



OPEN

Genomic epidemiology of SARS-CoV-2 Omicron variants in the Republic of Korea

Dong-Wook Lee^{1,3}, Jeong-Min Kim^{2,3}, Ae Kyung Park², Da-Won Kim¹, Ji-Yun Kim¹, Noori Lim¹, Hyeokjin Lee², Il-Hwan Kim², Jeong-Ah Kim², Chae young Lee², Jung-Hoon Kwon^{1✉} & Eun-Jin Kim^{2✉}

The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic since 2019. Variants of concern (VOCs) declared by the World Health Organization require continuous monitoring because of their possible changes in transmissibility, virulence, and antigenicity. The Omicron variant, a VOC, has become the dominant variant worldwide since November 2021. In the Republic of Korea (South Korea), the number of confirmed cases increased rapidly after the detection of Omicron VOC on November 24, 2021. In this study, we estimated the underlying epidemiological processes of Omicron VOC in South Korea using time-scaled phylodynamic analysis. Three distinct phylogenetic subgroups (Kor-O1, Kor-O2, and Kor-O3) were detected in South Korea. The Kor-O1 subgroup circulated in the Daegu region, whereas Kor-O2 and Kor-O3 circulated in Incheon and Jeollanam-do, respectively. The viral population size and case number of the Kor-O1 subgroup increased more rapidly than those of the other subgroups, indicating the rapid spread of the virus. The results indicated the multiple introductions of Omicron sub-lineages into South Korea and their subsequent co-circulation. The evolution and transmission of SARS-CoV-2 should be continuously monitored, and control strategies need to be improved to control the multiple variants.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread globally; as of March 2, 2022, there have been 437,333,859 confirmed cases, including 5,960,972 deaths worldwide¹. New variants of SARS-CoV-2 should be constantly observed because some contain mutations that increase their transmissibility, virulence, and ability to evade immune responses^{2–4}. These variants have been classified as Variants of Concern (VOCs) by the World Health Organization⁵. These VOCs include B.1.1.7 (Alpha), first identified in England⁶, B.1.351 (Beta) in South Africa⁷, P.1 (Gamma) in Brazil⁸, B.1.617.2 (Delta) in India⁹, and B.1.1.529 (Omicron) in South Africa and Botswana^{10,11}. Global data show that the Alpha variant was the dominant variant until May 2021. Thereafter, the number of Delta variants increased and peaked from October 2021 to November 2021. After the emergence of Omicron in November 2021, the Omicron variant became the dominant global variant, and the weekly number of cases rapidly increased^{12,13}. Omicron has many mutations compared with other variants. In particular, the mutations concentrated in the S protein enable it to evade neutralizing antibodies or vaccine-induced immunity^{14–16}.

In the Republic of Korea (South Korea), the number of confirmed cases has continuously increased since the first influx of Omicron on November 24, 2021, surpassing 10,000 on January 26, 2022, and by mid-March 2022 more than 300,000 cases had been reported¹⁷. The increased transmissibility of Omicron VOC was suspected as a major factor in the exponential increase of cases in South Korea and worldwide; however, other factors, including vaccine efficacy and level of social distancing, may contribute to the viral spread. With an ongoing exponential increase in confirmed cases, simultaneous outbreaks occurring in all states of South Korea, and the continuous introduction of SARS-CoV-2 from abroad, tracing all the contacts of infected patients is impossible. Thus, epidemiological investigations based on contact tracking methods could not present significant epidemiological patterns of viruses. Consequently, molecular epidemiology based on phylogenetic analysis has become critical for understanding the outbreak of SARS-CoV-2. To investigate the recently increasing number of Omicron strains circulating in South Korea, we investigated the evolutionary and geographical relationships of Omicron in South Korea, based on the phylogenetic analysis of full-length viral genome sequences.

¹College of Veterinary Medicine, Kyungpook National University, Daegu 41566, Republic of Korea. ²Division of Emerging Infectious Diseases, Bureau of Infectious Disease Diagnosis Control, Korea Disease Control and Prevention Agency, Cheongju-Si 28159, Republic of Korea. ³These authors contributed equally: Dong-Wook Lee and Jeong-Min Kim. ✉email: Junghoon.kwon@knu.ac.kr; ekim@korea.kr

Results

Maximum-likelihood phylogenetic tree reveals three different virus groups transmitted in South Korea. A total of 582 sequences of the Omicron variant were obtained from November 24, 2021, to January 11, 2022. The proportion of the Omicron variant rapidly increased during this period and replaced the previously dominant Delta variant in South Korea. Of the various sub-lineages, the BA.1 sub-lineage was dominant (Fig. 1a).

We constructed approximately maximum-likelihood phylogenetic trees of BA.1 sub-lineages and detected 128 distinct monophyletic subgroups sharing a common ancestral node with other Korean viruses, with a local support value of >0.7 . Most subgroups caused only single or short local outbreaks; however, three subgroups—Kor-O1 ($n=273$), Kor-O2 ($n=68$), and Kor-O3 ($n=86$)—caused multiple local outbreaks in South Korea (Fig. 1b). In the Kor-O1 subgroup, two inbound travelers were detected, which included imported cases from Ivory Coast on December 30, 2021, and Congo on December 29, 2021; however, no inbound travelers were detected in the Kor-O2 or Kor-O3 subgroups (Table S1). The Kor-O1 subgroup also included 35 sequences detected outside South Korea, indicating the possible spread of viruses from South Korea to other countries.

The Kor-O1 and Kor-O3 subgroups belong to the BA.1.1 sub-lineage of Omicron, and the Kor-O2 subgroup belongs to BA.1. Korean Omicron sub-lineages have additional mutations from the original Omicron strain, hCoV-19/South Africa/NICD-N20868/2021, collected on November 11, 2021. All three subgroups have T5730C in ORF 1a and T462A in the ORF3a gene. The Kor-O1 subgroup mutated A603T and G604C in the N gene, Kor-O2 mutated G955A in the N gene, and The Kor-O3 subgroup mutated C2470T in the ORF 1a gene and C473T in the M gene (Fig. S1).

Genomic epidemiology reveals that these three subgroups have different tendencies for geographic transmission. To investigate the time to most recent common ancestor (tMRCA), mutation rates, and transmission trends, we constructed a time-scaled phylogenetic tree. The tMRCA results showed that all three Korean subgroups were introduced into South Korea between mid-November and mid-December, and Kor-O2 was estimated to be the first subgroup. For the Kor-O1 and Kor-O3 subgroups, the difference between the first isolation date and the mean of tMRCA was 9 days, and for Kor-O2, it was 12 days. The mutation rates for all three subgroups were between 1.0×10^{-3} and 1.3×10^{-3} substitutions/site/year and did not show significant differences between the subgroups (Table 1).

To estimate the transmission dynamics of the Omicron variant circulating in South Korea, we reconstructed ancestral locations and inferred migration of viruses for three Korean subgroups (Kor-O1, Kor-O2, and Kor-O3). Our results showed that the three subgroups had different geographic transmission tendencies. The Kor-O1 subgroup spread to 13 out of 17 states. On December 17, 2021, the Kor-O1 subgroup was first detected in Gyeonggi-do, and it was the major site of the outbreak until December 24, 2021. On December 31, 2021, it was primarily spread in Daegu, with an increasing rate of transmission from Daegu and Gyeongsangbuk-do to other regions (Fig. 2, Table 2, and Video S1).

In the case of the Kor-O2 subgroup, the virus was primarily detected in three states: Incheon, Gangwon-do, and Busan. Most transmissions started from Incheon to Gangwon-do before mid-December, located in the northern region of South Korea. However, after mid-December, transmission mainly proceeded in Gangwon-do, and nearly all transmissions were directed to Busan, the southern region of South Korea (Fig. 3, Table 3, Video S2).

Finally, in the case of Kor-O3, we detected that the transmission started in Jeollabuk-do by the end of December. Most transmissions were observed in Jeollanam-do, Jeollabuk-do, and Gwangju, which are in the western region of South Korea. We observed that Kor-O3 in Jeollabuk-do spread mainly to Jeollanam-do, Seoul, and Daegu and that in Jeollanam-do spread mainly to Gyeongsangbuk-do, Gwangju, and Incheon. From the beginning of January 2022, many transmissions were observed from Gyeongsangbuk-do to Gyeonggi-do and Daejeon (Fig. 4, Table 4, Video S3).

Population size dynamic analysis showed that the Kor-O2 subgroup was discovered first, followed by Kor-O3 and Kor-O1. Whereas the population size of the Kor-O2 subgroup was constant, the population size of the Kor-O1 and Kor-O3 subgroups continuously increased. Furthermore, the Kor-O2 subgroup was not detected after the last case detected on January 5, 2022 (Fig. 5).

Discussion

The mutations of the Omicron variant were concentrated in the spike protein, including four deletions and one insertion ($\Delta 69-70$, $\Delta 143-145$, $\Delta 211-212$, ins214EPE)^{18,19}. Mutations in the spike protein can enhance the ability of the variant to evade current vaccines^{16,20}. In addition, mutations in the receptor binding domain could enhance receptor binding ability, which could affect virus transmissibility^{21,22}. The Omicron variant replaced the previous dominant lineage, Delta, and has become the dominant variant worldwide.

Omicron is classified into several sub-lineages (BA.1–5) based on its mutations in PANGOLIN (<https://pangolin.cog-uk.io/>)²³. In addition, several recombinant lineages between Delta and Omicron or between BA.1 and BA.2 have been continuously reported^{24,25}. The recombinant lineages were not detected in South Korea in this study because we analyzed data collected in the very early stages of the Omicron outbreak. Although BA.2 sequences were also detected, most cases in the early stages of the Omicron outbreak in South Korea belong to BA.1. Therefore, this study focused on BA.1 and its sub-lineage BA.1.1.

South Korea has implemented a 10-day quarantine policy for all inbound travelers since December 3, 2021, to prevent Omicron spread and designated 11 African countries as “Enhanced Quarantine Required Countries” (<http://ncov.mohw.go.kr/duBoardList.do#>). Despite national border control, our results showed that at least 128 distinct phylogenetic subgroups belonging to BA.1 were introduced in South Korea. Of these, 113 subgroups were introduced into South Korea after implementing border control. The detection of the Kor-O1 monophyletic

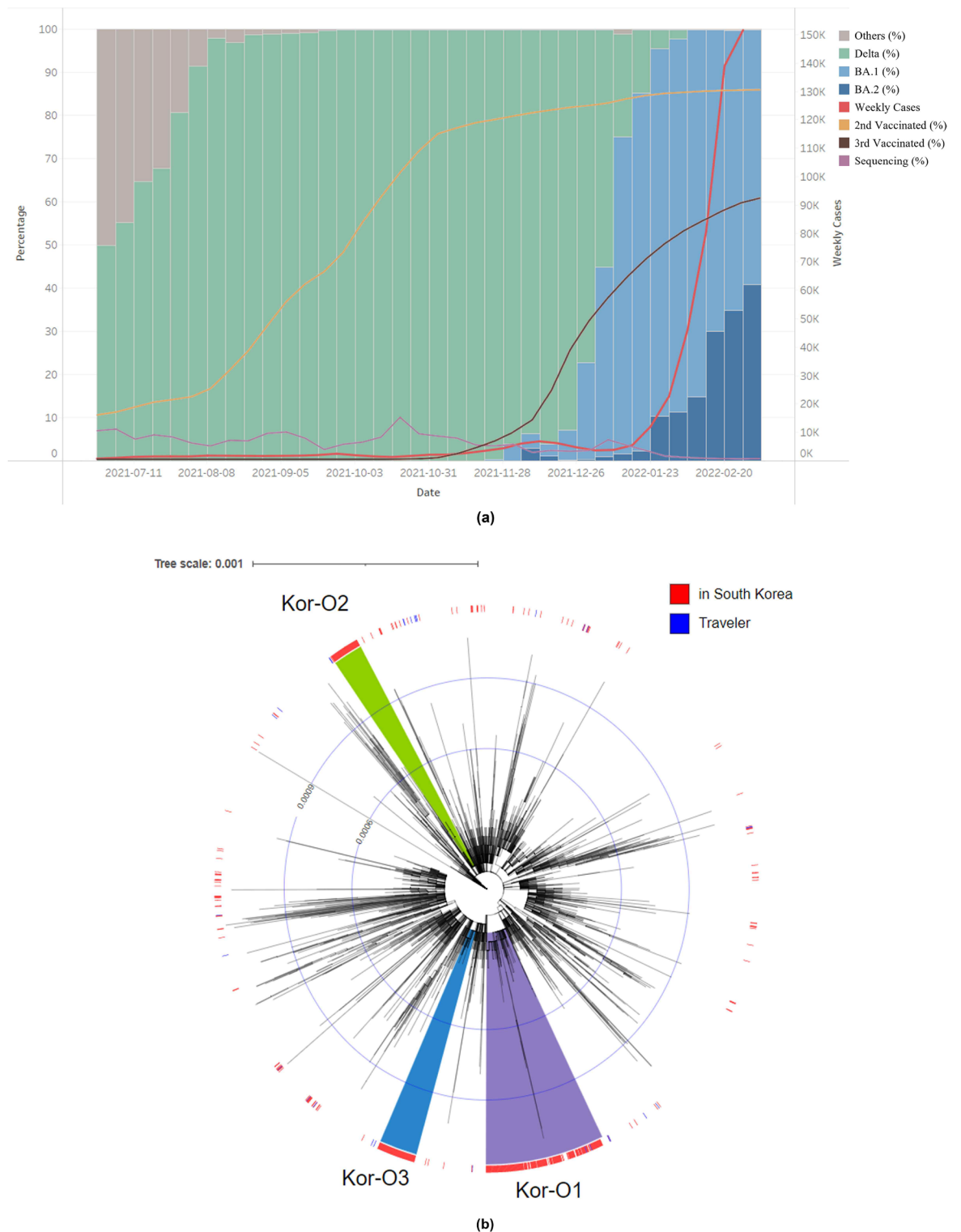


Figure 1. (a) Proportion of lineages and number of weekly cases of Korean sequences. Each column indicates the proportion of SARS-CoV-2 lineages of Korean sequences from July 7, 2021, to February 20, 2022. The red line indicates the number of weekly cases in South Korea from July 7, 2021, to February 20, 2022, and the yellow, brown, and purple lines indicate the 2nd vaccination rate, 3rd vaccination rate, and sequencing percentage. (b) Maximum-likelihood phylogenetic tree constructed by FastTree of 3330 foreign sequences and 582 Korean sequences. The SARS-CoV-2 viruses detected from travelers entering South Korea or their direct contacts were annotated by blue color strips, and viruses detected from South Korean residents with no travel history were annotated by red color strips. The Kor-O1 subgroup is highlighted in purple, Kor-O2 is highlighted in green, and Kor-O3 is highlighted in blue.

Subgroup	Date of first detection in South Korea	tMRCA ¹ (95% HPD) ²	Mutation rate ³ (95% HPD)
Kor-O1	Dec. 21, 2021	Dec. 12, 2021 (Nov. 29, 2021–Dec. 19, 2021)	1.2227×10^{-3} (7.2495×10^{-4} – 1.9052×10^{-3})
Kor-O2	Nov. 29, 2021	Nov. 17, 2021 (Oct. 27, 2021–Nov. 28, 2021)	1.2952×10^{-3} (3.8827×10^{-4} – 2.6098×10^{-3})
Kor-O3	Dec. 06, 2021	Nov. 27, 2021 (Nov. 21, 2021–Dec. 3, 2021)	1.0314×10^{-3} (5.2968×10^{-4} – 1.6441×10^{-3})

Table 1. First detection date, mutation rates, and time to most recent common ancestor of each Korean Omicron subgroup. ¹Time to most recent common ancestor. ²95% HPD, 95% height posterior density. ³Substitutions/site/year.

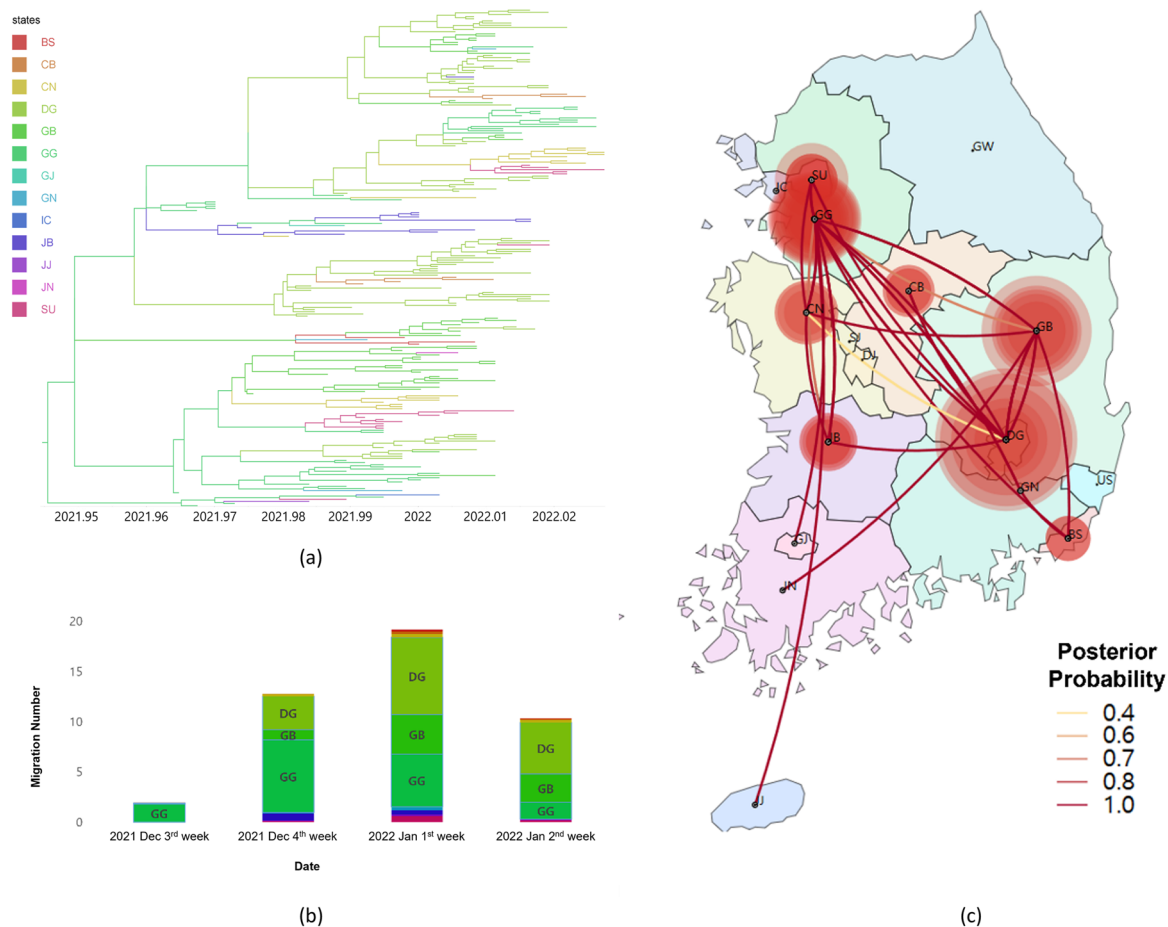


Figure 2. Phylogeographic result of the Kor-O1 subgroup. (a) Time-scaled maximum clade credibility tree of the Kor-O1 subgroup. Branches are colored according to their location in South Korea, and the branch thickness indicates the posterior probabilities of the inferred ancestral location. The phylogenetic tree with the virus name is available in Fig. S2. (b) Mean number of migration events of the Kor-O1 subgroup from major sources in Korea through time, which was estimated and computed by Bayesian phylogeographic analysis. (c) Transmissions of the Kor-O1 subgroup in South Korea visualized by Spred3. The color of the lines indicates the posterior probability. The diameters of the circles represent the number of branches maintaining a particular location state in each period. The video of the Kor-O1 transmission history is available on Video S1. (BS Busan, CB Chungcheongbuk-do, CN Chungcheongnam-do, DG Daegu, GB Gyeongsangbuk-do, GG Gyeonggi-do, GJ Gwangju, GN Gyeongsangnam-do, IC Incheon, JB Jeollabuk-do, JJ Jeju-do, JN Jeollanam-do, SU Seoul).

subgroup from foreign countries near the same time indicated the possibility of the multiple spread of viruses from South Korea to other nations or vice versa. The results indicate that multiple cross-country transmissions of Omicron VOC occurred, despite active border control.

Our study highlights the three Omicron subgroups that co-circulated in South Korea in different regions. These semi-separated circulations might have contributed to the continuous outbreak of each subgroup. As shown in the spread of the Alpha, Delta, and Omicron VOCs, a higher transmissible variant could replace the prevailing strain²⁶. Here, the KOR-O2 subgroup was not detected at the end of this study. However, the

From	To	Migration rates ¹ (95% HPD) ²	Migration number ³ (95% HPD)	Bayes factor	Posterior probability
GG	JB	1.32 (0–2.83)	2.31 (0–5)	131.15	92.07
	JJ	0.28 (0–1.08)	0.90 (0–1)	13.69	54.80
	DG	2.47 (0–4.85)	3.92 (2–7)	618.18	98.21
DG	SU	1.25 (0–2.84)	3.37 (1–7)	58.22	83.76
	CB	0.58 (0–1.86)	1.94 (0–5)	15.20	57.38
	CN	0.93 (0–2.62)	2.61 (0–7)	19.76	63.64
GB	GN	0.66 (0–2.25)	1.13 (0–3)	13.36	54.20
	JN	0.55 (0–1.58)	0.92 (0–1)	48.39	81.08
JB	GJ	0.60 (0–2.04)	0.36 (0–1)	19.42	63.23

Table 2. Migration rates, migration number, Bayes factor, and posterior probability of spread of the Kor-O1 subgroup in South Korea. Only well-supported viral migrations (posterior probability > 0.5 and Bayes factor > 3) are displayed in this table. *GG* Gyeonggi-do, *JB* Jeollabuk-do, *JJ* Jeju-do, *DG* Daegu, *SU* Seoul, *CB* Chungcheongbuk-do, *CN* Chungcheongnam-do, *GB* Gyeongsangbuk-do, *GN* Gyeongsangnam-do, *JN* Jeollanam-do, *GJ* Gwangju. ¹Actual migration rates were calculated as rate × indicator. ²95% HPD, 95% height posterior density. ³Number of migrations among regions in phylogenetic tree (Markov jumps) were estimated using stochastic mapping techniques implemented in the BEAST package.

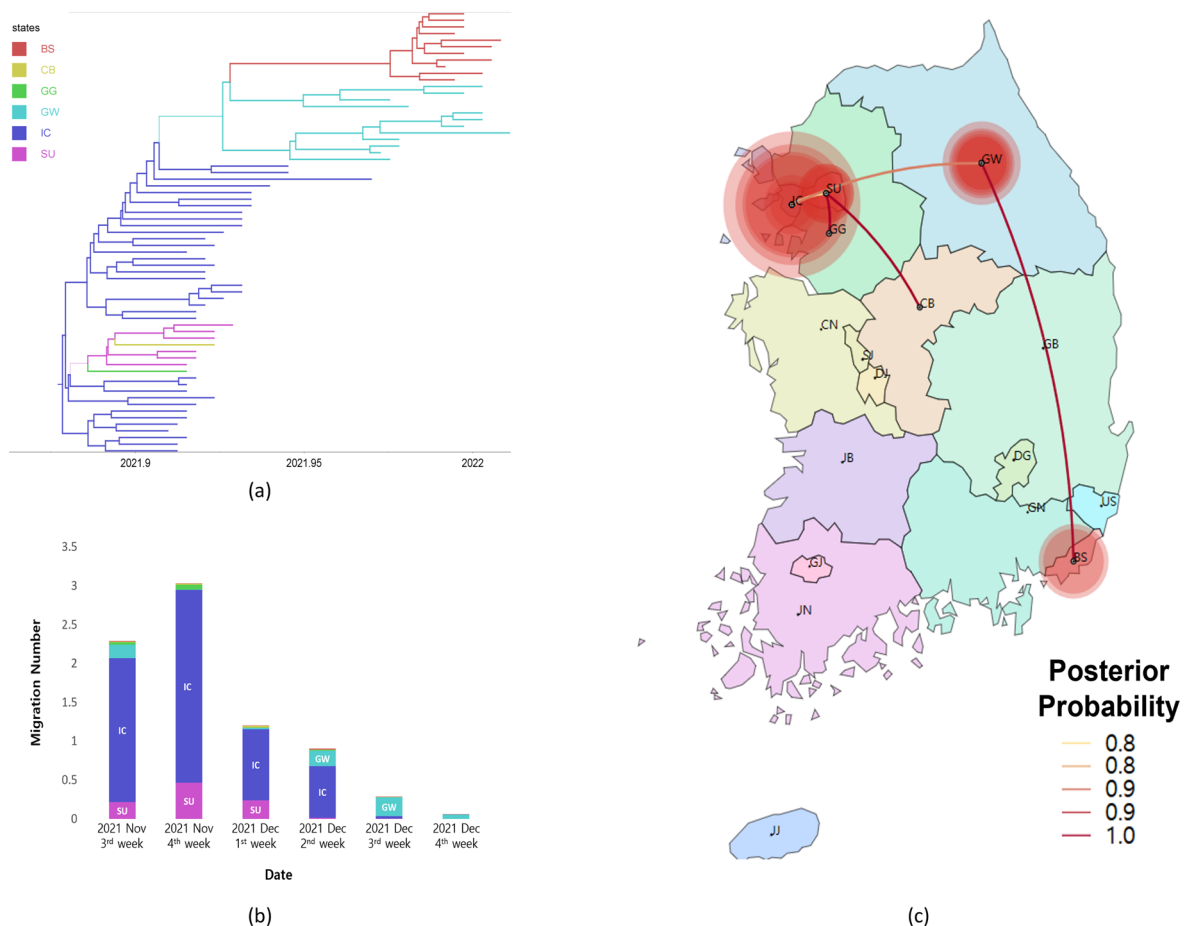


Figure 3. Phylogeographic result of the Kor-O2 subgroup. (a) Time-scaled maximum clade credibility tree of the Kor-O2 subgroup. Branches are colored according to their location in South Korea, and the branch thickness indicates the posterior probabilities of the inferred ancestral location. The phylogenetic tree with the virus name is available in Fig. S3. (b) Mean number of migration events of the Kor-O2 subgroup from major sources in Korea through time, which was estimated and computed by Bayesian phylogeographic analysis. (c) Transmissions of the Kor-O2 subgroup in South Korea visualized by SpredD3. The color of the lines indicates the posterior probability. The diameters of the circles represent the number of branches maintaining a particular location state in each period. The video of the Kor-O2 transmission history is available on video S2 (BS Busan, CB Chungcheongbuk-do, GG Gyeonggi-do, GW Gangwon-do, IC Incheon, SU Seoul).

From	To	Migration rates ¹ (95% HPD) ²	Migration number ³ (95% HPD)	Bayes factor	Posterior probability (%)
IC	SU	1.27 (0–3.34)	1.63 (0–5)	33.79	88.78
	CB	0.64 (0–2.29)	0.62 (0–1)	6.77	61.34
	GG	0.79 (0–2.42)	0.78 (0–1)	13.38	75.81
	GW	0.97 (0–2.82)	1.16 (0–3)	19.13	81.75
GW	BS	0.71 (0–2.50)	0.51 (0–1)	8.35	66.17
SU	CB	0.64 (0–2.66)	0.35 (0–1)	4.39	50.72

Table 3. Migration rates, migration number, Bayes factor, and posterior probability of spread of the Kor-O2 subgroup in South Korea. Only well-supported viral migrations (posterior probability > 0.5 and Bayes factor > 3) are displayed in this table. *IC* Incheon, *SU* Seoul, *CB* Chungcheongbuk-do, *GG* Gyeonggi-do, *GW* Gangwon-do, *BS* Busan. ¹Actual migration rates were calculated as rate × indicator. ²95% HPD, 95% height posterior density. ³Number of migrations among regions in phylogenetic tree (Markov jumps) were estimated using stochastic mapping techniques implemented in the BEAST package.

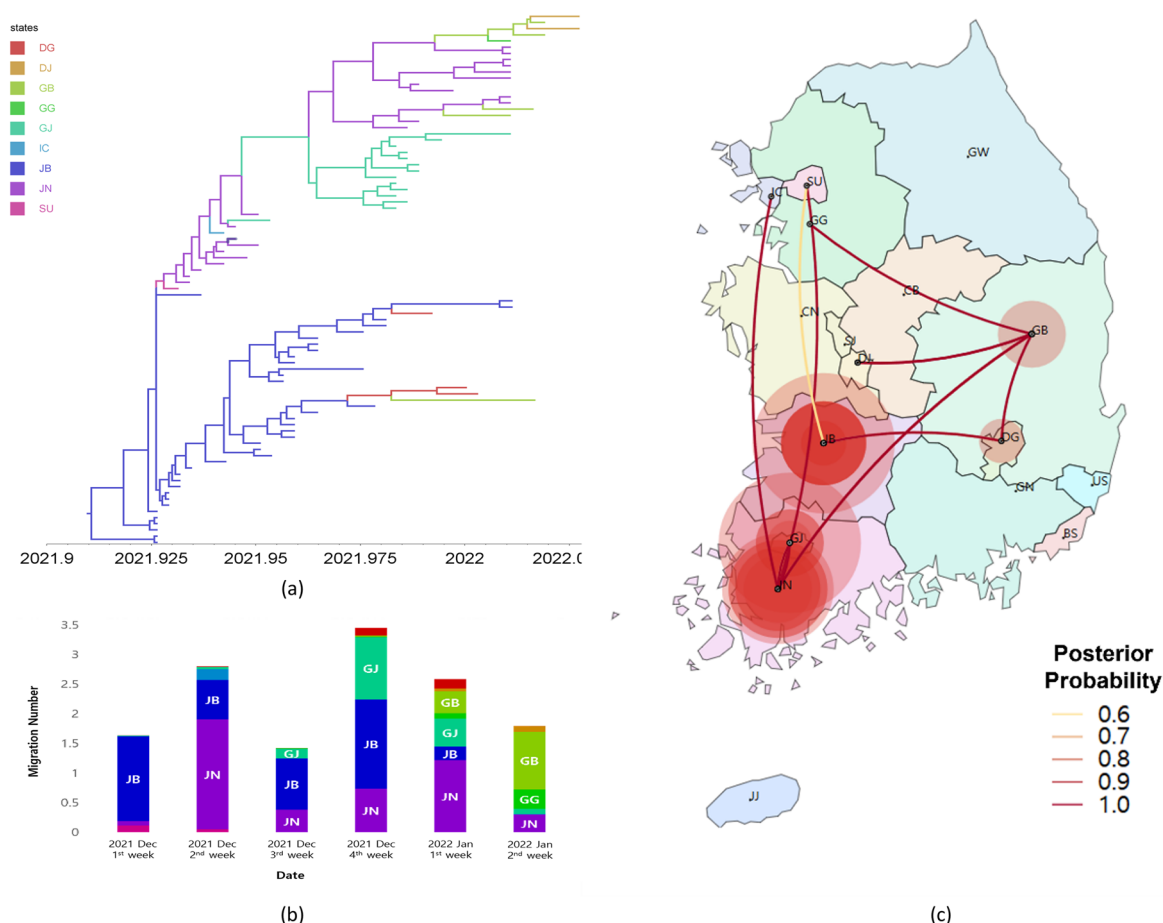


Figure 4. Phylogeographic result of the Kor-O3 subgroup. **(a)** Time-scaled maximum clade credibility tree of the Kor-O3 subgroup. Branches are colored according to their location in South Korea, and the branch thickness indicates the posterior probabilities of the inferred ancestral location. The phylogenetic tree with the virus name is available in Fig. S4. **(b)** Mean number of migration events of the Kor-O3 subgroup from major sources in Korea through time, which was estimated and computed by Bayesian phylogeographic analysis. **(c)** Transmissions of the Kor-O3 subgroup in South Korea visualized by Spred3. The color of the lines indicates the posterior probability. The diameters of the circles represent the number of branches maintaining a particular location state in each period. The video of the Kor-O3 transmission history is available on video S3 (DG Daegu, DJ Daejeon, GB Gyeongsangbuk-do, GG Gyeonggi-do, GJ Gwangju, IC Incheon, JB Jeollabuk-do, JN Jeollanam-do, SU Seoul).

From	To	Migration rates ¹ (95% HPD) ²	Migration number ³ (95% HPD)	Bayes factor	Posterior probability
JB	JN	0.95 (0–2.64)	1.48 (0–4)	32.23	81.57
	SU	0.63 (0–1.91)	0.81 (0–1)	24.34	76.97
	DG	1.16(0–2.67)	1.99 (1–2)	249.76	97.17
JN	GB	1.12 (0–3.27)	1.62 (0–4)	16.17	68.95
	GJ	0.99 (0–2.80)	1.38 (0–4)	29.71	80.31
	IC	0.88 (0–2.26)	0.97 (0–1)	113.13	93.95
GB	GG	0.69 (0–2.54)	0.51 (0–1)	9.56	56.77
	DJ	0.93 (0–2.98)	0.78 (0–2)	15.81	68.46

Table 4. Migration rates, migration number, Bayes factor, and posterior probability of spread of the Kor-O3 subgroup in South Korea. Only well-supported viral migrations (posterior probability > 0.5 and Bayes factor > 3) are displayed in this table. *JB* Jeollabuk-do, *JN* Jeollanam-do, *SU* Seoul, *DG* Daegu, *GB* Gyeongsangbuk-do, *GJ* Gwangju, *IC* Incheon, *GG* Gyeonggi-do, *DJ* Daejeon. ¹Actual migration rates were calculated as rate × indicator. ²95% HPD, 95% height posterior density. ³Number of migrations among regions in phylogenetic tree (Markov jumps) were estimated using stochastic mapping techniques implemented in the BEAST package.

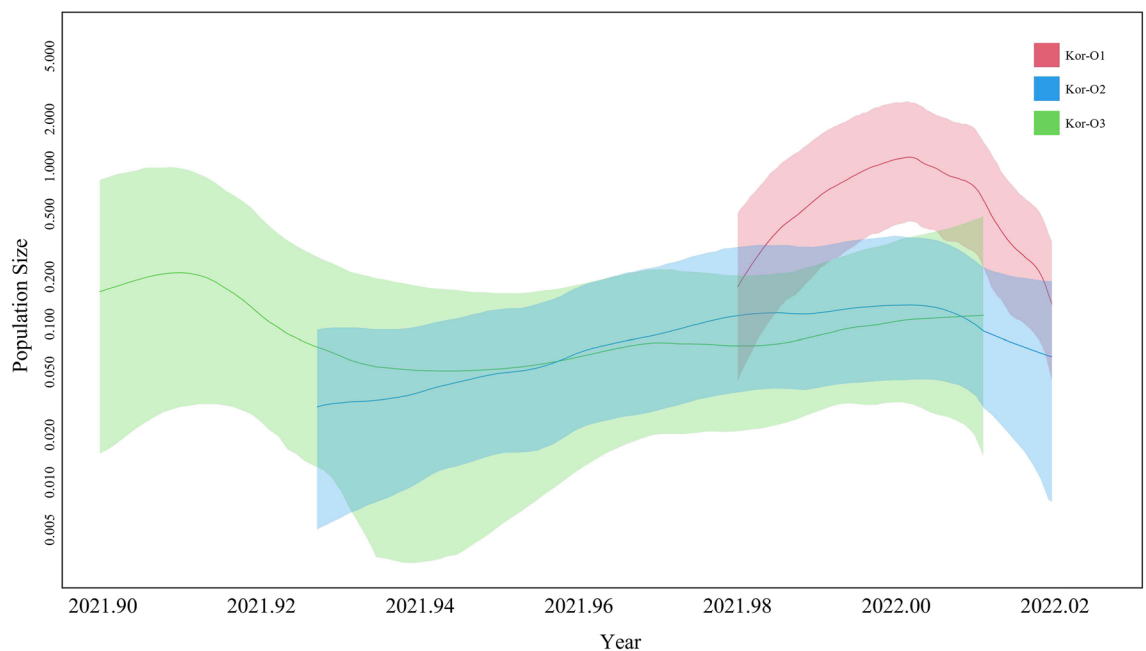


Figure 5. Viral population size change of each Omicron subgroup detected in South Korea. A line indicates the mean value of the viral population size, and 95% HPD is indicated by the solid area.

co-circulation of multiple variants should be carefully monitored. The co-circulation of different variants could be an obstacle to disease control using vaccines and therapeutics. More extended monitoring and continued genomic epidemiology will be required to identify circulating variants.

The Kor-O1 subgroup caused the largest outbreak and spread to most of the regions in South Korea. While the spread of Kor-O2 and Kor-O3 started in Incheon and Jeollabuk-do, respectively, the Kor-O1 subgroup started in Gyeonggi-do. Seoul and Gyeonggi-do are the major metropolitan areas in South Korea and have the largest population and movement of population. Outbreaks in metropolitan regions may contribute to the more frequent and rapid spread of Kor-O1 than that of other subgroups. During our investigation period, the vaccination rate in South Korea was approximately 80%; nevertheless, the Omicron variant rapidly spread and caused a large outbreak. Vaccination could not completely prevent SARS-CoV-2 infection, but the effect of vaccination on the severity and transmission of the disease should be carefully evaluated by further studies.

The ancestral reconstruction and phylogeographic analysis used in this study can be affected by sampling bias²⁷. We selected viruses for sequencing based on field epidemiological investigation results to minimize the impact of sampling bias. In addition, we did not conduct down-sampling and used all available sequences to avoid other sampling biases. However, the potential role of unsampled cases and less sampled populations cannot be entirely excluded. Owing to the large outbreak after the introduction of Omicron, sequence analysis was conducted for only approximately 3% of confirmed cases. The possible existence of undetected sub-clinical

infections also could not be analyzed. Such sampling biases and low sequencing coverage can affect the results of the phylogeographic analysis. However, this study presents the most probable transmission dynamics of the Omicron variant in South Korea by reconstructing ancestors in phylogenetic analysis, using all of the available sequence data.

As the number of Omicron cases in South Korea increases, it is essential to estimate the spread of variants using molecular epidemiology. However, the early detection of novel variants among the numerous cases can be limited, owing to the restriction of sequencing coverage. We estimate the time of introduction of the virus based on a time-scaled phylogenetic tree, and the results show that South Korea's current sequence surveillance system could detect new viral introductions within 9 to 12 days.

Omicron variant sub-lineages continue to be detected worldwide, and viruses are evolving into various sub-lineages. As SARS-CoV-2 is becoming endemic in most countries, the Korean government took steps to relax its quarantine policy and reduce the level of social distancing for a gradual return to normal life. Although the number of confirmed cases is gradually decreasing, novel variants of SARS-CoV-2 continuously threaten public health. Because multiple introductions and co-circulation of variants were detected in this study, the relaxed quarantine policy will require improvements in the variant monitoring system and control strategies.

Methods

Virus detection and whole-genome sequencing. Nasopharyngeal and oropharyngeal swabs and sputum samples were collected from symptomatic patients to detect SARS-CoV-2 by real-time reverse transcriptase (RT)-PCR. According to the manufacturer's protocol, RNA was extracted from 140 µL of the sample using a Qiagen viral RNA mini kit (Qiagen, Hilden, Germany). Real-time RT-PCR was performed on the extracted RNA and cycle threshold²⁸.

The samples were chosen for sequencing from the outbreak data based on epidemiological links. In addition, we selected samples from sporadic cases and a few random representative samples from epidemiologically linked large outbreaks. Whole-genome sequences of selected viruses were analyzed using the ARTIC protocol (<https://artic.network/ncov-2019>). Libraries were prepared using the Nextera DNA Flex Library Prep Kit (Illumina, USA), and sequencing was performed by the MiSeq instrument using the MiSeq Reagent Kit V2 (Illumina, USA).

After lineage assignment in PANGOLIN (<https://pangolin.cog-uk.io/>)²³, we obtained 582 complete genome sequences of Omicron VOC from November 24, 2021, to January 11, 2022, including 550 Korean resident cases and 32 inbound traveler cases (Table S1).

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institution and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Institutional Review Board of the Korea Disease Control and Prevention Agency (approval number: 2020-03-01-P-A) and was designated a service to public health during the pandemic. Therefore, the Institutional Review Board waived the requirement for written informed consent, as outlined in the Title Laboratory Response to COVID-19.

Identify phylogenetic subgroups. BA.1 and BA.1.1 sub-lineages of Omicron full-genome sequences, identified as of January 31, 2022, in other countries, with >29,000 nt and <1% undefined bases, were downloaded from the GISAID EpiCoV™ database (<https://www.gisaid.org/>) for reference sequences (n = 7053). For efficient computation and analysis, we selected 3330 sequences (1543 from Europe, 951 from North America, 543 from Asia, 212 from Oceania, 54 from Africa, and 27 from South America) based on the nucleotide sequence identified at the 99.97% level using the program CD-HIT²⁹. We used Nextclade v1.13.2 (<https://clades.nextstrain.org/>) for quality control (QC) of 582 sequences isolated in South Korea. Nextclade uses an algorithm that considers missing data, mixed sites, private mutations, mutation clusters, stop codons, and frameshifts in calculating a score. A total of 32 sequences that exceeded the QC score of 100 were excluded, resulting in 550 sequences remaining.

Using Geneious Prime software (<https://www.geneious.com/>), each sequence was aligned to the reference sequence Wuhan-Hu-1 (GenBank ID MN908947). Aligned sequences were trimmed to equal lengths (29,409 bp) from the start codon of ORF1ab to the stop codon of ORF10. Frame shifting insertions in <1% of sequences were considered erroneous sequencing results and were deleted through mask alignment in the Geneious software. The approximately maximum-likelihood phylogenetic tree was constructed using FastTree in Galaxy Europe (<http://usegalaxy.eu/>) under the general time reversible (GTR) model for the effective computation of big genomic data. The phylogenetic tree was visualized and annotated using iTOL (<https://itol.embl.de/>). In the FastTree result, we divided Korean sequences into three subgroups with the monophyletic cluster sharing a common ancestral node with other Korean sequences showing a local support value of >0.7. This resulted in a Kor-O1 subgroup with 273 sequences, a Kor-O2 subgroup with 68 sequences, and a Kor-O3 subgroup with 86 sequences.

To investigate the mutations of each subgroup of Korean sequences, we compared the Korean sequences with the first isolated Omicron, hCoV-19/South Africa/NICD-N20868/2021, collected on November 11, 2021, as a reference sequence.

Phylogeography and population size dynamics of each subgroup. For each Korean sub-lineage, we reconstructed the maximum-likelihood phylogenetic tree using Geneious Prime software, and the temporal signal of the dataset was analyzed using the TempEst program (<https://beast.community/tempEst>)³⁰. Outlier sequences in the root-to-tip regression were excluded. In addition, we deleted 2 sequences of inbound travelers and 35 reference foreign sequences from the data to focus on the geographical relationships in virus transmission in South Korea. The final data set of the Kor-O1 subgroup with 231 sequences, the Kor-O2 subgroup with 67 sequences, and the Kor-O3 subgroup with 86 sequences were used for time-scaled phylogeographic analysis.

To estimate the transmission and introduction of Omicron in South Korea, we constructed a time-scaled phylogenetic tree for three subgroups (Kor-O1–Kor-O3) using BEAST v1.10.4³¹. The sequences of each subgroup were coded as 17 states and other geographic locations by the administrative district in South Korea. Posterior phylogenetic tree distribution was estimated using Bayesian phylogenetic inference. The general time reversible model, along with a gamma-distributed rates (GTR + γ) nucleotide substitution model was selected. To estimate the viral population size change, the unweighted pair group method with an arithmetic starting tree model and Bayesian Skygrid coalescent tree prior was used with an uncorrelated relaxed clock model. For each subgroup, Markov-chain Monte Carlo was run in parallel for four chains, each with 150 million steps. The parameters and trees were sampled every 10,000 steps, yielding a total of 60,000 parameter states and posterior trees. The parameters were analyzed using TRACER v1.5 (<https://beast.community/tracer>)³² and burned in 10%–20% of each result. All parameters had an effective sample size of > 200 (Table S2). The log and tree files of each subgroup's resulting data were combined with LogCombiner v1.10.4 (<https://beast.community/logcombiner>). The combined log file was analyzed using TRACER 1.7.2 to reconstruct the effective population size over time that was estimated by the Gaussian Markov random field skyride model. A time-scaled maximum clade credibility tree was generated using TreeAnnotator v1.10.4 (<https://beast.community/treeannotator>) in BEAST and visualized using FigTree 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>), and the migration routes were visualized using Spread3 v0.9.6 (<https://beast.community/spread3>)³³.

The rate and number of transitions among the regions (Markov jumps) were estimated using stochastic mapping techniques implemented in the BEAST package³⁴. Posterior trees were analyzed using the PACT program (<http://www.trevorbedford.com/pact>) to compute the number of transition events between the regions through time. The posterior trees were broken into multiple temporal sections (one week per section). The migrations of the location state at the tree nodes were counted at each time window for each posterior tree.

Data availability

The sequences were shared through the GISAID EpiCoV (<https://www.gisaid.org/>) database. The list of accession numbers of South Korean SARS-CoV-2 sequences used in this study is available in Table S1.

Received: 20 June 2022; Accepted: 20 December 2022

Published online: 27 December 2022

References

- World Health Organization (WHO). *Coronavirus Disease (COVID-19) Pandemic*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (2022).
- Grint, D. J. *et al.* Case fatality risk of the SARS-CoV-2 variant of concern B. 1.1. 7 in England, 16 November to 5 February. *Euro Surveill.* **26**, 2100256 (2021).
- Davis, C. *et al.* Reduced neutralisation of the Delta (B. 1.617. 2) SARS-CoV-2 variant of concern following vaccination. *PLoS Pathog.* **17**, e1010022 (2021).
- Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science* **372**, eabg3055 (2021).
- World Health Organization (WHO). *Tracking SARS-CoV-2 Variants*. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants> (2022).
- Wise, J. Covid-19: New coronavirus variant is identified in UK. *BMJ* **371**, m4857 (2020).
- Tegally, H. *et al.* Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021).
- Fujino, T. *et al.* Novel SARS-CoV-2 variant in travelers from Brazil to Japan. *Emerg. Infect. Dis.* **27**, 1243 (2021).
- Cherian, S. *et al.* SARS-CoV-2 spike mutations, L452R, T478K, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. *Microorganisms* **9**, 1542 (2021).
- World Health Organization (WHO). *Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern*. [https://www.who.int/news/item/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news/item/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern) (2021).
- Prevention, C. o. D. C. a. *Science Brief: Omicron (B.1.1.529) Variant*. <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-omicron-variant.html#print> (2021).
- World Health Organization (WHO). *Director-General's Introductory Remarks at Member State Information Session on COVID-19–13 January 2022*. <https://www.who.int/director-general/speeches/detail/director-general-s-introductory-remarks-at-member-state-information-session-on-covid-19---13-january-2022> (2022).
- World Health Organization (WHO). *Director-General's Opening Remarks at the Media Briefing on COVID-19–25 June 2021*. <https://www.who.int/director-general/speeches/detail/director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-25-june-2021> (2021).
- He, X., Hong, W., Pan, X., Lu, G. & Wei, X. SARS-CoV-2 Omicron variant: Characteristics and prevention. *MedComm* **2**, 838–845 (2021).
- Wilhelm, A. *et al.* Reduced neutralization of SARS-CoV-2 omicron variant by vaccine sera and monoclonal antibodies. *medRxiv*. <https://doi.org/10.1101/2021.12.07.21267432> (2021).
- Chen, J., Wang, R., Gilby, N. B. & Wei, G.-W. Omicron variant (B. 1.1. 529): Infectivity, vaccine breakthrough, and antibody resistance. *J. Chem. Inf. Model.* **62**, 42–422 (2022).
- World Health Organization (WHO). *The Republic of Korea, Current COVID-19 Situation*. <https://www.who.int/countries/kor/> (2022).
- Bansal, K. & Kumar, S. Mutational cascade of SARS-CoV-2 leading to evolution and emergence of omicron variant. *Virus Res.* **315**, 198765 (2022).
- Papanikolaou, V. *et al.* From Delta to Omicron: S1-RBD/S2 mutation/deletion equilibrium in SARS-CoV-2 defined variants. *Gene* **814**, 146134 (2022).
- Piccoli, L. *et al.* Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042 (2020).
- Tian, F. *et al.* N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *Elife* **10**, e69091 (2021).
- Lupala, C. S., Ye, Y., Chen, H., Su, X.-D. & Liu, H. Mutations on RBD of SARS-CoV-2 Omicron variant result in stronger binding to human ACE2 receptor. *Biochem. Biophys. Res. Commun.* **590**, 34–41 (2022).
- O'Toole, Á. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* **7**, veab064 (2021).

24. World Health Organization (WHO). *Weekly Epidemiological Update on COVID-19–29 March 2022*, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20220329-weekly-epi-update_85.pdf?sfvrsn=16aaa557_4&download=true (2022).
25. Mahase, E. Covid-19: What do we know about the delta omicron recombinant variant?. *BMJ* **376**, o792 (2022).
26. Mishra, S. *et al.* Changing composition of SARS-CoV-2 lineages and rise of Delta variant in England. *EClinicalMedicine* **39**, 101064 (2021).
27. Vakulenko, Y., Deviatkin, A. & Lukashev, A. The effect of sample bias and experimental artefacts on the statistical phylogenetic analysis of picornaviruses. *Viruses* **11**, 1032 (2019).
28. Kim, J.-M. *et al.* Identification of coronavirus isolated from a patient in Korea with COVID-19. *Osong Public Health Res. Perspect.* **11**, 3 (2020).
29. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158> (2006).
30. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
31. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
32. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904. <https://doi.org/10.1093/sysbio/syy032> (2018).
33. Bielejec, F. *et al.* SpreaD3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol. Biol. Evol.* **33**, 2167–2169 (2016).
34. Minin, V. N. & Suchard, M. A. Fast, accurate and simulation-free stochastic mapping. *Phil. Trans. R. Soc. Lond. B* **363**, 3985–3995 (2008).

Acknowledgements

We gratefully acknowledge the authors of the originating and submitting laboratories who have deposited and shared genome data on the GISAID EpiCoV (<https://www.gisaid.org/>) database. The GISAID acknowledgment tables for the sequences used in this study can be found at EPI_SET-ID = EPI_SET_20220323mn (a full list of GISAID accession numbers is available for download as an archived file on this site). This study was funded by the Korea Disease Control and Prevention Agency (No. 6300-6331-301) and partially supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry (IPET) through the Animal Disease Management Technology Advancement Support Program, funded by the Ministry of Agriculture, Food and Rural Affairs (1220160-2).

Author contributions

J.-M.K., A.-K.P., H.L., I.-H.K., J.-A.K., and C.-Y.L. contributed to the sequencing analysis and data curation. D.-W.L., J.-M.K., A.-K.P., D.-W.K., J.-Y.K., and N.L. analyzed the sequence data and conducted genomic epidemiology and visualization of the results. D.-W.L. wrote the original draft of the main manuscript. J.-M.K., J.-H.K., and E.-J.K. conceptualized and supervised the study. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-26803-w>.

Correspondence and requests for materials should be addressed to J.-H.K. or E.-J.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022