

# **Research** Article

# Whole-Genome Comparison of Representatives of All Variants of SARS-CoV-2, Including Subvariant BA.2 and the GKA Clade

Ida B. K. Suardana (),<sup>1</sup> Bayu K. Mahardika (),<sup>2</sup> Made Pharmawati (),<sup>3</sup> Putu H. Sudipa (),<sup>4</sup> Tri K. Sari (),<sup>1</sup> Nyoman B. Mahendra (),<sup>5</sup> and Gusti N. Mahardika (),<sup>1,2</sup>

<sup>1</sup>Virology Laboratory, The Faculty of Veterinary Medicine, Udayana University, Denpasar, Bali, Indonesia

<sup>2</sup>The Animal Biomedical and Molecular Biology Laboratory, Udayana University, Jl. Sesetan-Markisa 6A, Denpasar 80223, Bali, Indonesia

<sup>3</sup>The Biology Study Program, The Faculty of Mathematic and Natural Science, Udayana University, Kampus Bukit Jimbaran, Badung, Bali, Indonesia

<sup>4</sup>Veterinary Bacteriology and Mycology Laboratory, The Faculty of Veterinary Medicine, Udayana University, Denpasar, Bali, Indonesia

<sup>5</sup>The Department of Obstetrics and Genecology, The Faculty of Medicine, Udayana University, Kuta Selatan, Bali, Indonesia

Correspondence should be addressed to Gusti N. Mahardika; gnmahardika@unud.ac.id

Received 5 May 2022; Revised 10 December 2022; Accepted 17 February 2023; Published 9 March 2023

Academic Editor: Majid Jabir

Copyright © 2023 Ida B. K. Suardana et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since its discovery at the end of 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has rapidly evolved into many variants, including the subvariant BA.2 and the GKA clade. Genomic clarification is needed for better management of the current pandemic as well as the possible reemergence of novel variants. The sequence of the reference genome Wuhan-Hu-1 and approximately 20 representatives of each variant were downloaded from GenBank and GISAID. Two representatives with no track of in-definitive nucleotides were selected. The sequences were aligned using muscle. The location of insertion/deletion (indel) in the genome was mapped following the open reading frame (ORF) of Wuhan-Hu-1. The phylogeny of the spike protein coding region was constructed using the maximum likelihood method. Amino acid substitutions in all ORFs were analyzed separately. There are two indel sites in ORF1AB, eight in spike, and one each in ORF3A, matrix (MA), nucleoprotein (NP), and the 3'-untranslated regions (3'UTR). Some indel sites and residues/substitutions are not unique, and some are variant-specific. The phylogeny shows that Omicron, Deltacron, and BA2 are clustered together and separated from other variants with 100% bootstrap support. In conclusion, whole-genome comparison of representatives of all variants revealed indel patterns that are specific to SARS-CoV-2 variants or subvariants. Finally, the higher transmissibility of BA.2 might be due at least in part to the 48 nucleotide deletions in the 3'UTR, while the seem-to-be extinction of GKA clade is due to the lack of genetic advantages as a consequence of amino acid substitutions in various genes.

# 1. Introduction

The rapid evolution of severe acute respiratory syndrome 2 virus (SARS-CoV-2), the causative agent of the coronavirus 2019 (COVID-19) pandemic, requires immediate scientific clarification to better manage the current pandemic and to serve as a reference for the possible emergence of novel variants. Since it was discovered at the end of 2019 [1], the

original virus has evolved into many variants, which could be attributed to specific clinical consequences [2]. Established variants of concern (VOCs) are Alpha, Beta, Gamma, Delta, and Omicron; Lambda and Mu are variants of interest (VOIs), and GH/490R is a variant under monitoring (VUM). Some other clusters of viruses that are closely related to Omicron, the so-called BA.2 and GKA clades, which carry molecular markers of Delta and Omicron variants, require special attention. Originally, detection of the GKA clade, popularly known as Deltacron, led to arguments that it may be the result of a sequencing error [3]. However, there is an increasing number of SARS-CoV-2 whole-genome sequences labelled as clade GKA with the notification "this submission requires investigation! It appears to contain markers of multiple lineages from both Delta and Omicron variants" in the database.

The capacity to spread globally as well as other biological properties of each variant must be encoded, at least in part, by its genome. Whole-genome comparison should also be able to confirm the establishment of the BA.2 subvariant and GKA clade. Based on previous publications [4, 5], the whole genome organization of SARS-CoV-2 after the ORF annotation of Wuhan-Hu-1 and adding 5'- and 3'UTR and intergenic sequence (IGS), is 5'UTR-ORF1AB- IGS-Spike-IGS-ORF3A-ORF3B-IGS-Protein E-IGS-membrane (MA)-IGS-ORF6-ORF7A-ORF7B-ORF8-IGS-Nucleoprotein NP-ORF-10-3'UTR. Intergenic sequences (IGSs) have been identified previously [6–8].

Large submissions of whole genome sequences pose a major computational challenge, and some portions of submitted sequences contain a long track of nondefinitive nucleotides. Here, we identify insertion/deletion (indel) and amino acid substitution patterns in the whole genome of representative variants, including the BA.2 subvariant and the GKA clade.

#### 2. Materials and Methods

The sequence of the reference genome of SARS-CoV-2 strain Wuhan-Hu-1 (accession number NC\_045512) was downloaded from GenBank. Ten to 20 complete sequences of each definitive variant as well as the subvariant BA.2 and GKA clade were selected randomly from GISAID and downloaded. Two representatives of each variant with no undetermined nucleotide of "N" or other IUPAC nucleotide codes were selected. The dataset identifier is EPI\_SET ID: EPI\_SET\_230223kx; doi: 10.55876/gis8.230223kx. Sequences with a single N or in-definitive nucleotide were accepted. The sequences were aligned using muscle in MEGA-X software [9]. The locations of deletions/insertions in the genome of SARS-CoV-2 were mapped following the open reading frame of Wuhan-Hu-1, as available in the GenBank file. ORFs were analyzed separately to determine the effects of mutations and deletion/insertion.

Using the corresponding ORF of the coding region of Wuhan-Hu-1, the first 15 nucleotides of the 5'-terminus were searched, and the sequence prior to the marked sequence was deleted. The last 15 nucleotides of Wuhan-Hu-1 were used. The selected sequences were translated into amino acid sequences and aligned using MEGA-X software [9]. Using the same software, the data were exported in Mega format and analyzed further for polymorphic amino acids. We identified amino acids that were consistently substituted from Wuhan-Hu-1 across all variants and amino acids shared by the Omicron, BA2, and GKA lineages, Omicron and BA.2, Delta and GKA. The final fasta file of the data set is available in Supplementary Material 1.

The phylogeny of the spike protein coding region of the representatives of variants was constructed using the maximum likelihood method and JTT matrix-based model [10] conducted in MEGA-X software [9]. The phylogenetic tree was rooted to Wuhan-Hu-1 sequence.

#### 3. Results

The indel pattern and its location in the whole genome of representatives of various variants of SARS-CoV-2 are presented in Table 1. There are two indel sites in ORF1AB, eight in spike, and one each in ORF3A, MA, NP, and 3'UTR. No indel occurs in IGS. Some indel sites are not unique, as they occur in more than one variant. D21605-21613 is unique to Omicron and DA.2 variants, and D21965-21967 is unique to the Alpha variant; I21968-21971 and D26143-26146 are unique to the Mu variant, and D28351-28359 is unique to the Omicron and DA.2 variants; and D29723-29748 is unique to the DA.2 variant. Some indels occur in one representative of the variant.

All polymorphic amino acids of all proteins of two representatives of each variant of SARS-CoV-2 are listed in Supplementary Material 2. A summary of unique amino acids across entire genes in at least one of the representative strains of SARS-CoV-2 variants is presented in Table 2. Amino acids consistently substituted from Wuhan-Hu-1 across all variants are ORF1AB P4715L/F and spike D618G. Regarding the amino acids shared by Omicron, BA2, and GKA, there are 10 in ORF1AB, 21 in spike, and one each in ORF3A, ORF6, and NP. Three deletions in NP are unique to Omicron and BA.2. Exclusive to Delta and GKA are two deletions in ORF1AB, four in spike, three in NP, and two in ORF7A and ORF8; spike amino acids shared by Delta, Omicron, BA.2, and GKA occur only once. Three insertions, Ins216E, Ins217P, and Ins218E, are unique to Omicron and GKA. GH/490 harbors 16 variants specific to Wuhan-Hu-1, namely, five in ORF1AB, eight in spike, two in NP, and one in ORF3A. Unique amino acid substitutions from Wuhan-Hu-1 to GKA clade were nine amino acids in ORF1AB, namely, E352D, A1306S, P2046L, A2529V, I2820V, V2930L, T3646A, P4715F, and A6319V, one in NP, namely, G215C, two in ORF3A, namely, S92L and D155Y, and one in ORF7B of T40I.

The topology of phylogenetic analysis of the spike protein gene of two representatives of each variant of SARS-CoV-2 is presented in Figure 1. The phylogeny shows that Omicron, Deltacron, and BA2 are clustered together and separated from other variants with 100% bootstrap support.

#### 4. Discussion

The genetic diversity of coronaviruses occurs through mutation and recombination, as it has been described for SARS-CoV-2 too [11]. Although the RNA-dependent RNA polymerases of coronaviruses possess proof-reading capacity [12], the virus still undergoes mutation, which might lead to amino acid replacement. Such changes impact the biology of the virus as well as the clinical manifestation of its infection. Recombination involves viral RNA merging with other

		Tabli	3 1: Insertion	ı/deletion p:	attern and i	ts location :	at the whol	e genome o	f representa	ttives of va	rious varian	ts of SARS-	CoV-2.		
	Position*	D6485- 6487	D11260- 11268	D21605- 21613	D21959- 21967	D21965- 21967	121968- 21971	D22005- 22010	D22170- 22172	I22181- 22189	D22266- 22274	D26143- 26146	D27055- 27067	D28351- 28359	D29723- 29748
No	Location at the genome <sup>**</sup> Virus strain	ORF1AB	ORF1AB	Spike	Spike	Spike	Spike	Spike	Spike	Spike	Spike	ORF3A	MA	NP	3'UTR
-	WH-1	I	I	I	I	I	I	I	I	I	I	I	I	I	I
2	A-1	I	+	I	I	+	I	I	I	I	I	I	I	I	I
Э	A-2	I	+	I	I	+	I	I	I	I	I	I	I	I	I
4	B-1	I	+	I	I	Ι	Ι	I	Ι	I	+	Ι	Ι	Ι	I
S	B-2	I	+	I	I	I	I	I	I	T	I	I	I	I	I
9	G-1	I	+	I	I	Ι	I	I	Ι	I	Ι	Ι	Ι	Ι	I
~	G2	I	+	I	I	I	I	I	I	I	I	I	I	I	I
8	D-1	I	I	I	I	I	I	+	I	I	I	I	I	I	I
6	D2	I	I	I	I	I	I	I	I	I	I	I	+	I	I
10	L-1	I	I	I	I	I	I	I	I	I	I	I	I	I	I
11	L-2	I	I	I	I	I	I	I	I	I	I	I	I	I	I
12	M-1	I	I	I	I	I	+	I	I	I	I	+	I	I	I
13	M-2	I	I	I	I	I	+	I	I	I	I	+	I	I	I
14	0-1	I	+	+	I	Ι	Ι	I	Ι	I	Ι	Ι	Ι	+	I
15	O-2	+	+	I	+	Ι	Ι	I	+	+	Ι	Ι	Ι	+	I
16	BA2-1	I	+	+	I	Ι	Ι	I	+	+	Ι	Ι	Ι	+	+
17	BA2-2	I	+	+	I	I	I	I	I	I	I	I	I	+	+
18	GKA-1	I	I	I	I	I	I	I	I	I	I	I	I	I	I
19	GKA-2	I	I	I	I	Ι	I	+	I	I	I	Ι	Ι	I	I
20	GH-1	Ι	I	I	I	Ι	I	I	Ι	I	Ι	Ι	Ι	Ι	I
21	GH-2	I	I	I	I	Ι	Ι	I	Ι	Ι	I	Ι	I	Ι	I
N: no Virus Gamn Omici and G	insertion/deletion strains and GISAL a EPI ISL 87219, I on EPI ISL 901092 H-2 is GH/490 IS	: P: positive in D codes WH- D-1 is Delta El A, O-2 is Om: L 7702915.	sertion/deletic 1 is Wuhan-H PI ISL 1544014 icron EPI ISL 9	nı; *Position u-1, A-1 is Al , D-2 is Delta 042199, BA2·	based on aligr pha EPI ISL 1 EPI ISL 1838 -1 is BA2 LSP	ned sequence (489174, A-2 i 642, L-1 is La) A-3B08C6A,	starting from is Alpha EPI mbda EPI ISI BA2-2 is BA.	1 nucleotide n. ISL 1489173, L 111127, L-1 2 HSLL-3B08(	o. 1 of Wuhar B-1 is Beta EH 2 is Lambda E 857, GKA-1 i	ı-Hu-1; **Lc 1 ISL 660190 21 ISL 11113 2 EPI ISL 902	ocation in the J, B-2 is Beta ] 341, M-1 is M1 21856, GKA-2	genome was b EPI ISL 66061 1 EPI ISL 2339 is HDF-IPP0 is HDF-IPP0	ased on GenF 3, G-1 is Gam 9799, M-2 is N 8027, GH-1 is	ank data of W ma EPI ISL 8 1u EPI ISL 23 GH/490 EPI	/uhan-Hu-1; 1149, G-2 is 39801, O-1 is ISL 6316493,

# Advances in Virology

Concietant emberitistion to							
Wuhan-Hu-1	Specific for omicron, BA2, and GKA	Specific for omicron and BA2	Specific for delta and GKA	Specific to delta, omicron, BA.2, and GKA	Specific to omicron and GKA	Unique to GH	Unique to GKA clade
ORFIAB (1)	ORF1AB (10)		ORF1AB (2)			ORF1AB (5)	ORF1AB (9)
							E352D, A1306S.
	S135R, T842I,					E452G,	P2046L,
	G1307S, L3027F,		067020			L2119F,	A2529V,
P4715L/F	T3090I, L320IF,		P5401L			A2355S, 52140E	I2820V,
	I5967V, T6594I					32149F, R3164H	73646A,
							P4715F, A6319V
Spike* (1)	Spike <sup>*</sup> (21) T101 A275 G142D		Spike <sup>*</sup> (4)	Spike <sup>*</sup> (1)	Spike <sup>*</sup> (3)	Spike <sup>*</sup> (8)	NP (1)
	V214G G343D						
	S375F/L S377D					P9L, E96Q,	
	S379F T380A					I211T,	
	D409N R412S		T19R, E157G,		Ins216E,	R350S,	
D618G	DAAR FA88A		F158del,	T482K	Ins217P,	N398S,	G215C
	O497P O507P		R159del		Ins218E	Y453N,	
	V500H N683K					F494R,	
	N768K, D800Y, 0958H, N973K					D940H	
	NP (1)	NP (3)	NP (3)			NP (2)	ORF3A (2)
		E31del,	D63G,			D22Y.	
	<b>S</b> 413K	K32del, S33del	K203M, D377V			E378Q	S92L, D155Y
		12DCCC					
	UKF3A (1) T772I		UKF/A (2) 1707 A T1201			UKF3A (I) T22I	UKF/B (I) TAM
	16221		V 02A, 112UI			1701	1401
	ORF6 (1)		ORF8 (2)				
	D61L		D119del, F120del				

4



<sup>0.00050</sup> 

FIGURE 1: The phylogeny of the spike protein coding region of the representatives of variants of SARS-CoV-2. The phylogeny was constructed using the maximum likelihood method and JTT matrix-based model [10] conducted in MEGA-X software [9]. The phylogenetic tree was rooted to Wuhan-Hu-1 sequence. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches.

RNAs, either its own RNA, the RNA of other viruses, or cellular RNA; thus, template switching occurs during transcription [13]. This process leads to RNA indels. Mutations in SARS-CoV-2 prior to the emergence of variants have been reported [6]. In HIV, the deletions occurred by at least three different mechanisms: (i) misalignment of the growing point; (ii) incorrect synthesis and termination in the primer-binding sequence during the synthesis of the plusstrand strong-stop DNA; and (iii) incorrect synthesis and termination before the primer-binding sequence during synthesis of the plus-strand strong-stop DNA [14].

Previous whole-genome comparisons have been conducted, including for Omicron [15]. However, that work focused on phylogeny and did not cover the recently identified BA.2 and GPA lineages, which are colloquially known as Deltacron. Indels and amino acid substitutions unique to specific variants were not described.

Through random selection of variant representatives with definitive sequences across the genome, we managed to identify unique patterns of indels and amino acid substitutions. Even with only two representatives for each variant, which is indeed the limitation of this study, we identified quasispecies or, in the case of a variant, quasivariant. Viral quasispecies refers to a population structure that consists of extremely large numbers of variant genomes, termed mutant spectra, mutant swarms, or mutant clouds [16]. For SARS-CoV-2, this phenomenon has been discovered even in single infected individuals [17–21]. We proposed the term quasivariant, as many indels and amino acid substitutions occur in one of only two representatives. We believe that we will find more variation if we analyze more variant representatives.

Amino acids consistently substituted from Wuhan-Hu-1 across all variants are ORF1AB P4715L/F and spike D618G. The D618G has been covered in previous works [6, 22–29]. ORF1AB P4715L/F has also been described [30, 31]. A database-wide survey is needed to understand the frequency of those substitutions.

The variant that harbors the most variant-specific substitution from Wuhan-Hu-1 is VUM GH/490. Both representatives show five, eight, two, and one amino acid substitutions in ORF1AB, spike, NP, and ORF3A, respectively. This VUM is being tracked in Europe, Africa, Asia, and America; however, the genome frequency for access of GISAID dated March 30, 2022, is lower than 0.3%.

The GKA clade does not comprise a unique variant. It harbors no unique indels or substitutions compared with Wuhan-Hu-1, but it does share 34 amino acid replacements from Wuhan-Hu-1 with Omicron and BA2, 13 with Delta, one in spike with Delta, Omicron, and BA.2. Three insertions in the spike in one representative of Ins216E, Ins217P, and Ins218E of the GKA clade are shared with Omicron. The molecular signatures of Delta and Omicron are obvious in the GKA clade. It is plausible that the GKA clade is an Omicron subvariant. We suggest that the clade is not the result of sequencing errors, as previously thought [3].

The GKA clade seems to have no genetic advantage, so it becomes extinct shortly after its discovery. There are only 89 full genome sequences tagged with GKA clade upon access to the GISAIS database on May 3<sup>rd</sup>, 2022. The collection date of the earliest sequence was dated on January 20, 2022, and the last one was dated on March 21, 2022. This clade poses nine amino acid changes from the reference strain of Wuhan-Hu-1 in ORF1AB, two in ORF3A, and one in NP and ORF7B. No unique amino acid change was observed in the spike protein. The clade seems to be suppressed by antibodies to other variants following previous natural infection and/or vaccination.

Interestingly, we identified a truncated ORF3A in the Mu variant. Deletion of four nucleotides generates a stop codon; thus, ORF3A in this variant is 257 amino acids in length, whereas the others are 275 residues long. This accessory protein contributes to the pathogenesis of SARS-CoV-2 by inducing pathological apoptosis [32]. The effect of the Mu variant at the cellular level has not yet been described. One article on this variant covered the neutralization effect of antibodies [33]. According to the GISAID database accessed on March 30, 2022, this variant has been identified in many countries, with a maximum global genome frequency of less than 1%, which has declined recently.

BA.2 differs from Omicron in the deletion of 48 nucleotides from the 3'UTR. The 3'UTR of coronaviruses contains all cis-acting sequences necessary for viral replication and binds to cellular as well as the viral components nsp1 and N proteins [34], which are required for minus-strand RNA synthesis [35]. This has also been described in SARS-CoV-2, whereby the 3'UTR is involved in genomic dimerization and interacts with cellular microRNA [36]. BA.2 has recently increased in frequency in multiple regions of the world, suggesting that it has a selective advantage over Omicron [37-40]. The genome frequency of BA.2 has increased exponentially to 90% of total Omicron submissions, as based on GISAID accessed on the previous date. As the original SARS-CoV-2 has a basic reproduction number (R0) of 2.4-3 [41], Delta has an R0 of 5 [42], and Omicron has an R0 of estimated to be higher than 10 or three times greater than Delta [43]; additionally, BA.2 subvariant might have an R0 of 15 or higher. The higher transmissibility of BA.2 might be attributed, at least in part, to the shorter 3'UTR, which results in a higher speed of viral replication, which needs to be investigated further. However, because the coding region across the whole genome, particularly for the spike protein of BA.2, is very close to that of Omicron, people who survived Omicron infection should be naturally protected against BA.2.

Phylogenetic analysis (Figure 1) demonstrated that the BA2 and GKA subvariants are Omicron variant. The phylogeny shows that Omicron, GKA/Deltacron, and BA2 are clustered together and separated from other variants with 100% bootstrap support.

# 5. Conclusion

Whole-genome comparison of representatives of all variants revealed indel patterns that are specific to SARS-CoV-2 variants or subvariants. Polymorphic amino acid comparison across all coding regions also showed amino acid residues shared by specific groups of variants. Finally, the higher transmissibility of BA.2 might be due at least in part to the 48 nucleotide deletions in the 3'UTR, which result in a higher speed of viral replication, while the seem-to-be extinction of GKA clade is due to the lack of genetic advantage as a consequence of amino acid substitutions in various genes.

### **Data Availability**

All genome sequences and associated metadata in this dataset are published in the GISAID's EpiCov database. The final dataset is available at GISAID with identifier EPI\_S-ET\_230223kx, doi: 10.55876/gis8.230223kx. To view the contributors and each individual sequence with details such as the accession number, virus name, collection date, originating lab and submitting lab, and the list of authors, we need to visit 10.55876/gis8.230223kx. All polymorphic amino acids of all proteins of two representatives of each variant of SARS-CoV-2 are listed in Supplementary Material 1.

#### Disclosure

An earlier version of the manuscript has been presented as a preprint in https://www.researchsquare.com/article/rs-1526043/v1.

# **Conflicts of Interest**

The authors declare that they have no conflicts of interest.

## **Authors' Contributions**

Ida B. K. Suardana and Gusti N. Mahardika contributed in the conception. Ida B. K. Suardana, Bayu K. Mahardika, Made Pharmawati, and Putu H. Sudipa contributed in the acquisition, analysis, and interpretation of data. Tri K. Sari and Nyoman B. Mahendra drafted the manuscript. All authors reviewed the manuscript and approved the final version of the manuscript for publication.

#### Acknowledgments

The English language of the manuscript has been edited by Springer Nature Author Services (support@as.springernature.com). This study was supported by the Research, Technology, and Higher Education (RISTEKDIKTI) of Indonesia through world class research.

### **Supplementary Materials**

All polymorphic amino acids of all proteins of two representatives of each variant of SARS-CoV-2. (*Supplementary Materials*)

#### References

- H. Wang, X. Li, T. Li et al., "The genetic sequence, origin, and diagnosis of SARS-CoV-2," *European Journal of Clinical Microbiology and Infectious Diseases*, vol. 39, no. 9, pp. 1629–1635, 2020.
- [2] F. A. AlMalki, S. Albukhaty, A. A. Alyamani, M. N. Khalaf, and S. Thomas, "The relevant information about the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) using the five-question approach (when, where, what, why, and how) and its impact on the environment," *Environmental Science and Pollution Research International*, pp. 1–25, 2022.
- [3] F. Kreier, "Deltacron: the story of the variant that wasn't," *Nature*, vol. 602, no. 7895, p. 19, 2022.
- [4] F. Wu, S. Zhao, B. Yu et al., "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020.
- [5] Y. Jin, H. Yang, W. Ji et al., "Virology, epidemiology, pathogenesis, and control of COVID-19," *Viruses*, vol. 12, no. 4, p. 372, 2020.
- [6] D. Mercatelli and F. M. Giorgi, "Geographic and genomic distribution of SARS-CoV-2 mutations," *Frontiers in Mi*crobiology, vol. 11, p. 1800, 2020.
- [7] A. Rahimi, A. Mirzazadeh, and S. Tavakolpour, "Genetics and genomics of SARS-CoV-2: a review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection," *Genomics*, vol. 113, no. 1, pp. 1221–1232, 2021.

- [8] I. Jungreis, R. Sealfon, and M. Kellis, "SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes," *Nature Communications*, vol. 12, no. 1, p. 2642, 2021.
- [9] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, and X. Mega, "Mega X: molecular evolutionary genetics analysis across computing platforms," *Molecular Biology and Evolution*, vol. 35, no. 6, pp. 1547–1549, 2018.
- [10] D. T. Jones, W. R. Taylor, and J. M. Thornton, "The rapid generation of mutation data matrices from protein sequences," *Bioinformatics*, vol. 8, no. 3, pp. 275–282, 1992.
- [11] S. Pollett, M. A. Conte, M. Sanborn et al., "A comparative recombination analysis of human coronaviruses and implications for the SARS-CoV-2 pandemic," *Scientific Reports*, vol. 11, no. 1, Article ID 17365, 2021.
- [12] F. Robson, K. S. Khan, T. K. Le et al., "Coronavirus RNA proofreading: molecular basis and therapeutic targeting," *Molecular Cell*, vol. 80, no. 6, pp. 1136–1138, 2020.
- [13] M. J. Kim and C. Kao, "Factors regulating template switch in vitro by viral RNA-dependent RNA polymerases: implications for RNA-RNA recombination," vol. 98, no. 9, pp. 4972–4977, 2001.
- [14] G. A. Pulsinelli and H. M. Temin, "Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer," *Journal of Virology*, vol. 65, no. 9, pp. 4786–4797, 1991.
- [15] M. Kandeel, M. E. M. Mohamed, H. M. Abd El-Lateef, K. N. Venugopala, and H. S. El-Beltagi, "Omicron variant genome evolution and phylogenetics," *Journal of Medical Virology*, vol. 94, no. 4, pp. 1627–1632, 2021.
- [16] E. Domingo and C. Perales, "Viral quasispecies," PLoS Genetics, vol. 15, no. 10, Article ID 1008271, 2019.
- [17] F. Sun, X. Wang, S. Tan et al., "SARS-CoV-2 quasispecies provides an advantage mutation pool for the epidemic variants," *Microbiology Spectrum*, vol. 9, no. 1, Article ID 26121, 2021.
- [18] A. Ghorbani, S. Samarfard, A. Ramezani et al., "Quasi-species nature and differential gene expression of severe acute respiratory syndrome coronavirus 2 and phylogenetic analysis of a novel Iranian strain," *Infection, Genetics and Evolution*, vol. 85, Article ID 104556, 2020.
- [19] B. T. Lau, D. Pavlichin, A. C. Hooker et al., "Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangenome to individual infection quasispecies," *medRxiv*, vol. 11, no. 2, Article ID 20224816, 2020.
- [20] A. Armero, N. Berthet, and J. C. Avarre, "Intra-host diversity of SARS-cov-2 should not Be neglected: case of the state of victoria, Australia," *Viruses*, vol. 13, no. 1, p. 133, 2021.
- [21] A. Jary, V. Leducq, I. Malet et al., "Evolution of viral quasispecies during SARS-CoV-2 infection," *Clinical Microbiol*ogy and Infections, vol. 26, no. 11, pp. 1560.e1–1560.e4, 2020.
- [22] C. Andres, M. Pinana, B. Borras-Bermejo et al., "A year living with SARS-CoV-2: an epidemiological overview of viral lineage circulation by whole-genome sequencing in Barcelona city (Catalonia, Spain)," *Emerging Microbes and Infections*, vol. 11, no. 1, pp. 172–181, 2022.
- [23] K. Leung, Y. Pei, G. M. Leung, T. T. Lam, and J. T. Wu, "Estimating the transmission advantage of the D614G mutant strain of SARS-CoV-2, December 2019 to June 2020," *Euro Surveillance*, vol. 26, no. 49, Article ID 2002005, 2021.
- [24] Y. W. Cheng, T. L. Chao, C. L. Li et al., "D614G substitution of SARS-CoV-2 spike protein increases syncytium formation

and virus titer via enhanced furin-mediated spike cleavage," *mBio*, vol. 12, no. 4, Article ID 58721, 2021.

- [25] T. J. Yang, P. Y. Yu, Y. C. Chang, and S. T. D. Hsu, "D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation," *Journal of Biological Chemistry*, vol. 297, no. 4, Article ID 101238, 2021.
- [26] M. Bhattacharya, S. Chatterjee, A. R. Sharma, G. Agoramoorthy, and C. Chakraborty, "D614G mutation and SARS-CoV-2: impact on S-protein structure, function, infectivity, and immunity," *Applied Microbiology and Biotechnology*, vol. 105, no. 24, pp. 9035–9045, 2021.
- [27] A. S. Mat Yassim, M. F. F. Asras, A. M. Gazali, M. S. Marcial-Coba, U. A. Zainulabid, and H. F. Ahmad, "COVID-19 outbreak in Malaysia: decoding D614G mutation of SARS-CoV-2 virus isolated from an asymptomatic case in pahang," *Materials Today Proceedings*, vol. 48, pp. 828–836, 2022.
- [28] M. Eaaswarkhanth, A. Al Madhoun, and F. Al-Mulla, "Could the D614G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality?" *International Journal of Infectious Diseases*, vol. 96, pp. 459-460, 2020.
- [29] R. Muttineni, N. Kammili, T. C. Bingi et al., "Clinical and whole genome characterization of SARS-CoV-2 in India," *PLoS One*, vol. 16, no. 2, Article ID 246173, 2021.
- [30] C. De Marco, N. Marascio, C. Veneziano et al., "Wholegenome analysis of SARS-CoV-2 in a 2020 infection cluster in a nursing home of Southern Italy," *Infection, Genetics and Evolution*, vol. 99, Article ID 105253, 2022.
- [31] S. Banerjee, S. Seal, R. Dey, K. K. Mondal, and P. Bhattacharjee, "Mutational spectra of SARS-CoV-2 orf1ab polyprotein and signature mutations in the United States of America," *Journal of Medical Virology*, vol. 93, no. 3, pp. 1428–1435, 2021.
- [32] Y. Ren, T. Shu, D. Wu et al., "The ORF3a protein of SARS-CoV-2 induces apoptosis in cells," *Cellular and Molecular Immunology*, vol. 17, no. 8, pp. 881–883, 2020.
- [33] K. Uriu, I. Kimura, K. Shirakawa et al., "Genotype to phenotype Japan C: neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine serum," *New England Journal of Medicine*, vol. 385, no. 25, pp. 2397–2399, 2021.
- [34] D. Yang and J. L. Leibowitz, "The structure and functions of coronavirus genomic 3' and 5' ends," *Virus Research*, vol. 206, pp. 120–133, 2015.
- [35] Y. J. Lin, X. Zhang, R. C. Wu, and M. M. Lai, "The 3' untranslated region of coronavirus RNA is required for subgenomic mRNA transcription from a defective interfering RNA," *Journal of Virology*, vol. 70, no. 10, pp. 7236–7240, 1996.
- [36] J. A. Imperatore, C. L. Cunningham, C. J. Frye, K. A. Pellegrene, J. D. Evanseck, and M.-R. Mihailescu, "Conserved elements in the 3'-UTR of SARS-CoV-2: involvement in genomic dimerization and interactions with cellular micrornas," *Biophysical Journal*, vol. 120, no. 3, p. 313, 2021.
- [37] J. Yu, A. r Y. Collier, M. Rowe et al., "Comparable neutralization of the SARS-CoV-2 Omicron BA.1 and BA.2 variants," *medRxiv*, vol. 2, no. 6, Article ID 22270533, 2022.
- [38] V. C. C. Cheng, J. D. Ip, A. W. H. Chu et al., "Rapid spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron subvariant BA.2 in a single-source community outbreak," *Clinical Infectious Diseases*, vol. 75, no. 1, pp. e44–e49, 2022.

- [39] E. Mahase, "Omicron sub-lineage BA.2 may have "substantial growth advantage", UKHSA reports," *BMJ*, vol. 376, p. 263, 2022.
- [40] P. A. Desingu, K. Nagarajan, and B. A. Omicron, "Omicron BA.2 lineage spreads in clusters and is concentrated in Denmark," *Journal of Medical Virology*, vol. 94, no. 6, pp. 2360–2364, 2022.
- [41] M. D'Arienzo and A. Coniglio, "Assessment of the SARS-CoV-2 basic reproduction number, R 0, based on the early phase of COVID-19 outbreak in Italy," *Biosafety and Health*, vol. 2, no. 2, pp. 57–59, 2020.
- [42] Y. Liu and J. Rocklov, "The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus," *Journal of Travel Medicine*, vol. 28, no. 7, p. 124, 2021.
- [43] K. Ito, C. Piantham, and H. Nishiura, "Relative instantaneous reproduction number of Omicron SARS-CoV-2 variant with respect to the Delta variant in Denmark," *Journal of Medical Virology*, vol. 94, no. 5, pp. 2265–2268, 2022.