Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biosafety and Health 4 (2022) 171–178

Contents lists available at ScienceDirect

Biosafety and Health

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

Genetic differentiation and diversity of SARS-CoV-2 Omicron variant in its early outbreak

Shenghui Weng [a,b,1], Jingzhe Shang [a,b,1], Yexiao Cheng [a,b,1], Hangyu Zhou [a,b], Chengyang Ji [a,b], Rong Yang [a,b], Aiping Wu [a,b,*]

[a] Institute of Systems Medicine, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China
[b] Suzhou Institute of Systems Medicine, Suzhou 215123, China

ARTICLE INFO

Article history:
Received 16 February 2022
Revised 17 April 2022
Accepted 22 April 2022
Available online 25 April 2022

Keywords:
SARS-CoV-2
Omicron
Genetic diversity

ABSTRACT

The recently emerged Omicron variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has quickly spread around the world. Although many consensus mutations of the Omicron variant have been recognized, little is known about its genetic variation during its transmission in the population. Here, we comprehensively analyzed the genetic differentiation and diversity of the Omicron variant during its early outbreak. We found that Omicron achieved more structural variations, especially deletions, on the SARS-CoV-2 genome than the other four variants of concern (Alpha, Beta, Gamma, and Delta) in the same timescale. In addition, the Omicron variant acquired, except for 50 consensus mutations, seven great new non-synonymous nucleotide substitutions during its spread. Three of them are on the S protein, including S_A701V, S_L1081V, and S_R346K, which belong to the receptor-binding domain (RBD). The Omicron BA.1 branch could be divided into five divergent groups spreading across different countries and regions based on these seven novel mutations. Furthermore, we found that the Omicron variant possesses more mutations related to a faster transmission rate than the other SARS-CoV-2 variants by assessing the relationship between the genetic diversity and transmission rate. The findings indicated that more attention should be paid to the significant genetic differentiation and diversity of the Omicron variant for better disease prevention and control.

1. Introduction

Since the emergence of the coronavirus disease 2019 (COVID-19) was firstly reported in December 2019, the frequent emerging events of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants have raised significant concerns [1]. To prioritize the monitoring of noteworthy SARS-CoV-2 variants, the World Health Organization (WHO) divided highlighted variants into three categories: variants of concern (VOCs), variants of interest (VOIs), and variants under monitoring (VUMs). Previously, four VOCs were highlighted, including Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), and Delta (B.1.617.2) [2]. On November 19, 2021, a new variant was detected in S-gene target failure (SGTF) samples in South America that was were genetically distinct from all previous SARS-CoV-2 strains [3]. On November 24, 2021, this SARS-CoV-2 variant was defined as a new PANGO lineage (B.1.1.529), and two days later, this branch was classified as a VOC by the WHO and named the Omicron variant. As of February 16, 2022, the Omicron variant had spread to at least 140 countries and regions, leading to another COVID-19 case spike.

Omicron is currently the variant of concern with the most mutations, carrying 50 characteristic mutations, 31 of which are on the S protein [4]. Some of these characteristic mutations are also present in other variants [5], while others are unique mutations, such as S_G339D, S_S375F, S_G446S, and S_Q498R [6]. In the first month of Omicron circulating in the human population, it was split into four lineages, including B.1.1.529, BA.1, BA.2, and BA.3. At the beginning of the Omicron variant's spread, BA.1 was the dominant lineage. However, recently, in Denmark, the amount of sequenced BA.2 genomes increased rapidly and the BA.2 lineage has become the dominant strain [7]. These phenomena indicated that the Omicron variant can quickly evolve and differentiate at a high transmission speed.

Furthermore, some studies have confirmed that the Omicron variant has a significantly greater immune escape capability than the SARS-CoV-2 strain reported in 2019. Vaccinated people and previously infected people are still at an extremely high risk of being infected with Omicron [8–11]. Although a preliminary understanding of Omicron mutations has been achieved, internal dynamic evolution and genetic differentiation remained unknown during its transmission in

* Corresponding author: Suzhou Institute of Systems Medicine, Suzhou 215123, China.
E-mail address: wap@ism.cams.cn (A. Wu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bsheal.2022.04.004

HIGHLIGHTS

Scientific question

More mutations have been carried by the SARS-CoV-2 Omicron variant than previously reported variants. However, the genetic differentiation and diversity within Omicron variant that occurs during its early spread remains unclear.

Evidence before this study

At the end of 2021, a new SARS-CoV-2 variant Omicron appeared in South Africa. It had 50 consensus mutations, of which 31 mutations were in the S protein. Omicron had remarkable immune evasion ability and extremely fast transmission speed. As the infection cases increase, there is currently a large number of genome sequences from the early stage of the Omicron outbreak.

New findings

In this study, the genetic differentiation and diversity of the Omicron variant during its early outbreak has been comprehensively analyzed. More deletions on Omicron genome were accumulated than other four SARS-CoV-2 variants in the same timescale. Seven new notable non-synonymous mutations emerged in addition to 50 known consensus mutations. The rapid spread of the Omicron variant might lead to its high genetic differentiation and diversity in the population.

Significance of the study

Our study showed that Omicron had remarkably rapid genetic differentiation and mutational diversity with its rapid spread. The findings reminded us that more attention should be paid to the emerging Omicron sub-lineages in disease prevention and control.

the human population. Therefore, it is urgent to understand the evolutionary progress of Omicron in the early outbreak stage, which will be significantly helpful in the prevention and control of the Omicron variant's spread.

2. Materials and methods

2.1. Data source

The genome sequences of SARS-CoV-2 were downloaded from the GISAID [12]. The multiple sequence alignment (MSA) and spatiotemporal files (Metadata) were downloaded on December 12, 2021. Since the delayed updates of MSA file data, we downloaded raw Omicron sequence data and treated them as other sequences in the MSA file suffered. Excluding sequences with more than 5% unknown bases (N), there were 12,304 Omicron sequences until December 20, 2021. To further analyze the genetic diversity in the later period, we further downloaded 103,688 Omicron sequences collected in England on January 8, 2022. These sequences were individually aligned to the reference WIV04 (EPI_ISL_402124) by MAFFT [13]. The early-stage sequences of variants Alpha, Beta, Gamma, and Delta were extracted from the MSA file to form two datasets with the same sequence number or time duration as the Omicron dataset mentioned above. The initial time points for the other four variants (Alpha, Beta, Gamma, and Delta) were set at the day when they had 100 genome sequences. The mutation information of these sequences was extracted by an R package (https://github.com/wuaipinglab/genome_treatment). Mutations before the 300th and after the 29,000th bases were discarded for the low sequencing quality at the head and tail of the genome. Nucleotide substitutions occurring more than three times and insertions or deletions occurring more than once were kept. To remove the interference from low-quality sequencing data, we discarded sequences with N on the position between one base before and one base after each variant's consensus deletions/insertions. The final used sequence numbers of each strain were shown in supplementary Tables 4 and 7.

2.2. Phylogenic tree

The Phylogenic tree was downloaded from NextStrain [14], accessed on December 22, 2021, as a NEXUS file, including detailed lineage information. Characteristic mutations tables of the five variants were downloaded from the website Outbreak.info on December 25, 2021 [4]. Mutations in more than or equal to 75% of sequences would be treated as characteristic mutations. Characteristic mutations tables of the four Omicron lineages (B.1.1.529, BA.1, BA.2, and BA.3) and the five cluster groups (a, b, c, d, and e) were calculated with the Omicron sequences treated above. Only the mutations that appeared in more than half of the sequences were shown.

2.3. Omicron transmission network

The spread of SARS-CoV-2 was distributed scaleless. Many infected people could contribute to a large-scale virus spread through multiple gathering events in its early stages. In the scaleless network, a few nodes could connect to a large number of nodes. These key intermediate nodes might help to infer the transmission route of Omicron. We first extracted all the mutations (nucleotide substitutions, deletions, and insertions) in Omicron sequences to discover these key nodes. We then clustered these genome sequences based on their mutation similarity using an apcluster package in R [15]. Eventually, we had 253 clusters. Within each cluster, the earliest strains that appeared in different countries were selected as the representative sequences, and a total of 782 representative sequences were obtained. Two sequences were speculated to have a propagation relationship if there was only one nucleotide substitution difference, and a link was made between them. Nucleotide substitutions were discarded if there were more than 1000 N (from more than 10,000 sequences) at its location. Therefore, although some nucleotide substitutions occurred many times, they were not included in further analysis. Finally, these 782 sequences formed an omicron propagation network with 8,224 edges. We visualized this network in the software Gephi [16].

2.4. Mutations in the protein structure

We downloaded the monomer structure of S protein (QHD43416.pdb) from the website on December 22, 2021 (https://zhanggroup.org//COVID-19/). We visualized the S proteins using Pymol [17]. Mutations in Omicron were labeled on the S proteins.

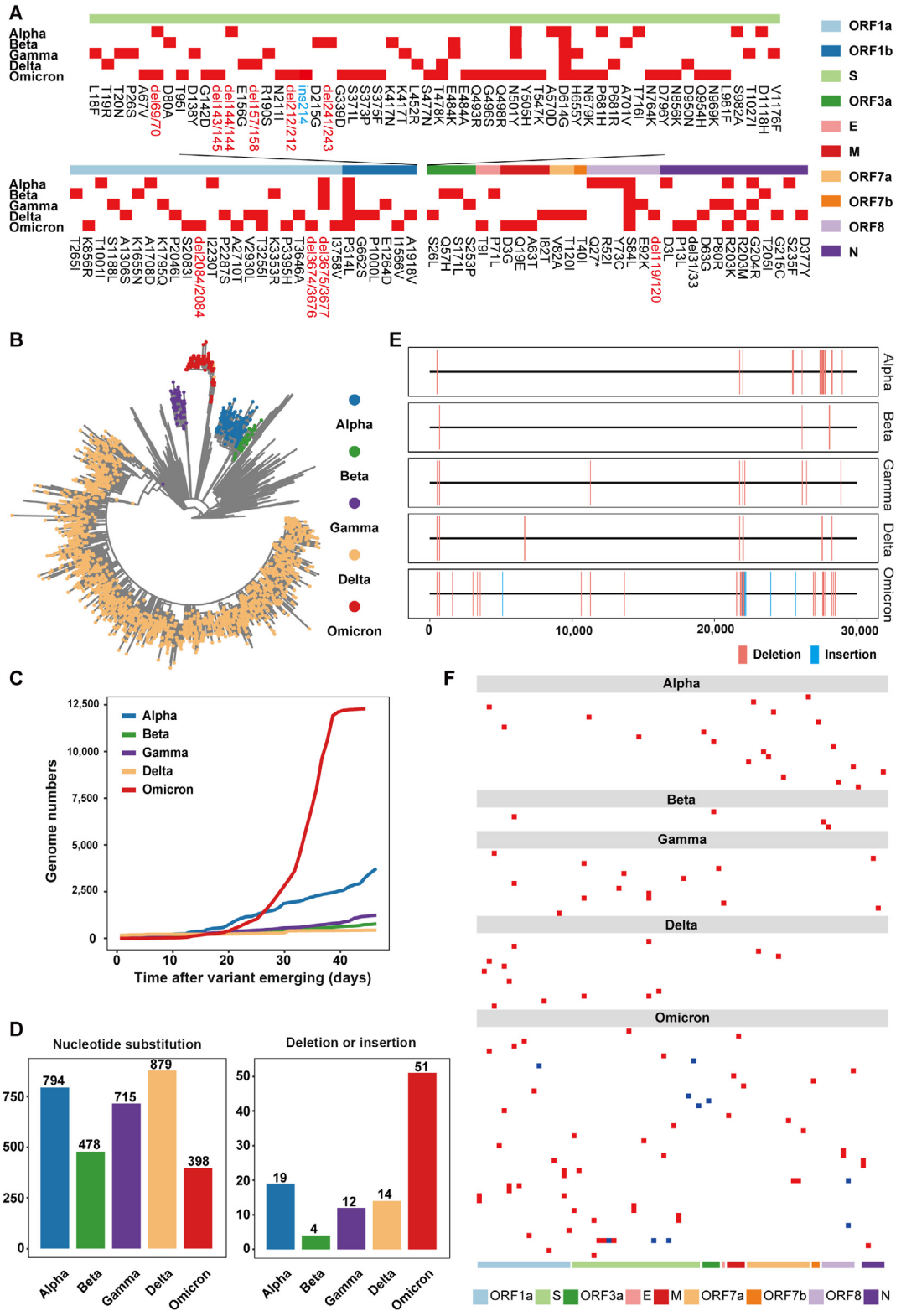
3. Results

3.1. Omicron had more mutations than other SARS-CoV-2 VOCs

The emergence of the Omicron variant has raised significant concerns about its vast genome mutations. The Omicron variant had 50 consensus mutations, including 43 nucleotide substitutions, six deletions, and one insertion. Of them, 27 nucleotide substitutions, three deletions, and one insertion were on the S protein (Fig. 1A and Table S1). The other four VOCs possessed relatively fewer mutations: Alpha had 22 mutations, Beta had 18 mutations, Gamma had 23 muta-

tions, and Delta had 29 mutations. In addition, systematic studies from the NextStrain website revealed that Omicron did not come from the previous dominant strain Delta [5,14,18] but was an individually emerging variant (Fig. 1B).

The accumulated genomes increased exponentially within 47 days, indicating that the Omicron variant had a relatively high speed spread worldwide (Fig. 1C). During the first 47 days after their emergence, Omicron, Alpha, Beta, Gamma, and Delta variants were reported with 12304, 3364, 733, 961, and 441 sequenced genomes, respectively (Fig. 1C). We compared the mutation number of these variants accumulated in their first 47 days. The result showed that Omicron contained 398 nucleotide substitutions, similar to that in the Beta variant and was half of the other three variants (Fig. 1D and Table S2). However, deletions or insertions in Omicron significantly happened more frequently than those in the other variants, up to twice as many (Fig. 1D and Table S3). Then we performed a systematic anal-



ysis of these deletions and insertions. Although all the five SARS-CoV-2 variants shared a similar deletion regional preference, deletions in variant Omicron had a wider distribution on the genome. The deletion regions in variant Omicron generally covered the regions where most deletions occurred in the other variants (Fig. 1E). Furthermore, more diverse deletion combinations were observed in Omicron (Fig. 1F).

3.2. Continuous genetic differentiation in early Omicron transmission

The Omicron variant was further divided into four lineages, namely B.1.1.529, BA.1, BA.2, and BA.3, based on the NextStrain [14]. The number of genome sequences of these four different Omicron branches in the first-47-day Omicron dataset was 23, 10,754, 20, and 8, respectively (Table S4). Finally, we displayed the consensus mutations, which occurred in over 50% of sequences of each branch. For example, the consensus mutations of the BA.1 branch covered completely that of the B.1.1.529 branch (Fig. 2A). At the same time, the BA.2 and BA.3 branches contained their unique consensus mutations (Fig. 2A and Table S5).

Because of the few sequenced genomes of B.1.1.529, BA.2, and BA.3 in the early stage of variant Omicron, we only used BA.1 sequences to further analyze the genetic diversity after the introduction of the Omicron variant into the human population. In addition to the 50 consensus mutations, we found ten other sites had relatively high-frequency mutations, including seven non-synonymous nucleotide substitutions (nsp3_V1069I, nsp4_V94A, nsp12_F685Y, S_R346K, S_A701V, S_I1081V, and ORF3a_L106F) and three synonymous nucleotide substitutions (Fig. 2B). After a cluster analysis by all nucleotide substitutions, including these ten high-frequency mutations, the early-stage BA.1 sequences could be divided into five groups, as group a–e (Fig. 2B). Except for Group a, each group had one or two nucleotide substitutions. Three nucleotide substitutions on the S protein belonged to group b, c, and d, respectively.

We then built a network. Two sequences were linked in this network if there was only one nucleotide substitution difference between them. The whole network presented a process of continuous diffusion from the center to the outside. We labeled group a–e in this network. We found that these five groups appeared in the different parts of the network (Fig. 2C). Group a was at the center of the network. Group c, d, and e were on the outside, connecting to group a through several nodes. Notably, group b did not connect with the other groups in the network. The intermediate nodes between group b and other groups were not included in our sequence dataset. When we mapped the detection time of each node into the network, we found that group c and d appeared earliest, followed by group b and e (Fig. 2D and Fig. S1). The spatiotemporal analyses showed that each of these groups had a unique distribution. Although in some countries, such as England and the United States, all groups were detected (Fig. S1A–S1D). These results indicated that the Omicron variant mutated and evolved during its early transmission.

3.3. Novel mutations appeared on the S protein of the Omicron variant

In the early-stage Omicron genome sequences, diverse mutations appeared in BA.1. These mutations divided BA.1 into one original group and four subgroups. We calculated the consensus mutations occurring in more than 50% of sequences within each group. Six unique nucleotide substitutions were notable (Fig. 3A and Table S6). Three unique nucleotide substitutions on the S protein belonged to groups b, c, and d. These mutations were S_R346K (group c), S_A701V (group d), and S_L1081V (group b), respectively. Besides, group b had L106F on the orf3a protein which could induce apoptosis [19]. The L106F has been reported in India and Brazil [20,21]. Group d had another mutation, V1069I, on the papain-like protease domain of the nsp3 protein, responsible for cleaving the polypeptide [22]. Finally, Ggroup e contained one unique nucleotide substitution, F685Y, located on the nsp12 encoding RNA-dependent RNA polymerase. Group c was later named Lineage BA.1.1. The Outbreak.info showed that the number of BA.1.1 cases increased rapidly, and the BA.1.1 has infected more than 20% of cases in Djibouti [4]. We mapped consensus Omicron mutations acquired from Outbreak.info together with three other novel nucleotide substitutions to the three-dimensional structure of the S protein [4]. The S_R346K mutation was on the receptor-binding domain (RBD) and the other two mutations, S_A701V and S_L1081V, were on the later part of the S2 region (Fig. 3B).

3.4. Faster spread led to higher genetic diversity in the Omicron variant

To determine whether more genome sequences could contribute to more mutations, we compared the internal diversity among different variants (Alpha, Delta, and Omicron) with the same number of sequences from England (Table S7). Until January 8, 2022, there were 103,688 sequences of the Omicron variant in England, and the genome of Omicron also accumulated faster than that of the other variants. The faster accumulation indicated that the spread speed of the Omicron variant was faster than the other variants in this region (Fig. 4A). In our results, the deletion diversity among Omicron increased, with a rapid increase in the number of sequenced genomes, significantly faster than that of the Alpha and Delta variants (Fig. 4B). In addition, a similar growth trend was shared between the internal diversity of nucleotide substitution and that of insertion or deletion in variant Alpha, Delta, and Omicron (Fig. 4C). When we mapped the location of deletions and insertions from different variants on the SARS-CoV-2 genome, we found that in the sequences with the same time duration after the initial day of each variant in England, which was labeled by a dotted line in Fig. 4A, the deletions of variant Omicron distributed wider on the genome than that of other variants (Fig. 4D). The wider distribution could result from a rapid sequence accumulation of variant Omicron in the early stage (Fig. 4A). However, when these variants came to have the same sequence number, their deletion distribution tended to be similar (Fig. 4E). The above results indicated that higher genetic diversity in the Omicron variant could be related to a faster spread in its early outbreak stage.

4. Discussion

Compared to the other VOCs, variant Omicron had almost four times the number of sequenced genomes within the same time duration after the initial day of the variants. The larger number of infection cases of this dominant variant could result in the more frequent emergence of mutations. During the first 47 days of spread, 398 nucleotide

Fig. 1. Comparing the diversity of mutations among the five variants of concern (VOCs) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). A) The consensus mutations of the variants Alpha, Beta, Gamma, Delta, and Omicron are shown in a heatmap. Word labels in the red highlight deletions, and those in blue highlight insertions. Nucleotide substitutions were labeled in black. B) A phylogenetic tree from the NextStrain is displayed, whose five VOCs are labeled in different colors. C) The number of the genome sequences of five VOCs uploaded to the GISAID grew overtime in the first 47 days. D) Among these sequences, the types of nucleotide substitutions and deletions/insertions were calculated in each variant of concern. E) The location distribution of these deletions (red) or insertions (blue) on the SARS-CoV-2 genome is shown in bars. F) In each line, every combination of deletion (red) or insertion (blue) in five variants is shown in the heatmap. Each column represents a unique deletion/insertion ordering by their location on the SARS-CoV-2 genome. Their respective genes are labeled in different colors.

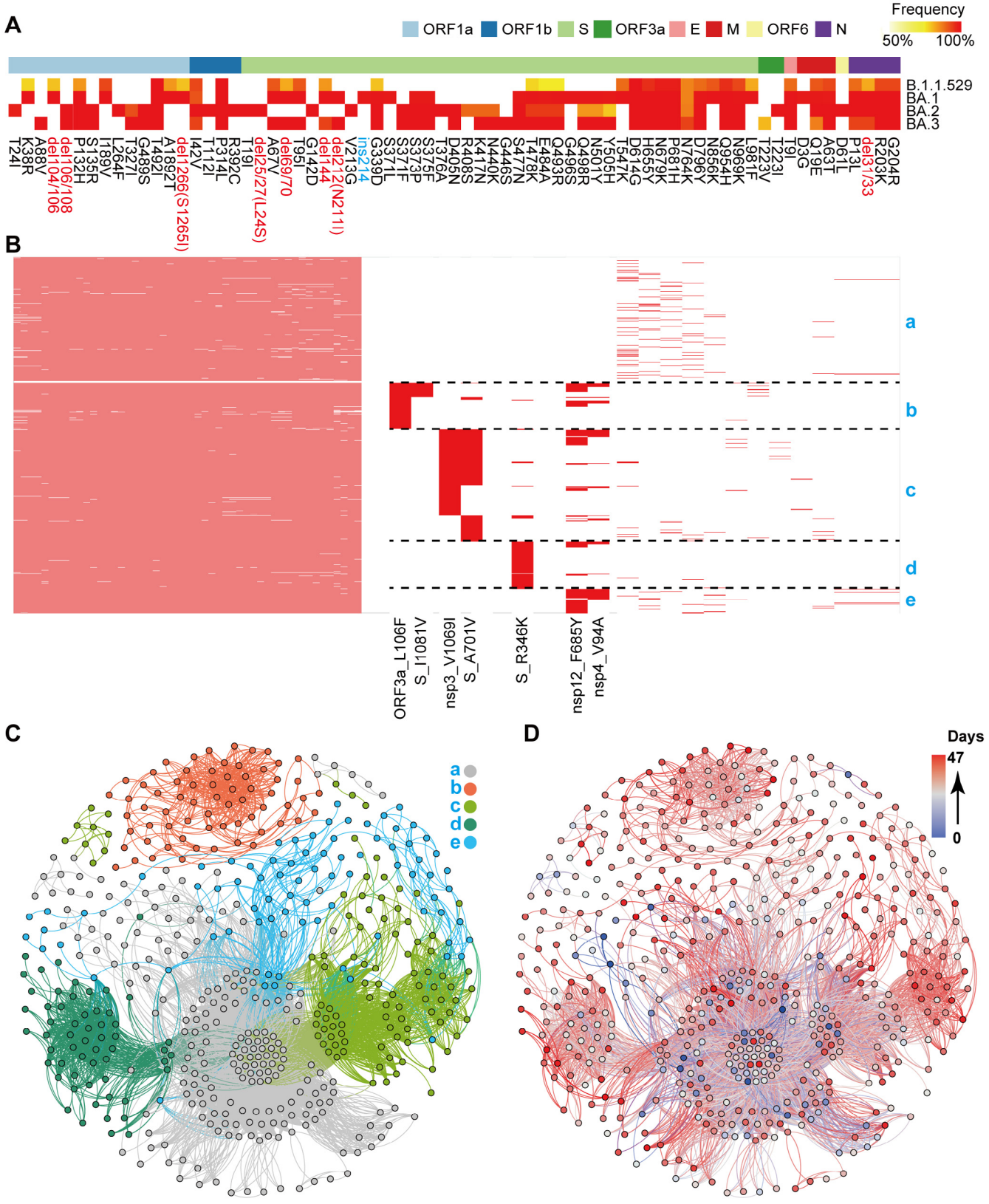


Fig. 2. Differentiation and diversity within the early sequences of Omicron. A) The characteristic mutations of each sub-lineages in early Omicron sequences are shown in the heatmap. Word labels in red highlight deletions, and those in blue highlight insertions. Nucleotide substitution labels were in black. The frequency of the mutations in each sub-lineage is shown from yellow to red as 50% to 100%. B) In the sequences from the BA.1 sub-lineage, their nucleotide substitution combinations are shown in each line in the heatmap and further grouped into a, b, c, d, and e group. C) A propagation net was created by early Omicron sequences after clustered. D) The time characteristic of these sequences is shown in the propagation net.

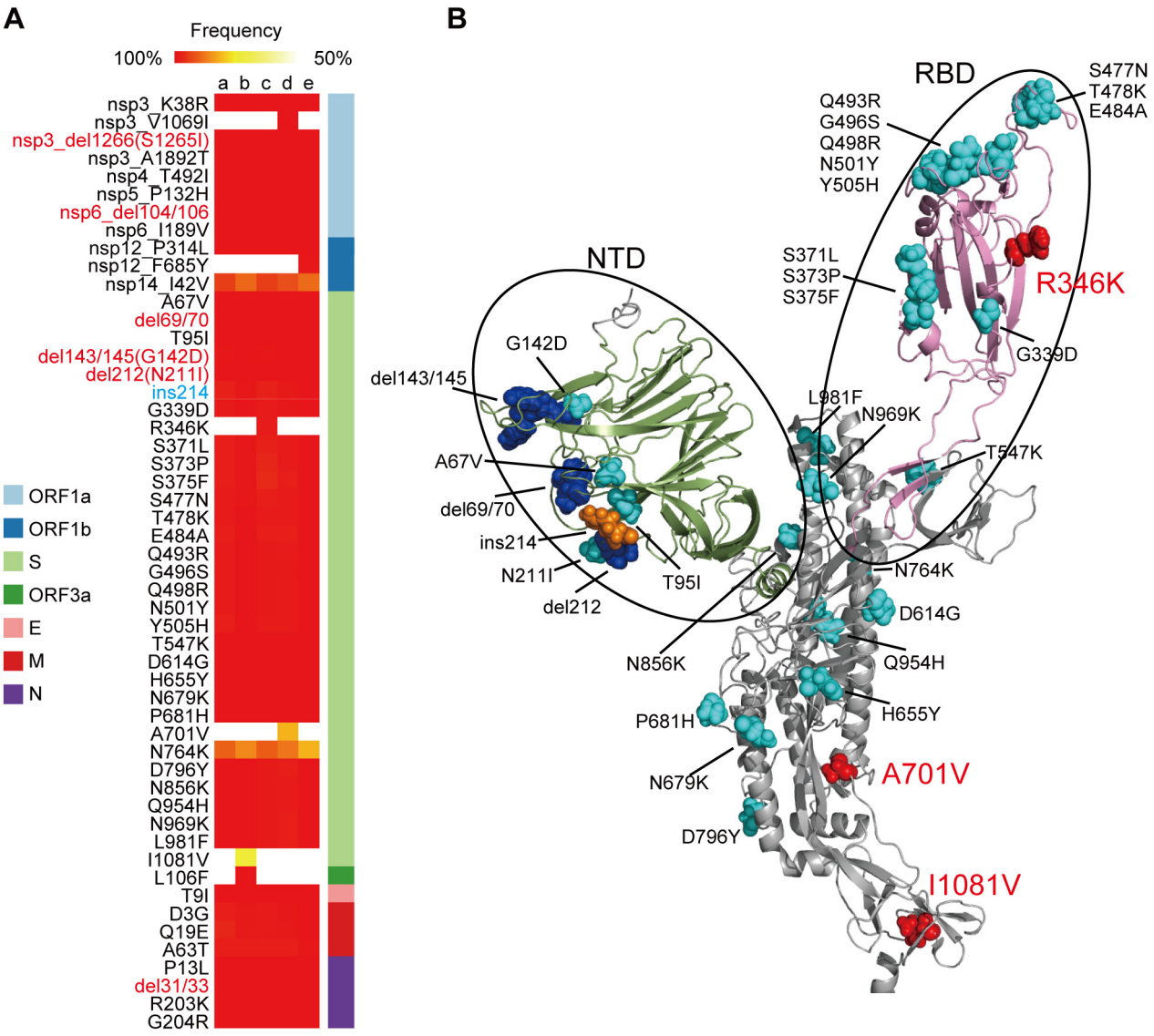


Fig. 3. The mutation distribution of the early Omicron sequences in different groups. A) The characteristic mutations of each group in early Omicron sequences are shown. Word labels in red highlight deletions, and those in blue highlight insertions. Nucleotide substitution labels were in black. The frequency of the mutations in each group is shown from yellow to red as 50% to 100%. B) Consensus mutations in Omicron and remarkable mutations in group a, b, c, d, and e are mapped on the structural model of the Spike protein. N-terminal domain (NTD) is marked in green, and receptor-binding domain (RBD) is marked in pink. Blue indicates deletion positions, yellow indicates insertion positions, cyan indicates nucleotide substitutions, and red indicates remarkable mutations.

substitutions and 51 deletions/insertions were identified in the Omicron sequences. Previous studies showed that deletions could affect the virus protein greater than single nucleotide substitutions [23]. A systemic analysis revealed that deletion in the SARS-CoV-2 had a regional preference. It was also illustrated that the recurrent deletions on the N-terminal Domain of the S protein partially covered the binding domain of some neutralizing antibodies indicating a potential role of the deletions in virus evolution [24,25]. Therefore, in preventing and controlling the COVID-19 pandemic, it was necessary to pay more attention to the internal genetic diversity, including nucleotide substitutions and deletions or insertions of the dominant variants.

Previous studies have shown that the Omicron variant consisted of four sub-lineages. These sub-lineages seemed to emerge at similar times, two of which (BA.1 and BA.2) had spread worldwide [26]. A recent study showed that the BA.2 lineage, which appeared later, might spread faster than BA.1 [7]. Our study showed that the BA.1 lineage continued to differentiate. We divided BA.1 into five groups by multiple mutations, including one original group and four subgroups. Each of these subgroups carried one novel nucleotide substitution. Group c, with S_R346K, increased rapidly and has been named the BA.1.1 lineage [27]. Group b and d possessed their own one characteristic nucleotide substitutions on the S protein. These mutations on the spike protein might potentially affect virus transmission. Finally, group e had one nucleotide substitution (nsp12_F685Y) on the nsp12. Since nsp12, encoding RNA-dependent RNA polymerase, participates in virus replication and translation, it was meaningful to figure out the function of this mutation.

On the S protein of the Omicron variant, there were 31 consensus mutations. Some of them, such as S477R, Q498R, and N501Y, have already been associated with an increased binding ability to the ACE2 receptor [28–31]. Another consensus nucleotide substitution, S_K417N, has been confirmed to be able to inactivate some therapeutic neutralizing antibodies [29–31]. The consensus deletion S_del69/70 has been proved to help the virus enter host cells [32]. Except for these notable mutations, we found that a series of novel mutations continued to emerge on the S protein during the spread of variant Omicron.

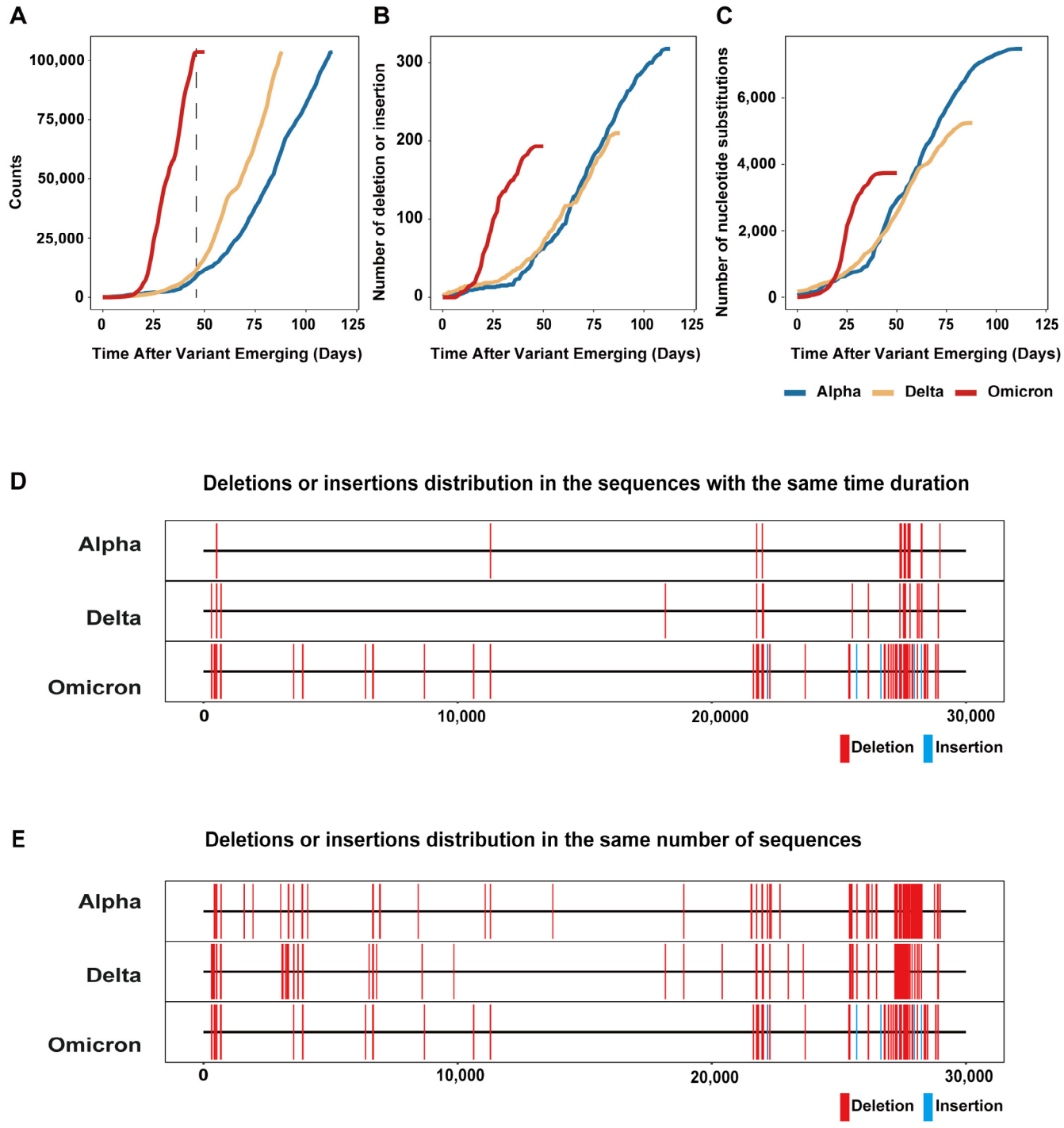


Fig. 4. The relationship between Omicron transmission and mutation accumulation in England. A) The genome sequence numbers of dominant variants (Alpha, Delta, and Omicron) uploaded to the GISAID grew over time in the early stage of Omicron in England. A dotted line marked the time duration which we used to show the deletions or insertions distribution in the sequences with the same time duration. B) and C) The types of deletion/insertion or nucleotide substitutions in these sequences were calculated over time. The location of these deletions or insertions were labeled separately on the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome, D) in the sequence with the same time duration or E) in the same number of sequences.

Three novel nucleotide substitutions (S_R346K, S_A701V, S_L1081V) were detected in these early-stage sequences, of which S_R346K was on the receptor-binding domain. S_R346K has been proved to slightly affect the binding between SARS-CoV-2 virus and class 2 antibodies [33]. Another nucleotide substitution, S_A701V, was one of the dominant mutations in the third pandemic wave in Malaysia [34]. In addition, many studies have proved that Omicron had a solid ability to escape several neutralizing antibodies [8], and previously infected people were also the susceptible population [11]. Therefore, it is critical to figure out the ability of not only consensus mutations, but also these emerging mutations of the Omicron variant on virus transmission ability and immune escape capability.

From January 2022 to February 2022, the Omicron variant has spread exponentially. Many infected people could lead to more confirmed cases with severe symptoms and a shortage of medical resources. Our studies implied that a rapid increase in the infected

patients might lead to a rapid increase in viral diversity. More dangerous mutations may even occur in the future. Therefore, the internal diversity among the dominant variants should be followed through more mutation surveillance. Therefore, reasonable pandemic prevention and control measures should be carried out cautiously to confront the Omicron variant's spread.

Acknowledgements

This work was supported by the National key research and development program (2021YFC2301300); the CAMS Innovation Fund for Medical Sciences (2021-I2M-1-061); the National Natural Science Foundation of China (92169106, 31900472); the special research fund for central universities, Peking Union Medical College (2021-PT180-001); Suzhou science and technology development plan (szs2020311).

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Author contributions

Shenghui Weng: Conceptualization, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing. **Jingzhe Shang:** Conceptualization, Data Curation, Writing – Original Draft, Writing – Review & Editing. **Yexiao Cheng:** Investigation. **Hangyu Zhou:** Conceptualization. **Chengyang Ji:** Methodology. **Rong Yang:** . **Aiping Wu:** Conceptualization, Writing – Review & Editing, Validation.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bsheal.2022.04.004.

References

[1] T. Li, T. Huang, C. Guo, A. Wang, X. Shi, X. Mo, Q. Lu, J. Sun, T. Hui, G. Tian, Genomic variation, origin tracing, and vaccine development of SARS-CoV-2: A systematic review, Innovation 2 (2) (2021), 100116. https://doi.org/10.1016/j.xinn.2021.100116.

[2] WHO, Tracking SARS-CoV-2 variants. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants, 2022 (accessed 21 January 2022).

[3] R. Viana, S. Moyo, D.G. Amoako, H. Tegally, W.T. Choga, O. Lesetedi-Mafoko, T. Mohale, S. Gaseitsiwe, A. von Gottberg, T. de Oliveira, et al, Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa, Nature 603 (7902) (2022) 679–686, https://doi.org/10.1038/s41586-022-04411-y.

[4] J. Tsueng, A. Latif, M. Alkuzweny, M. Cano, E. Haag, J. Zhou, M. Zeller, E. Hufbauer, N. Matteson, et al., outbreak.info. https://outbreak.info/situationreports/omicron, 2022 (accessed 21 January 2022).

[5] L. Wang, G. Cheng, Sequence analysis of the emerging Sars-CoV-2 variant omicron in South Africa, J. Med. Virol. 94 (4) (2021) 1728–1733, https://doi.org/10.1002/jmv.27516.

[6] P. Du, G.F. Gao, Q. Wang, The mysterious origins of the Omicron variant of SARS-CoV-2, Innovation 3 (2) (2022), 100206. https://doi.org/10.1016/j.xinn.2022.100206.

[7] E. Mahase, Omicron sub-lineage BA.2 may have "substantial growth advantage, UKHSA reports, BMJ 376 (2022), o263. https://doi.org/10.1136/bmj.o263.

[8] Y. Cao, J. Wang, F. Jian, T. Xiao, W. Song, X. Hao, X. Wang, J. Xiao, Y. Wang, X.S. Xie, et al, Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies, Nature 602 (7898) (2022) 657–663, https://doi.org/10.1038/s41586-021-04385-3.

[9] D. Planas, N. Saunders, P. Maes, F. Guivel-Benhassine, C. Planchais, E. Simon-Loriere, T. Bruel, H. Mouquet, E. André, O. Schwartz, et al, Considerable escape of SARS-CoV-2 Omicron to antibody neutralization, Nature 602 (7898) (2022) 671–675, https://doi.org/10.1038/s41586-021-04389-z.

[10] E. Dolgin, Omicron thwarts some of the world's most-used COVID vaccines, Nature 601 (7893) (2022) 311, https://doi.org/10.1038/d41586-022-00079-6.

[11] J.R. Pulliam, C. van Schalkwyk, N. Govender, A. von Gottberg, C. Cohen, M.J. Groome, J. Dushoff, K. Mlisana, H. Moultrie, Increased risk of SARS-CoV-2 reinfection associated with emergence of the Omicron variant in South Africa, Science (2022), eabn4947. https://doi.org/10.1126/science.abn4947.

[12] Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data–from vision to reality, Euro Surveill. 22 (13) (2017), 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

[13] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[14] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, J. Kelso, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics 34 (23) (2018) 4121–4123, https://doi.org/10.1093/bioinformatics/bty407.

[15] U. Bodenhofer, A. Kothmeier, S. Hochreiter, APCluster: an R package for affinity propagation clustering, Bioinformatics 27 (17) (2011) 2463–2464, https://doi.org/10.1093/bioinformatics/btr406.

[16] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, Proceedings of the International AAAI Conference on Web and Social Media 3 (1) (2009) 361–362, https://doi.org/10.13140/2.1.1341.1520.

[17] W.L. DeLano, Pymol: An open-source molecular graphics tool. http://legacy.ccp4.ac.uk/newsletters/newsletter40/11_pymol.html, 2002 (accessed 8 February 2022).

[18] W. Ma, J. Yang, H. Fu, C. Su, C. Yu, Q. Wang, A.T.R. de Vasconcelos, G.A. Bazykin, Y. Bao, M. Li, Genomic perspectives on the emerging SARS-CoV-2 omicron variant [Preprint], Genomics Proteomics Bioinf. (2022), S1672-0229(22)00002-X. https://doi.org/10.1016/j.gpb.2022.01.001.

[19] Y. Ren, T. Shu, D. Wu, J. Mu, C. Wang, M. Huang, Y. Han, X.Y. Zhang, W. Zhou, Y. Qiu, X. Zhou, The ORF3a protein of SARS-CoV-2 induces apoptosis in cells, Cell. Mol. Immunol. 17 (8) (2020) 881–883, https://doi.org/10.1038/s41423-020-0485-9.

[20] B. Sahoo, P.K. Maurya, R.K. Tripathi, J. Agarwal, S. Tiwari, Diversity of SARS-CoV-2 genome among various strains identified in Lucknow, Uttar Pradesh [Prerprint], bioRxiv (2021), https://doi.org/10.1101/2021.10.05.463185.

[21] U. Souza, R. Santos, A. Belmok, F. Melo, J.D. Galvao, S.B. Damasceno, T. Rezende, M. Andrade, B.M. Ribeiro, J.R. Junior, Detection of potential new SARS-CoV-2 Gamma-related lineage in Tocantins shows the spread and ongoing evolution of P.1 in Brazil [Preprint], bioRxiv (2021), https://doi.org/10.1101/2021.06.30.450617.

[22] L.A. Armstrong, S.M. Lange, V. Dee Cesare, S.P. Matthews, R.S. Nirujogi, I. Cole, A. Hope, F. Cunningham, R. Toth, R. Mukherjee, Biochemical characterization of protease activity of Nsp3 from SARS-CoV-2 and its inhibition by nanobodies, PLoS One 16 (7) (2021), e0253364. https://doi.org/10.1371/journal.pone.0253364.

[23] M. Wellenreuther, C. Mérot, E. Berdan, L. Bernatchez, Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification, Mol. Ecol. 28 (6) (2019) 1203–1209, https://doi.org/10.1111/mec.15066.

[24] S. Weng, H. Zhou, C. Ji, L. Li, N. Han, R. Yang, J. Shang, A. Wu, Conserved pattern and potential role of recurrent deletions in SARS-CoV-2 evolution, Microbiol. Spectr. 10 (2) (2022), e0219121. https://doi.org/10.1128/spectrum.02191-21.

[25] K.R. McCarthy, L.J. Rennick, S. Nambulli, L.R. Robinson-McCarthy, W.G. Bain, G. Haidar, W.P. Duprex, Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape, Science 371 (6534) (2021) 1139–1142, https://doi.org/10.1126/science.abf6950.

[26] S. Mallapaty, Where did Omicron come from?, Three key theories, Nature 602 (7895) (2022) 26–28, https://doiorg/10.1038/d41586-022-00215-2.

[27] C. Roemer, Omicron sublineage with potentially beneficial mutation S:346K. https://github.com/cov-lineages/pango-designation/issues/360, 2021 (accesed 5 December 2021).

[28] Q. Li, J. Wu, J. Nie, L.i. Zhang, H. Hao, S. Liu, C. Zhao, Q.i. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang, Y. Wang, The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity, Cell 182 (5) (2020) 1284–1294.e9, https://doi.org/10.1016/j.cell.2020.07.012.

[29] T.N. Starr, A.J. Greaney, S.K. Hilton, D. Ellis, K.H.D. Crawford, A.S. Dingens, M.J. Navarro, J.E. Bowen, M.A. Tortorici, A.C. Walls, N.P. King, D. Veesler, J.D. Bloom, Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding, Cell 182 (5) (2020) 1295–1310.e20, https://doi.org/10.1016/j.cell.2020.08.012.

[30] A.J. Greaney, A.N. Loes, K.H.D. Crawford, T.N. Starr, K.D. Malone, H.Y. Chu, J.D. Bloom, Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies, Cell Host Microbe 29 (3) (2021) 463–476.e6, https://doi.org/10.1016/j.chom.2021.02.003.

[31] P. Wang, M.S. Nair, L. Liu, S. Iketani, M. Sobieszczyk, C.A. Kyratsous, L. Shapiro, Z. Sheng, Y. Huang, D.D. Ho, et al, Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7, Nature 593 (7857) (2021) 130–135, https://doi.org/10.1038/s41586-021-03398-2.

[32] B. Meng, S.A. Kemp, G. Papa, R. Datir, I.A. Ferreira, S. Marelli, W.T. Harvey, S. Lytras, A. Mohamed, G. Gallo, Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the variant of concern lineage B, Cell Rep. 35 (13) (2021), 109292. https://doi.org/10.1016/j.celrep.2021.109292.

[33] F. Fratev, The R346K mutation in the mu variant of SARS-CoV-2 alter the interactions with monoclonal antibodies from Class 2: A free energy of perturbation study, J. Chem. Inf. Model. 62 (3) (2022) 627–631, https://doi.org/10.1021/acs.jcim.1c01243.

[34] J. Suppiah, K. Kamel, Z. Mohd-Zawawi, M. Afizan, H. Yahya, S. Md-Hanif, R. Thayan, Phylogenomic analysis of SARS-CoV-2 from third wave clusters in Malaysia reveals dominant local lineage B. 1.524 and persistent spike mutation A701V, Trop. Biomed. 38 (3) (2021) 289–293, https://doi.org/10.47665/tb.38.3.070.