1.526, B.1.617.1, C.37, B.1.621, B.2.75, XBB.1.16, XBB.1.16.16, JG.3, BA.2.86.1, and XBB.2.3) are listed in Table 1. As shown in this table, variants other than BA.2.86.1 have epicenters, with collection sites concentrated in specific areas.

The divergence in the collection locations of BA.2.86.1 is statistically significant both in terms of the number of areas where the first 50 samples were collected ($p = 0.017$) and the low occupancy rate of the most sampled area ($p = 0.014$), assuming a normal distribution.

Mutation spectra of the whole genome and the spike genome from the original Wuhan strain to BA.1 and from BA.2 to BA.2.86.1 are shown in Figure 2. The mutation spectrum of human SARS-CoV-2 [33] is also shown for reference. Previous studies have shown that the mutation spectrum from the original Wuhan strain to BA.1 differs significantly from that of human SARS-CoV-2 and is more similar to that of mice [5]. The mutation spectra of the whole genome and spike genome from BA.2 to BA.2.86.1 also differ significantly from that of human SARS-CoV-2, with *p*-values of 0.0073 (whole genome) and 0.010 (spike) based on the G-test.

Discussion

As Figure 1 shows, BA.2.86 and BA.2.86.1, which are quite distant variants from the previous ones, suddenly emerged simultaneously across the globe. Since it is unlikely that similar viruses with the same 30 spike mutations appeared independently from multiple sources, a mechanism to rapidly spread the same new strain worldwide must have existed. BA.2.86 and BA.2.86.1 are not infectious enough to spread as quickly via human-to-human transmission as Omicron BA.1 did at the onset of its emergence, as indicated by the small total number of entries in GISAID. An additional L455S mutation was needed for the new strain (JN.1) to infect a large population globally. Therefore, natural community infection cannot explain the rapid spread of BA.2.86 and BA.2.86.1 across the world.

Based on the mutation spectrum of BA.2.86 from BA.2, it is most likely that the evolution of BA.2.86 occurred in an animal host other than humans. Immune escape in a heavily vaccinated human population or in immunocompromised human patients cannot explain the deviation in the mutation spectrum from that of humans. Evolution in independent non-human hosts worldwide is also implausible, as it is unlikely that the same 30 mutations occurred simultaneously in multiple locations by coincidence.

One possible theory to explain the emergence of a completely new variant, deviating from human evolution, is a lab-grown virus in an animal host. If a virus strain was grown in animal cell cultures or lab animals were infected with a new strain of the virus, either deliberately or accidentally, and then sent to collaborating laboratories worldwide for experimental purposes, it could lead to the global emergence of the same new virus strain that is completely different from previous variants, with a non-human mutation spectrum.

Unfortunately, activities in virology laboratories have been quite opaque both before and during the COVID-19 pandemic. Many research accidents have been repeatedly covered up [34]. To clarify what happened, not only in the genesis of BA.2.86 and BA.2.86.1 but also in the emergence of other SARS-CoV-2 variants, some of which have been indicated as unnatural [35-37], thorough investigations into all virology laboratories that have dealt with SARS-CoV-2 are necessary. If any of the variants are found to have originated in a laboratory, the researchers responsible for their release should be held accountable and prosecuted.

Data Availability

The source code used for this study is available at

https://visual-media-lab.github.io/data/Mutation_Spectrum/index.html

Conflicts of interest

The author declares no conflict of interest exists.

References

[1] C. Mallapaty, The hunt for the origin of Omicron. Nature 602, 26-28 (2022).
DOI: 10.1038/d41586-022-00215-2

[2] B. Choi, M. C. Choudhary, J. Regan, et al., persistence and evolution of SARS-CoV-2 in an immunocompromised host. The New England Journal of Medicine 383, 2291-2293 (2020).
DOI: 10.1056/NEJMc2031364

[3] S. A. Kemp, D. A. Collier, R. P. Datier, et al., SARS-CoV-2 evolution during treatment of chronic infection. Nature 592, 277-282 (2021).
DOI: 10.1038/s41586-021-03291-y

[4] T. T. Truong, A. Ryutov, U. Pandey, et al., Increased viral variants in children and young adults with impaired humoral immunity and persistent SARS-CoV-2 infection: A consecutive case series. EBioMedicine May;67:103355 (2021).
DOI: 10.1016/j.ebiom.2021.103355

[5] C. Wei, K. J. Shan, W. Wang, et al., Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. J Genet Genomics 48, 1111-1121 (2021).
DOI: 10.1016/j.jgg.2021.12.003

[6] W. Zhang, K. Shi, Q. Geng, et al., Structural basis for mouse receptor recognition by SARS-CoV-2 omicron variant. PNAS 119, e2206509119 (2022).

DOI: 10.1073/pnas.2206509119

[7] S. Piplani, P. K. Singh, D. A. Winkler, et al., In silico comparison of SARS-CoV-2 spike protein-ACE2 binding affinities across species and implications for virus origin. Scientific Reports 11, 13063 (2021).

DOI: 10.1038/s41598-021-92388-5

[8] H. Kakeya, H. Arakawa, Y. Matsumoto. Multiple probabilistic analyses suggest non-natural origin of SARS-CoV-2 Omicron variant. Zenodo (2023).

DOI: 10.5281/zenodo.7470652

[9] J. E. Pekar, A. Magee, E. Parker, et al., The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science 377, 960-966 (2022).

DOI:10.1126/science.abp8337

[10] M. Worobey, J. Levy, L. M. Serrano, et al., The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science 377, 951-959 (2022).

DOI:10.1126/science.abp8715

[11] S. E. Massey, A. Jones, D. Zhang, et al. Unwarranted exclusion of intermediate lineage A-B SARS-CoV-2 genomes is inconsistent with the two-spillover hypothesis of the origin of COVID-19. Microbiology Research 14, 448-453 (2023).

DOI: 10.3390/microbiolres14010033

[12] Z. Daoyu, D. Gilles, J. Adrian, et al., Zoonosis at the Huanan Seafood Market: a critique. Zenodo (2022).

DOI: 10.5281/zenodo.7169296

[13] C. Luo, L. Li, Y. Gu, et al., Receptor binding and structural basis of raccoon dog ACE2 binding to SARS-CoV-2 prototype and its variants. PLoS Pathogen 20(12): e1012713 (2024).
DOI: 10.1371/journal.ppat.1012713

[14] L. Du Plessis, J. T. McCrone, A. E. Zarebski, et al., Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science 371, 708-712 (2021).
DOI: 10.1126/science.abf2946

[15] E. Volz, V. Hill, J. T. McCrone, et al., Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell 184, 64-75 (2021).
DOI: 10.1016/j.cell.2020.11.020

[16] B. Korber, W. M. Fischer, S. Gnanakaran, et al., Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812-827 (2020).
DOI: 10.1016/j.cell.2020.06.043

[17] S. S. Hassan, V. Kodakandla, E. M. Redwan, et al., Non-uniform aspects of the SARS-CoV-2 intraspecies evolution reopen question of its origin. International Journal of Biological Macromolecules 222, 972-993 (2022).
DOI: 10.1016/j.ijbiomac.2022.09.184

[18] M. Nikolaidis, A. Papakyriakou, K. Chlichlia, et al., Comparative analysis of SARS-CoV-2 variants of concern, including Omicron, highlights their common and distinctive amino acid substitution patterns, especially at the spike ORF. Viruses 14, 707 (2022).
DOI: 10.3390/v14040707

[19] A. Mella-Torres, A. Escobar, C. Barrera-Avalos, et al., Epidemiological characteristics of Omicron and Delta SARS-CoV-2 variant infection in Santiago, Chile. Front Public of Health 10, 98443 (2022).

DOI:10.3389/fpubh.2022.984433

[20] L. Liu, S. Chiou, P. Chen, et al., Epidemiology and analysis of SARS-CoV-2 Omicron subvariants BA.1 and 2 in Taiwan. Scientific Reports 13, 16583 (2023).

DOI: 10.1038/s41598-023-43357-7

[21] A. P. Lamarca, U. Souza, U.J.B.d., Moreira, F.R.R., et al., The Omicron Lineages BA.1 andBA.2 (Betacoronavirus SARS-CoV-2) Have Repeatedly Entered Brazil through a Single Dispersal Hub. Viruses 15, 888 (2023).

DOI:10.3390/v15040888

[22] M. T. S. Freitas, L. O. C. Sena, K. F. Fukutani, et al., The increase in SARS-CoV-2 lineages during 2020–2022 in a state in the Brazilian Northeast is associated with a number of cases. Front Public of Health (2023).

DOI:10.3389/fpubh.2023.1222152

[23] E. S. Rodrigues, S. N. Slavov, D. G. L. de La Roque, et al., Epidemiology of the SARS-CoV-2 Omicron Variant Emergence in the Southeast Brazilian Population. Microorganisms 12, 449 (2024).

DOI:10.3390/microorganisms1230449

[24] S. Zárate, B. Taboada, M. Rosales-Rivera, et al., Omicron-BA.1 Dispersion Rates in Mexico Varied According to the Regional Epidemic Patterns and the Diversity of Local Delta Subvariants. Viruses 15, 243 (2023).

DOI:10.3390/v15010243

[25] J. Tsui, J. T. McCrone, B. Lambert, et al., Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1. Science 381, 336–343 (2023).

DOI: 10.1126/science.adg6605

[26] P. Elliott, O. Eales, N. Steyn, et al., Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England. Science 376, 1432 (2022).
DOI: 10.1126/science.abq4411

[27] C. Tong, W. Shi, G. H. Siu, et al., Understanding spatiotemporal symptom onset risk of Omicron BA.1, BA.2 and hamster-related Delta AY.127. Frontiers in Public Health 10, 978052 (2022).
DOI: 10.3389/fpubh.2022.978052

[28] H. Tegally, M. Moir, J. Everatt, et al., Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. Nature Medicine 28, 1785-1790 (2022).
DOI: 10.1038/s41591-022-01911-2.

[29] A. C. Chrysostomou, B. Vrancken, C. Haralambous, et al., Unraveling the Dynamics of Omicron (BA.1, BA.2, and BA.5) Waves and Emergence of the Deltacron Variant: Genomic Epidemiology of the SARS-CoV-2 Epidemic in Cyprus (Oct 2021–Oct 2022). Viruses 15, 1933 (2023).
DOI: 10.3390/v15091933

[30] L. Lopes, K. Pham, F. Klaassen, et al., Combining genomic data and infection estimates to characterize the complex dynamics of SARS-CoV-2 Omicron variants in the US. Cell Reports 43, 114451 (2024)
DOI:10.1016/j.celrep.2024.114451

[31] H. Kakeya. Anomalous US-wide prevalence of reversion mutants in the emergence of Omicron BA.1. Research Square (2024).
DOI: 10.21203/rs.3.rs-4919461/v1

[32] H. Kakeya. Anomalies in regional and chronological distributions of SARS-CoV-2 Omicron BA.1.1 lineage in the United States. medRxiv (2024).
DOI: 10.1101/2024.08.14.24311991

[33] K. J. Shan, C. Wei, Y. Wang, et al. Host-specific asymmetric accumulation of mutation types reveals that the origin of SARS-CoV-2 is consistent with a natural process. Innovation 2(4), 100159 (2021). DOI: 10.1016/j.xinn.2021.100159

[34] A. Young. Pandora's gamble: lab leaks, pandemics, and a world at risk. Center Street (2023).

[35] H. Kakeya and Y. Matsumoto. A probabilistic approach to evaluate the likelihood of artificial genetic modification and its application to SARS-CoV-2 Omicron variant. ISPJ Transactions on Bioinformatics, 15, 22–29 (2022). DOI: 10.2197/ipsjtbio.15.22

[36] A. Tanaka and T. Miyazawa. Unnaturalness in the evolution process of the SARS-CoV-2 variants and the possibility of deliberate natural selection. Zenodo (2023). DOI: 10.5281/zenodo.8361577

[37] H. Arakawa. The natural evolution of RNA viruses provides important clues about the origin of SARS-CoV-2 variants. SynBio, 2(3), 285–297 (2024). DOI: 10.3390/synbio2030017

Table 1. Locations of the first 50+ sample collections (the number exceeds 50 when multiple samples are collected on the same day). The regions of the world are divided into Africa, the Middle East, Central and South Asia, East Asia, Europe, North America, South America, and Oceania. The top area rate shows the occupancy rate of the most sampled region, divided by the total number of samples.

			Locations of first 50+ sample collections										
Pongo Lineage	N (~2023)	VOC	Africa	Mid East	C&S Asia	E Asia	Europe	N Amer.	S Amer.	Oceania	# area	First 50+ collection dates	Top area rate
B.1.351	39506	Beta	51	1	1			1			**4**	5/6-9/4,2020	**0.94**
B.1.525	10018	Eta	39	1			4	7			**4**	12/7-1/3,2020-21	**0.76**
B.1.526	45236	Iota						55			**1**	11/3-12/27,2020	**1.00**
B.1.617.1	5195	Kappa			50						**1**	12/2-2/3,2021	**1.00**
C.37	10428	Lambda						1	52		**2**	12/11-1/20,2021	**0.98**
B.1.621	14364	Mu					4	5	44		**3**	12/15-3/15,2021	**0.83**
BA.2.75	5999	BA2			50	3	2	4		1	**5**	5/26-6/13,2022	**0.83**
XBB.1.16	37145	XBB			43		2	7	1		**4**	11/23-2/13,2022-23	**0.81**
XBB.1.16.15	4537	XBB				1	43		7		**3**	5/15-6/30.2023	**0.84**
JG.3	8749	XBB					53	12			**2**	6/26-8/28,2023	**0.82**
BA.2.86.1	4952	BA2	2	1	1	2	30	14		1	**7**	7/2-8/29,2023	**0.59**
XBB.2.3	8503	XBB			39			8	2	1	**4**	12/9-1/12,2023-4	**0.78**

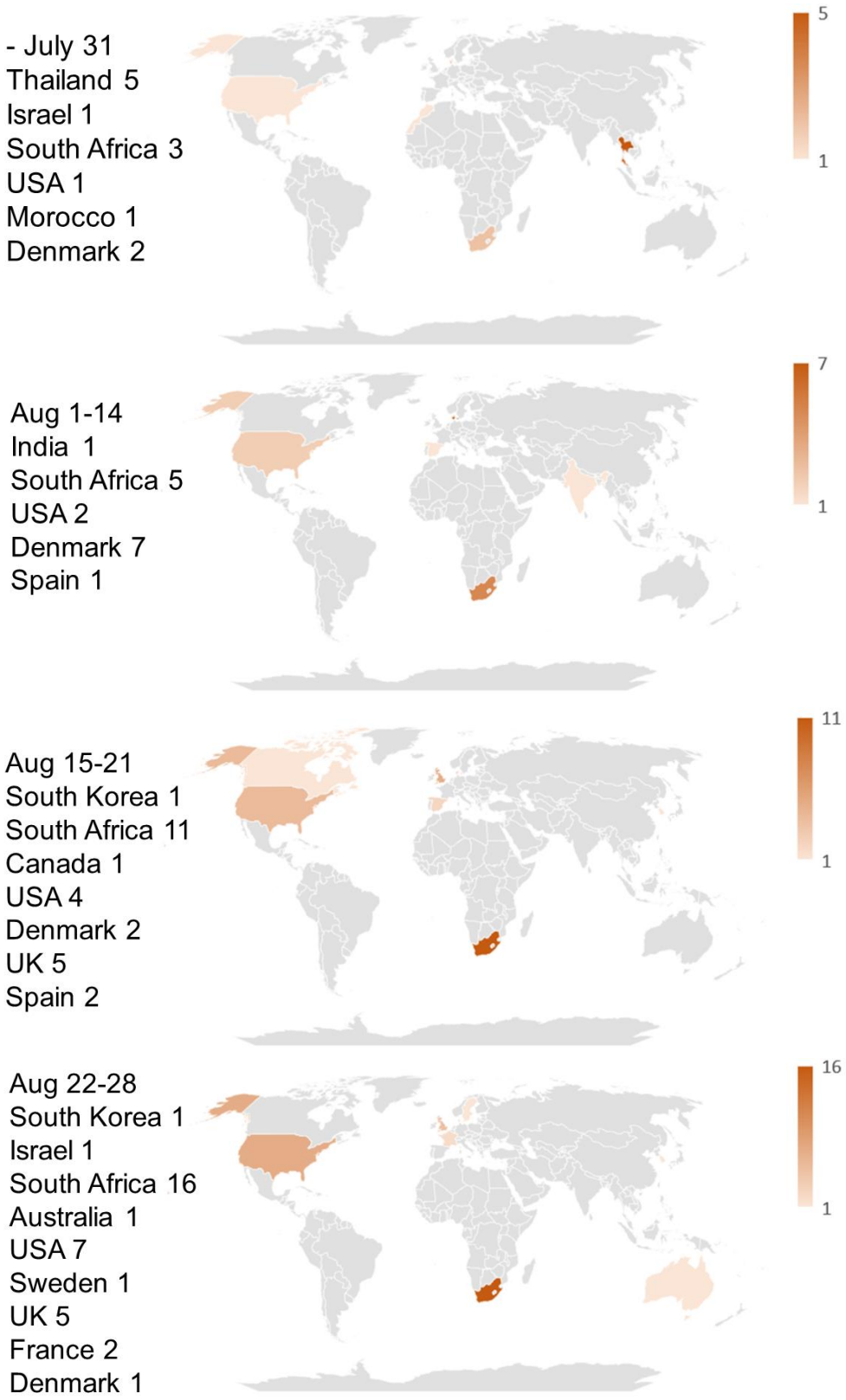
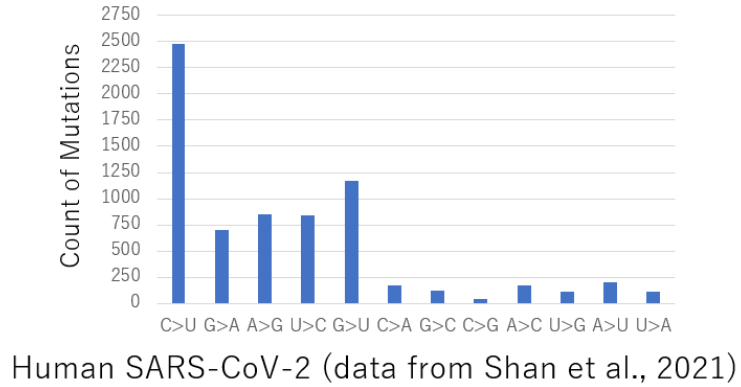


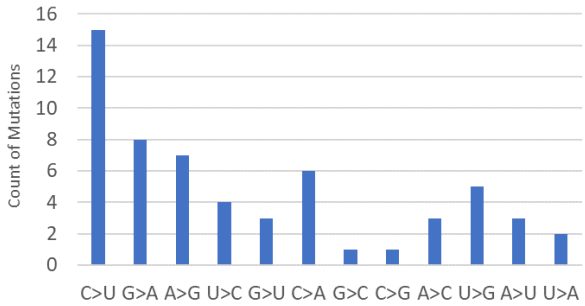
Figure 1. The heatmaps of the areas where BA.2.86 and BA.2.86.1 were sampled in July and August 2023.

A

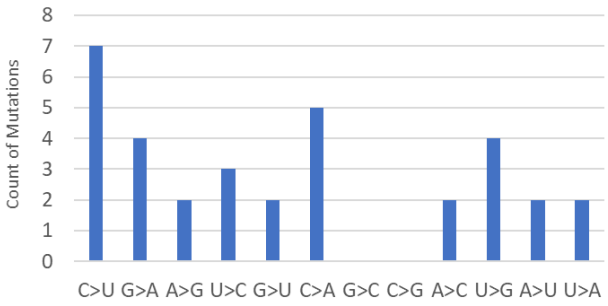


Human SARS-CoV-2 (data from Shan et al., 2021)

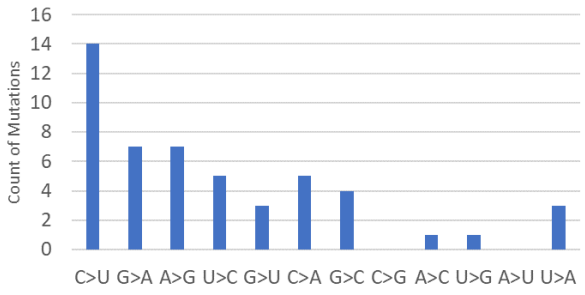
B



C



D



E

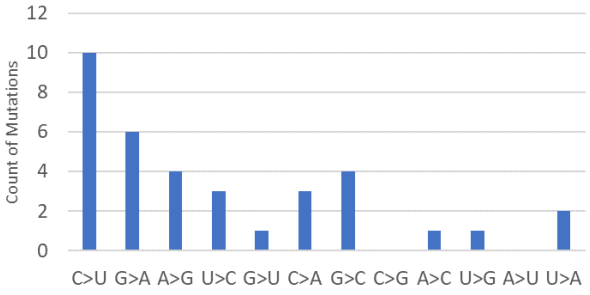


Figure 2. Mutation spectra of SARS-CoV-2 and its variants: (A) Mutation spectrum of human SARS-CoV-2 [35]; (B) Mutation spectrum of the whole genome from the Wuhan strain to Omicron BA.1; (C) Spike mutation spectrum from the Wuhan strain to Omicron BA.1; (D) Mutation spectrum of the whole genome from BA2 to BA.2.86.1; (E) Spike mutation spectrum from BA.2 to BA.2.86.1.