


RESEARCH ARTICLE

Omicron variant genome evolution and phylogenetics

Mahmoud Kandeel^{1,2}  | Maged E. M. Mohamed^{3,4} | Hany M. Abd El-Lateef^{5,6} |
Katharigatta N. Venugopala^{3,7} | Hossam S. El-Beltagi^{8,9} 

¹Department of Biomedical Sciences, College of Veterinary Medicine, King Faisal University, Al-Hofuf, Saudi Arabia

²Department of Pharmacology, Faculty of Veterinary Medicine, Kafrelshikh University, Kafrelshikh, Egypt

³Department of Pharmaceutical Sciences, College of Clinical Pharmacy, King Faisal University, Al-Ahsa, Saudi Arabia

⁴Department of Pharmacognosy, College of Pharmacy, Zagazig University, Zagazig, Egypt

⁵Department of Chemistry, College of Science, King Faisal University, Al-Ahsa, Saudi Arabia

⁶Department of Chemistry, Faculty of Science, Sohag University, Sohag, Egypt

⁷Department of Biotechnology and Food Science, Faculty of Applied Sciences, Durban University of Technology, Durban, South Africa

⁸Agricultural Biotechnology Department, College of Agriculture and Food Sciences, King Faisal University, Al-Ahsa, Saudi Arabia

⁹Biochemistry Department, Faculty of Agriculture, Cairo University, Giza, Egypt

Correspondence

Mahmoud Kandeel, Department of Biomedical Sciences, College of Veterinary Medicine, King Faisal University, Al-Hofuf, 31982 Al-Ahsa, Saudi Arabia.
Email: mkandeel@kfu.edu.sa

Funding information

Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia

Abstract

Following the discovery of the SARS-CoV-2 Omicron variant (B.1.1.529), the global COVID-19 outbreak has resurfaced after appearing to be relentlessly spreading over the past 2 years. This new variant showed marked degree of mutation, compared with the previous SARS-CoV-2 variants. This study investigates the evolutionary links between Omicron variant and recently emerged SARS-CoV-2 variants. The entire genome sequences of SARS-CoV-2 variants were obtained, aligned using Clustal Omega, pairwise comparison was computed, differences, identity percent, gaps, and mutations were noted, and the identity matrix was generated. The phylogenetics of Omicron variants were determined using a variety of evolutionary substitution models. The ultrametric and metric clustering methods, such as UPGMA and neighbor-joining (NJ), using nucleotide substitution models that allowed the inclusion of nucleotide transitions and transversions as Kimura 80 models, revealed that the Omicron variant forms a new monophyletic clade that is distant from other SARS-CoV-2 variants. In contrast, the NJ method using a basic nucleotide substitution model such as Jukes–Cantor revealed a close relationship between the Omicron variant and the recently evolved Alpha variant. Based on the percentage of sequence identity, the closest variants were in the following order: Omicron, Alpha, Gamma, Delta, Beta, Mu, and then the SARS-CoV-2 USA isolate. A genome alignment with other variants indicated the greatest number of gaps in the Omicron variant's genome ranging from 43 to 63 gaps. It is possible, given their close relationship to the Alpha variety, that Omicron has been around for much longer than predicted, even though they created a separate monophyletic group. Sequencing initiatives in a systematic and comprehensive manner is highly recommended to study the evolution and mutations of the virus.

1 | INTRODUCTION

Omicron was discovered in Botswana in early November. South Africa reported it to the World Health Organization on November 24, 2021, and it was designated as a variant of concern (VOC) on November 26, 2021.¹ Omicron contains a huge number of previously documented mutations in other VOCs, including at least 32

mutations in the spike protein alone, compared to the 16 mutations in the already highly infectious delta version, as well as other proteins required for viral replication such as NSP12 and NSP14.²

Several assumptions postulated the Omicron variant's probable emerging pattern, including (1) the potential circulation among chronically infected patients. (2) The introduction of the new variant in some South African countries during the winter wave, which went

unreported due to lower genome sequencing in some countries. (3) Spike mutations may have increased the Spike's ability to connect to the ACE2 receptor on host cells. (4) A hidden animal reservoir could be the culprit, owing to the high number of mutations found in the Omicron form. (5) Africa's low immunization rate may have contributed to the establishment of the Omicron variant.³

Because of the emergent nature of the Omicron variation, various concerns have been raised, including the source of emergence, the effect of mutations in Omicron in the response to vaccinations, the influence of mutations on modulation of host immunity, clinical data, Omicron spreading potency and lethality. In this study, an attempt was made to trace the phylogenetic relationships of the Omicron genome. To achieve the best fit of alignment of whole genomes, many methodologies were used.

2 | MATERIALS AND METHODS

2.1 | Collection of genomes and analytical programs

The genomes of CoV variants were retrieved from GISAID (<https://www.gisaid.org/>).⁴ The basic information of the used genomes are provided in Table 1.

The CLC Genomics Workbench 12.0 (QIAGEN)⁵ and Geneious prime⁶ software were used to handle the sequences.

2.2 | Alignment of genomes

The FASTA files containing entire genomes were uploaded to the Clustal Omega website at the European Bioinformatics Institute using the default parameters, and the results were analyzed. Using in-house software, the output files were imported, and the pairwise comparison matrix was produced. Differences and identity percent were calculated, as well as gaps and mutations were noted, and the identity matrix was generated.

2.3 | Phylogenetics

The creation of the phylogenetic tree was accomplished through the use of two algorithms: the neighbor-joining (NJ) method or the UPMA method. For distance measuring, the Jukes-Cantor (JC), Kimura 80 substitution models⁷ were employed. Bootstrap resampling with 100 replicates was applied.

3 | RESULTS AND DISCUSSION

The first sequenced genome of Omicron variant was used to trace its phylogenetic relations with other SARS-CoV-2 variants (Table 1). The sample was collected in November 11, 2021 in Botswana, the

TABLE 1 The list of SARS-CoV-2 variants used in this study, including virus name, accession numbers at GISAID website, collection date, submission date, host, location, and originating lab

Variant	Virus name	Accession ID	Collection date	Submission Date	Length	Host	Location	Originating lab
Omicron	hCoV-19/Botswana/R40B59_BHP_3321001248/2021	EPI_ISL_6640916	2021-11-11	2021-11-23	29684	Human	Africa/Botswana	Botswana Harvard
Alpha	hCoV-19/Japan/HiroYH02/2021	EPI_ISL_6756515	2021-08-02	2021-11-26	29763	Human	Asia/Japan/Tokyo	The Virology lab, Hiroshima University
Beta	hCoV-19/Japan/TY27-328-P0/2021	EPI_ISL_5416540	2021-07	2021-10-21	29764	Human	Asia/Japan/Tokyo	Department of Virology I, National Institute of Infectious Diseases
Gamma	hCoV-19/Japan/TY30-974-P0/2021	EPI_ISL_6228367	2021-08	2021-11-12	29768	Human	Asia/Japan/Tokyo	Department of Virology I, National Institute of Infectious Diseases
Delta	hCoV-19/Japan/TKY501334/2021	EPI_ISL_6832166	2021-10-25	2021-11-29	29769	Human	Asia/Japan/Tokyo	Tokyo Metropolitan Institute of Public Health
hCoV	hCoV-19/USA/WA-VAP5-531-1721017846/2021	EPI_ISL_6910522	2021-09-09	2021-12-01	29807	Human	North America/USA/Washington	Seattle VA Medical Center
Mu GH	hCoV-19/Japan/TY27-063-P0/2021	EPI_ISL_4470504	2021-07	2021-09-24	29781	Human	Asia/Japan	Department of Virology I, National Institute of Infectious Diseases
GH490R	hCoV-19/France/BRE-IPP36497/2021	EPI_ISL_6887009	2021-11-05	2021-11-30	29724	Human	Europe/France/Bretagne/Chateaulin	Labo Analyses Med

	1	2	3	4	5	6	7	
Omicron_variant_hCoV-19_Botswana_R40B59_BHP_3321001248_2021_EPI_ISL_6640916_selection	1		141	138	109	130	140	132
hCoV-19_USA_WA-VAPS-531-1721017846_2021_EPI_ISL_6910522_	2	99.53		19	95	100	104	93
Delta_variant_hCoV-19_Japan_TKYS01334_2021_EPI_ISL_6832166_	3	99.54	99.94		94	99	103	90
Alpha_variant_hCoV-19_Japan_HiroYH02_2021_EPI_ISL_6756515_	4	99.63	99.68	99.68		74	84	92
Gamma_variant_hCoV-19_Japan_TY30-974-P0_2021_EPI_ISL_6228367_-1_1-29768	5	99.56	99.66	99.67	99.75		70	84
Beta_variant_hCoV-19_Japan_TY27-328-P0_2021_EPI_ISL_5416540_	6	99.53	99.65	99.65	99.72	99.76		76
Mu_GH_variant_hCoV-19_Japan_TY27-063-P0_2021_EPI_ISL_4470504_	7	99.56	99.69	99.70	99.69	99.72	99.74	

FIGURE 1 Pairwise comparative matrix of Omicron with SARS-CoV-2 variants. The upper diagonal panel is the differences in nucleotide composition between two variants. The lower diagonal panel is the percent identity. The matrix color ranges from red (larger differences) to blue (smaller differences)

	1	2	3	4	5	6	7	
Omicron_variant_hCoV-19_Botswana_R40B59_BHP_3321001248_2021_EPI_ISL_6640916_selection	1		63	63	43	53	66	63
hCoV-19_USA_WA-VAPS-531-1721017846_2021_EPI_ISL_6910522_	2	29583		0	30	26	39	28
Delta_variant_hCoV-19_Japan_TKYS01334_2021_EPI_ISL_6832166_	3	29586	29682		30	26	39	28
Alpha_variant_hCoV-19_Japan_HiroYH02_2021_EPI_ISL_6756515_	4	29602	29618	29619		14	27	34
Gamma_variant_hCoV-19_Japan_TY30-974-P0_2021_EPI_ISL_6228367_-1_1-29768	5	29593	29618	29619	29635		13	20
Beta_variant_hCoV-19_Japan_TY27-328-P0_2021_EPI_ISL_5416540_	6	29583	29614	29615	29625	29639		25
Mu_GH_variant_hCoV-19_Japan_TY27-063-P0_2021_EPI_ISL_4470504_	7	29598	29628	29631	29629	29637	29641	

FIGURE 2 Pairwise comparative matrix of Omicron with SARS-CoV-2 variants. The upper diagonal panel is the number of gaps. The lower diagonal panel is the number of identical nucleotides in the genome. The matrix color ranges from red (larger differences) to blue (smaller differences)

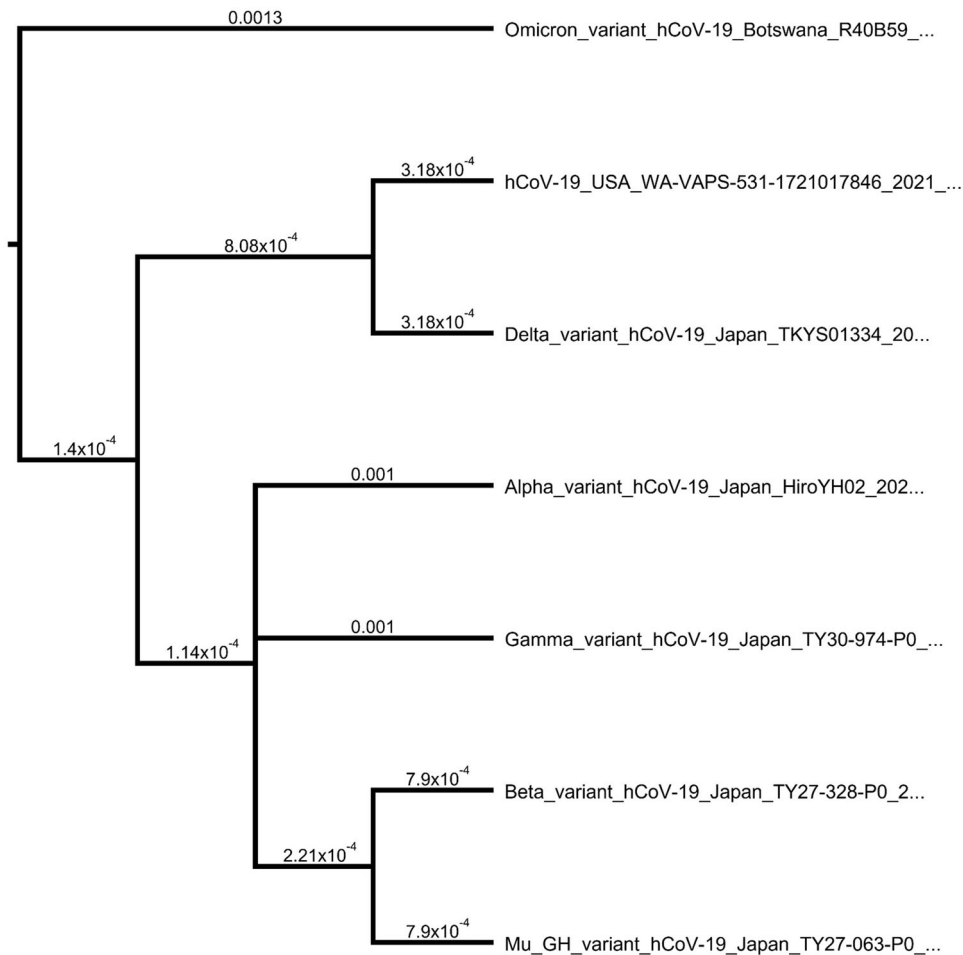


FIGURE 3 Phylogenetic tree using neighbor-joining method and Tamura substitution model. The figure was generated by Geneious prime software

genome sequence was submitted on November 23, 2021 with accession no. EPI_ISL_6640916. The other variants comprised Alpha, Beta, Gamma, Delta, Mu, GH49R, in addition to the SARS-CoV-2 USA isolate.

The FASTA files of variants were aligned by Clustal Omega. Pairwise comparison matrix revealed that the largest number of mutations were recorded with the Omicron variant. In comparison to

other variants, the number of nucleotide changes in the Omicron genome was in the following order: SARS-CoV-2 USA isolate > Mu variant > Beta variant > Delta variant > Gamma variant > Alpha variant > Omicron variant, with 141, 140, 138, 132, 130, and 109 mutations, respectively (Figure 1). The Alpha variant had the greatest identity rate with Omicron variant (99.63%), followed by Gamma and Mu variants (99.56%). The SARS-CoV-2 USA isolate has the lowest

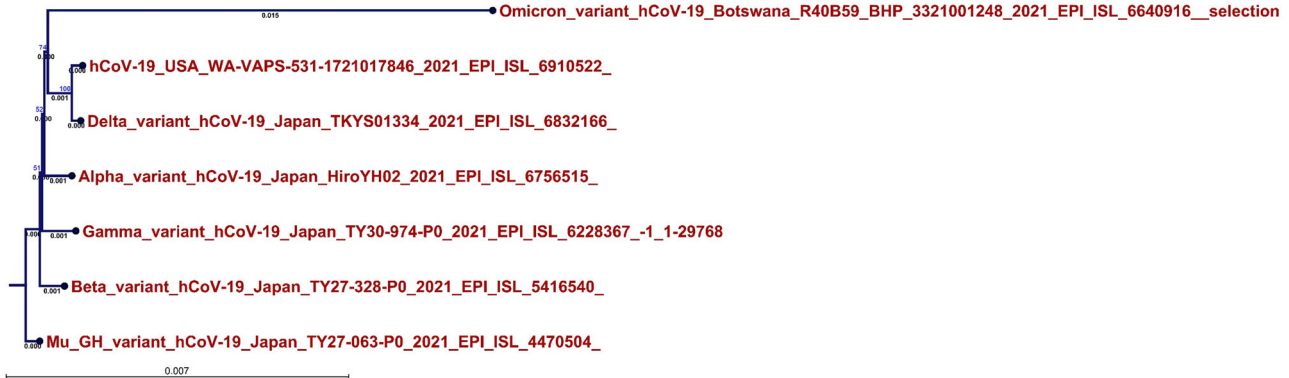


FIGURE 4 Phylogenetic tree using neighbor-joining method and Kimura 80 substitution model. The figure was generated by CLC genomics software



FIGURE 5 Phylogenetic tree using UPGMA method and Kimura 80 substitution model. The figure was generated by CLC genomics software



FIGURE 6 Phylogenetic tree using UPGMA method and Jukes-Cantor substitution model. The figure was generated by CLC genomics software

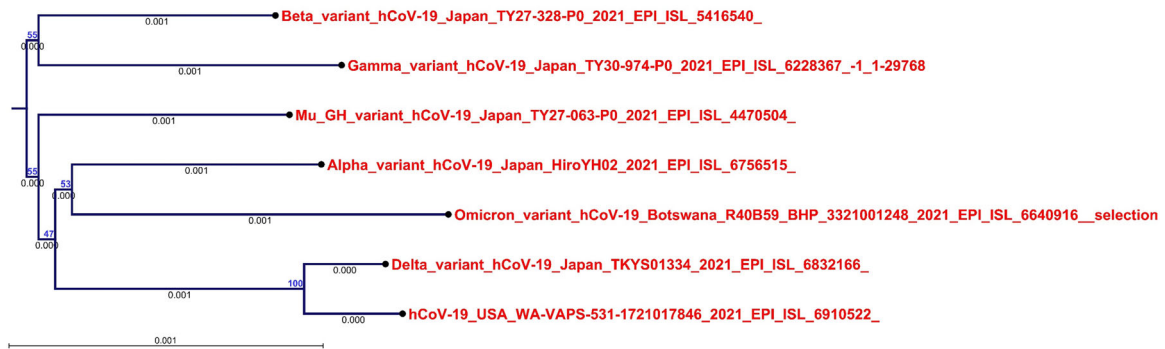


FIGURE 7 Phylogenetic tree using neighbor-joining method and Jukes–Cantor substitution model. The figure was generated by CLC genomics software

identity (99.53%). Furthermore, Omicron variant showed the greatest number of gaps during genome alignment with other viruses, ranging from 43 to 63 gaps (Figure 2).

The phylogenetic analysis of Omicron variant is provided in Figures 3–7. The NJ/Tamura (Figure 3), NJ/Kimura 80 (Figure 4), UPGMA/Kimura 80 (Figure 5), and UPGMA/JC (Figure 6) revealed that Omicron variant formed a new emergent group that was not originating with other variants. In contrast, the NJ/JC (Figure 7) revealed the close relation of Omicron variant with Alpha variant.

Bioinformatics and phylogeny tools are gold standards in microbial evolution and drug discovery against selected molecular targets.^{8–10} Analysis of SARS-CoV-2 genome constituents highlighted the forces affecting virus evolution.¹¹ In this study, we used a combination of tools to get insights into the evolution of Omicron variant. The UPGMA approach assumes that all lineages evolve at the same rate, and the mutation rate is not taken into account during tree construction. The tree construction depends on the pairwise distance. In contrast, the NJ considers the evolution rate during tree construction. The JC model of evolution considers all possible changes to nucleotides occurring with equal rates. While Kimura model assumes considers the transitions (e.g., changes of A to T or G to C) and transversions (e.g., changes from purines to pyrimidines). In virus evolution, a single evolution model cannot be assumed due to the complexity of virus evolution and variations even within single genes.¹² A NJ tree is expected to be insensitive to tree topology in the JC model, and a NJJC tree is thought to provide a good estimate of tree topology.

Based on sequence alignment differences and gaps, the Alpha variant has the fewest nucleotide alterations when compared to the Omicron variant. The close relationship between Omicron and the Alpha variant may indicate that the Omicron variant was in circulation for a lengthy period before it was discovered. The building of a phylogenetic tree using ultrametric distances between variants and equivalent evolution rates among branches using the UPGMA method revealed that the Omicron variant is phylogenetically distant from other variants, producing a monophyletic clade. Since UPGMA method merges pairs of sequences with small distance together, it was evident that Omicron variant was greatly distant from other

variants. The NJ method based on Kimura 80, which takes into account mutation rates as well as nucleotide transitions and transversions, ensured that the Omicron variant was significantly different from other variants and formed a distant monophyletic class. In contrast, a close relationship with the Alpha variant was obtained using a simplified model of JC replacement with identical likely nucleotide mutation rate.

ACKNOWLEDGMENT

This project is funded by the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this study work through the project number IFT20203.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

All data are within the manuscript.

ORCID

Mahmoud Kandeel  <http://orcid.org/0000-0003-3668-5147>

Hossam S. El-Beltagi  <http://orcid.org/0000-0003-4433-2034>

REFERENCES

- World Health Organization Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern. Accessed November 26, 2021. [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)
- Gao SJ, Guo H, Luo G. Omicron variant (B.1.1.529) of SARS-CoV-2, a global urgent public health alert!. *J Med Virol*. 2021. <https://onlinelibrary.wiley.com/doi/10.1002/jmv.27491>
- Kupferschmidt K. Where did 'weird' Omicron come from? *Science*. 2021;374(6572):1179.
- Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Euro Surveill*. 2017;22(13):30494.
- CLC Genomics Workbench 12.0 (QIAGEN). Accessed December 5, 2021. <https://digitalinsights.qiagen.com/>
- Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–1649.

7. Tamura K. Model selection in the estimation of the number of nucleotide substitutions. *Mol Biol Evol.* 1994;11(1):154-157.
8. Kandeel M, Albusadah K, Alaydaa SH, Albokhadaim I, Alhojaily S, Marzok M. Insulin from human, camel and farm animals: comparative bioinformatics and molecular dynamics studies. *Insulin.* 2021. http://www.pvj.com.pk/in_press/21-193.pdf
9. Kandeel M, Yamamoto M, Park BK, et al. Discovery of new potent anti-MERS CoV fusion inhibitors. *Front Pharmacol.* 2021;12:1241.
10. Aya-T MKE, Kwon H-J, Al-Nazawi M. MERS-CoV inhibitor peptides. United States Patent Office, Patent no. US10975126B1. 2021.
11. Kandeel M, Ibrahim A, Fayez M, Al-Nazawi M. From SARS and MERS CoVs to SARS-CoV-2: moving toward more biased codon usage in viral structural and nonstructural genes. *J Med Virol.* 2020;92(6):660-666.
12. Posada D, Crandall KA. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol Biol Evol.* 2001;18(6):897-906.

How to cite this article: Kandeel M, Mohamed MEM, Abd El-Lateef HM, Venugopala KN, El-Beltagi HS. Omicron variant genome evolution and phylogenetics. *J Med Virol.* 2022;94:1627-1632. doi:10.1002/jmv.27515