BRIEF REPORT

Consensus insertion/deletions and amino acid variations of all coding and noncoding regions of the SARS-CoV-2 Omicron clades, including the XBB and BQ.1 lineages

Hamong Suharsono[1] · Bayu K. Mahardika[2] · Putu H. Sudipa[3] · Tri K. Sari[4] · Ida B. K. Suardana[4] · Gusti N. Mahardika[2,4]

Abstract

The currently dominant Omicron variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has swiftly diverged into clades. To predict the probable impact of these clades, the consensus insertions/deletions (indels) and amino acid substitutions of the whole genome of clades were compared with the original SARS-CoV-2 strain. The evolutionary history of representatives of clades and lineages was inferred using the maximum-likelihood method and tested using the bootstrap method. The indels and polymorphic amino acids were found to be either clade-specific or shared among clades. The 21K clade has unique indels and substitutions, which probably represent reverted indels/substitutions. Three variations that appear to be associated with SARS-CoV-2 attenuation in the Omicron clades included a deletion in the nucleocapsid gene, a deletion in the 3'untranslated region, and a truncation in open reading frame 8. Phylogenetic analysis showed that the Omicron clades and lineages form three separate clusters.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will soon pass its three-year point. Despite mass vaccination with boosters in various countries, vaccines seem ineffective in curbing community transmission, as vaccine breakthrough is a common phenomenon observed in many parts of the world [1–3]. It is believed to occur due to the rapid evolution of SARS-CoV-2, which has led to the generation of many variants. The most recent and dominant variant is Omicron [1, 4, 5]. This variant has swiftly diverged into at least eight clades based on the data available in GISAID (https://nextstrain.org). According to the WHO Technical Advisory Group on SARS-CoV-2 Virus Evolution, some lineages, such as XBB and BQ.1, are a cause of concern because of their potential to bring on a new wave of cases and fatalities (www.who.int).

To predict the probable impact of these genetic variants, comparisons of whole-genome sequences of representatives of each clade or lineage are of global interest. However, as interclade or lineage variants are to be expected, preserved insertions/deletions (indels) and substitutions in comparison to the original SARS-CoV-2 strain are important genetic features of new clades/lineages. The genome organization of SARS-CoV-2, based on the open reading frame (ORF) annotation of the original Wuhan-Hu-1 isolate and adding the 5' and 3'untranslated regions (UTRs) and intergenic sequences (IGSs) is as follows: 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 [6, 7]. Intergenic sequences (IGSs) have been identified previously [8]. Transcription regulatory sequences (TRSs) have been identified at the junctions between these ORFs as well as at the 5' end of the genomic RNA downstream

✉ Gusti N. Mahardika
gnmahardika@unud.ac.id

1 Biochemistry Laboratory, The Faculty of Veterinary Medicine, Udayana University, Denpasar, Bali, Indonesia

2 The Animal Biomedical and Molecular Biology Laboratory, Udayana University, Jl. Sesetan—Markisa 6A, Denpasar 80223, Bali, Indonesia

3 Veterinary Bacteriology and Mycology Laboratory, The Faculty of Veterinary Medicine, Udayana University, Denpasar, Bali, Indonesia

4 Virology Laboratory, The Faculty of Veterinary Medicine, Udayana University, Denpasar, Bali, Indonesia

of the leader sequence [9]. Therefore, the 3' end of any SARS-CoV-2 gene can be critical for the translation of the next coding region. Accessory proteins should also be examined because they contribute to the pathogenesis of SARS-CoV-2 [10, 11]. The 5'- and 3'-UTRs should be examined, as they have been demonstrated to play important roles in viral fitness and pathogenesis [12].

Here, we compared the genome sequences of members of various clades of the Omicron variant as well as the XBB and BQ.1 lineages. The objective of this study was to identify unique consensus insertion/deletions and amino acid variations of all coding and noncoding regions of the SARS-CoV-2 Omicron clades, including the XBB and BQ.1 lineages.

Methods

Fifteen to 25 sequences of lineages 21K, 21L, 22A, 22B, 22C, 22D, 22E, and 22F, as well as XBB and BQ.1, were randomly selected from the GISAID Nextstrain phylogeny on October 31, 2022 and downloaded. The total number of sequences in the dataset was 203. The whole genome sequence was aligned with the original SARS-CoV-2 sequence of the Wuhan-Hu-1 isolate (GenBank accession no. NC_045512) using Clustal Omega, available online at EMBL-EBI (www.ebi.ac.uk). In whole-genome sequence alignments, sequences that caused long gaps due to the presence of a long track of Ns were excluded. Individual coding regions were selected based on the Wuhan-Hu-1 coding DNA sequence (CDSs) using MEGA11 software [13]. In comparisons of coding regions or open reading frames (ORFs), sequences with a track of more than two unidentified nucleotides (NNNs) were excluded. Therefore, the final number of sequences of each clade and or lineage varies, as shown in Tables 1, 2, and 3. The sequences were translated prior to alignment. Polymorphic amino acids as well as gaps were tabulated manually. To assess the genetic relatedness of clades and lineages, three representatives of clades and lineages from different countries were randomly selected from the dataset as above and aligned using Clustal Omega. The 5' and 3'ends were trimmed to produce sequences of equal length, with a total of 28,934 positions in the final dataset. The evolutionary history was inferred using the maximum-likelihood method and the Kimura 2-parameter model in MEGA 11 software [13]. The phylogeny was tested by the bootstrap method with 100 replications.

Results

The sequence dataset is available at GISAID with the identifier EPI_SET ID: EPI_SET_230327ca and https://doi.org/10.55876/gis8.230327ca. The number of sequences in the dataset of various clades and the XBB and BQ1 lineages and the number of sequences in each clade and lineage bearing consensus deletions and insertions are presented in Supplementary Material 1. The consensus indel pattern in all Omicron clades or lineages includes del-11260-11268 in ORF1AB and del-28346-28354 in NC. Other indels are unique to a clade or a lineage or shared between clades and/or lineages. The 21K clade has two unique deletions and an insertion (spike del-21960-21965, del-22167-22169, and ins-22178-22186). The deletion del-658-666 is unique to the 22A clade. Del-21606-21614 in ORF1AB and del-29732-29757 in the 3'-UTR are present in all clades except 21K. Del-21738-21743 in the spike coding region is not present in clades 21L, 22C, 22D, and 22F. The other deletion in the spike coding region, namely, 21966-21968, is present in 21K, 22F, and XBB.

Amino acid residues that are characteristic of all Omicron clades and the XBB and BQ.1 lineages in ORF1AB, the spike gene, and the combined ORF3A, envelope protein, matrix ORF6, ORF7B, ORF8, and NC are shown in Tables 1, 2, and 3, respectively. The substitutions that all of these clades and lineages have in common are I2235L, T3255I, P3395H, S3675Del, G3676Del, F3677Del, K3833N, P4715L, and I4175V in ORF1AB; G142D, S376P, S378F, K420N, N443K, S480N, T481K, E487A, Q501R, N504Y, Y508H, D617G, H658Y, N682K, P684H, N767K, D799Y, Q957H, and N972K in the spike protein; T9I in the envelope protein; Q19E and A63T in the matrix protein; and P13L, E31del, R32del, S33del and G204R in the NC. The amino acid substitutions/deletions present in all clades and lineages except the 21K clade are S135R, T842I, G1367S, L3207F, T3090I, R5716C, and T6564I in ORF1AB; T19I, L24del, P25del, P26del, A27S, V21G, T379A, D408N, and R411S in the spike protein; T22I in ORF3A; and S417R in the NC.

The unique substitutions in the 22K clade are K856R, L2084I, A2710T, and I3758V in ORF1AB and A67V, T95I, 143Vdel, 144Ydel, 145Ydel, N211del, L212I, ins215E, ins216P, ins217E, G499S, T550K, N859K, and L984F in the spike protein. The remaining clades resemble Wuhan-Hu-1 at those sites.

Substitutions found in both 22F and XBB include K47R in ORF1AB and V83A, Y144del, H146Q, Q183E, V213E, G255V, L371I, V448P, F489S, and F493S in the spike protein. Unique to the 22F clade and XBB lineage is the presence of a stop codon at position 8 of ORF8. Q556K, L3829F, Y4665H, M5557I, and N5592S in ORF1AB, as well as K447T in the spike protein are shared by 22E and BQ.1.

A phylogenetic tree is presented in Figure 1. The tree shows that the Omicron clades and lineages form three separated clusters with 100% bootstrap support. Clade 22K forms a unique cluster (cluster 1), while cluster 2 consists of clades

Table 1 Consensus amino acid substitutions/deletions in the ORF1AB protein of Omicron variant clades/lineages compared to Wuhan-Hu-1*

Strain/clade/lineage	Number of sequences	Position																			
		47	135	141–143	556	842	856	1221	1307	1640	2084	2235	2710	2909	3027	3090	3201	3255	3395	3674
Wuhan-Hu-1	1	K	**S**	KSF	Q	**T**	K	S	**G**	P	L	**I**	A	A	L	**T**	**L**	**T**	**P**	L
21K	19	K	**S**	KSF	Q	**T**	**R**	S	**G**	P	**I**	L	**T**	A	**L**	**T**	**L**	I	H	L
21L	20	K	R	KSF	Q	I	K	S	S	P	L	L	A	**V**	F	I	F	I	H	L
22A	14	K	R	**del**	Q	I	K	S	S	P	L	L	A	A	F	I	**L**	I	H	L
22B	14	K	R	KSF	Q	I	K	S	S	P	L	L	A	A	F	I	F	I	H	L
22C	12	K	R	KSF	Q	I	K	S	S	P	L	L	A	A	F	I	F	I	H	L
22D	11	K	R	KSF	Q	I	K	**L**	S	**S**	L	L	A	A	F	I	F	I	H	L
22E	9	K	R	KSF	**K**	I	K	S	S	P	L	L	A	A	F	I	F	I	H	**F**
22F	10	**R**	R	KSF	Q	I	K	S	S	P	L	L	A	A	F	I	F	I	H	L
XBB	25	**R**	R	KSF	Q	I	K	S	S	P	L	L	A	A	F	I	**L**	I	H	L
BQ.1	13	K	R	KSF	**K**	I	K	S	S	P	L	L	A	A	F	I	**L**	I	H	L

Strain/clade/lineage	Number of sequences	Position													
		3675–3677	3758	3829	3833	4060	4665	4715	5063	5360	5557	5592	5716	5967	6564
Wuhan-Hu-1	1	**SGF**	I	L	**K**	N	Y	**P**	G	S	M	N	**R**	**I**	**T**
21K	19	del	**V**	L	N	N	Y	L	G	S	M	N	**R**	V	**T**
21L	20	del	I	L	N	N	Y	L	G	S	M	N	C	V	I
22A	14	del	I	L	N	N	Y	L	G	S	M	N	C	V	I
22B	14	del	I	L	N	N	Y	L	G	S	M	N	C	V	I
22C	12	del	I	L	N	N	Y	L	G	S	M	N	C	V	I
22D	11	del	I	L	N	**S**	Y	L	**S**	S	M	N	C	V	I
22E	9	del	I	L	N	N	**H**	L	G	S	**I**	**S**	C	V	I
22F	10	del	I	L	N	N	Y	L	**S**	**P**	M	N	C	V	I
XBB	25	del	I	L	N	N	Y	L	S	S	M	N	C	V	I
BQ.1	13	del	I	**F**	N	N	**H**	L	G	S	**I**	**S**	C	V	I

* Residues unique to certain clades and/or lineages are highlighted

Table 2 Consensus amino acid substitutions/deletions in the spike protein of Omicron variant clades/lineages compared to Wuhan-Hu-1*

Amino acid position

Strain/clade/lineage	Number of Seq.¹	19	24–26	27	67	69–70	83	95	142	143–145	146	147	152	157	183	210	211	212	213	215–217	255	260	342	348	371	374	376	378	379	408
Wuhan Hu-1	1	**T**	**LPP**	A	A	HV	V	T	**G**	VYY	H	K	W	F	Q	I	N	L	**V**	--	G	G	G	R	L	**S**	**S**	**S**	**T**	**D**
21K	12	**T**	**LPP**	A	**V**	**del**	V	**I**	D	**del**	H	K	W	F	Q	I	**del**	**I**	V	**EPE**	G	G	G	**K**	L	**L**	P	F	**T**	**D**
21L	19	I	del	S	A	HV	V	T	D	VYY	H	K	W	F	Q	I	N	L	G	--	G	G	D	R	L	F	P	F	A	N
22A	16	I	del	S	A	**del**	V	T	D	VYY	H	K	W	F	Q	I	N	L	G	--	G	G	D	**T**	L	F	P	F	A	N
22B	14	I	del	S	A	**del**	V	T	D	VYY	H	K	W	F	Q	I	N	L	G	--	G	G	D	R	L	F	P	F	A	N
22C	22	I	del	S	A	HV	V	T	D	VYY	H	K	W	F	Q	I	N	L	G	--	G	G	D	R	L	F	P	F	A	N
22D	12	I	del	S	A	HV	V	T	D	VYY	H	**E**	**R**	**L**	Q	**V**	N	L	G	--	G	**S**	**H**	R	L	F	P	F	A	N
22E	11	I	del	S	A	**del**	V	T	D	VYY	H	K	W	F	Q	I	N	L	G	--	G	G	D	**T**	L	F	P	F	A	N
22F	10	I	del	S	A	HV	**A**	T	D	**V-Y**	**Q**	K	W	**E**	**E**	I	N	L	**E**	--	**V**	G	**H**	**T**	**I**	F	P	F	A	N
XBB	25	I	del	S	A	HV	**A**	T	D	**V-Y**	**Q**	K	W	**E**	**E**	I	N	L	**E**	--	**V**	G	**H**	**T**	**I**	F	P	F	A	N
BQ.1	16	I	del	S	A	HV	V	T	D	VYY	H	K	W	F	Q	I	N	L	G	--	G	G	D	**T**	L	F	P	F	A	N

Amino acid position

Strain/clade/lineage	411	420	443	447	448	449	455	463	480	481	487	489	493	496	499	501	504	508	550	617	658	661	682	684	707	767	799	859	957	972	984
Wuhan Hu-1	**R**	**K**	**N**	**K**	**V**	**G**	**L**	**N**	**S**	**T**	**E**	**F**	**Q**	**G**	**Q**	**N**	**Y**	**T**	**D**	**H**	**N**	**N**	**P**	**S**	**N**	**D**	**N**	**N**	**Q**	**N**	**L**
21K	**R**	N	K	K	V	**S**	L	N	N	K	A	**F**	**R**	**S**	R	Y	H	**K**	G	Y	K	K	H	H	K	Y	**K**	K	H	K	**F**
21L	S	N	K	K	V	G	L	N	N	K	A	**F**	**R**	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
22A	S	N	K	K	V	G	**R**	N	N	K	A	**V**	Q	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
22B	S	N	K	K	V	G	**R**	N	N	K	A	**V**	Q	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
22C	S	N	K	K	V	G	**Q**	N	N	K	A	F	**R**	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
22D	S	N	K	K	V	**S**	L	K	N	K	A	F	Q	G	R	Y	H	T	G	Y	K	K	**L**	H	K	Y	N	K	H	K	L
22E	S	N	K	**T**	**P**	**S**	**R**	K	N	K	A	**V**	Q	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
22F	S	N	K	**P**	**P**	**S**	L	K	N	K	A	**S**	Q	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
XBB	S	N	K	**T**	**P**	**S**	**R**	K	N	K	A	**S**	Q	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L
BQ.1	S	N	K	**T**	V	G	**R**	K	N	K	A	**V**	Q	G	R	Y	H	T	G	Y	K	K	H	H	K	Y	N	K	H	K	L

*Residues unique to certain clades and/or lineages are highlighted. ¹Number of sequences of the spike protein in each clade/lineage without a track of more than two 'Ns'

Table 3 Consensus amino acid substitutions/deletions of ORF3A, envelope protein, matrix, ORF6, ORF7B, ORF8, and nucleoprotein of Omicron variant clades/lineages compared to Wuhan-Hu-1[1]

Strain/ clade/ lineage	ORF3A			Envelope protein			Matrix				ORF6[3]		ORF7B		ORF8		Nucleoprotein				
	Number of seq.[2]	140	223	Number of seq[2]	9	11	Number of seq.[2]	3	19	63	Number of seq.[2]	61	Number of seq.[2]	11	Number of seq.[2]	8	Number of seq.[2]	13	31–33	204	417
Wuhan Hu-1	1	L	**T**	1	**T**	T	1	D	**Q**	**A**	1	D	1	L	1	G	1	**P**	**ERS**	**G**	**S**
21 K	21	L	I	21	I	T	17	**G**	E	T	21	D	19	L	19	G	19	L	del	R	S
21L	23	**F**	I	24	I	T	20	D	E	T	24	**L**	24	L	24	G	21	L	del	R	R
22A	24	L	I	24	I	T	21	D	E	T	21	**L**	21	**F**	15*	G	24	L	del	R	R
22B	21	L	I	21	I	T	19	**N**	E	T	21	D	21	L	21	G	21	L	del	R	R
22C	24	L	I	24	I	T	21	D	E	T	22	**L**	23	L	23	G[1]	24	L	del	R	R
22D	13	L	I	13	I	**A**	9	D	E	T	14	**L**	14	L	14	G	14	L	del	R	R
22E	12	L	I	13	I	T	9	**N**	E	T	13	D	11	L	11	G	12	L	del	R	R
22F	15	L	I	15	I	**A**	15	D	E	T	15	**L**	14	L	14	**Stop**	15	L	del	R	R
XBB	25	L	I	25	I	**A**	24	D	E	T	25	**L**	25	L	25	**Stop**	25	L	del	R	R
BQ.1	20	L	I	20	I	T	16	**N**	E	T	20	D	19	L	16	G	19	L	del	R	R

[1]No dominant (>50% of sequence data) amino acids, deletions, or insertions were found in ORF7A and ORF10. [2] Number of sequences in each clade/lineage without a track of more than two 'Ns'. [3]Stop codon at position 62 of ORF7A in one sequence. [3]Stop codon at position 8 in one 21 K sequence. [4]Stop codon at position 8 in one sequence. [5]Stop codon at position 8 in three sequence of clade 22F. [6] Stop codon at position 8 of 31 sequences of lineage XBB. Stop codon in ORF7a at position 62 in 21 K. * massive deletion in 5 of the 22A sequences. Residues unique to certain clades and/or lineages are highlighted

22B and 22E as well as the BQ lineage, and the other clades and lineages form cluster 3.

Discussion

Consensus indels and amino acid variations in all coding and noncoding regions of the SARS-CoV-2 Omicron variant are of global interest, as this variant is evolving rapidly and has become the global dominant circulating variant, suppressing others. Scientific explanation is paramount to understanding the potential threat of subvariants or lineages. Most reports on SARS-CoV-2 variants have emphasized changes in the spike protein [3, 4]. However, examination of the whole genome is important for assessing the impact of mutations in subvariants [14].

Some indels and polymorphic amino acids are specific to the Omicron variant. All Omicron clades and lineages contain del-11260-11268 in ORF1AB and del-28346-28354 in NC. However, del-11260-11268 is not unique, as it is also present in the Alpha, Beta, and Gamma variants [15].

The deletion in the NC of SARS-CoV-2 is a probable indirect signature of its attenuation. The SARS-CoV-2 NC is an abundantly expressed RNA-binding protein that is critical
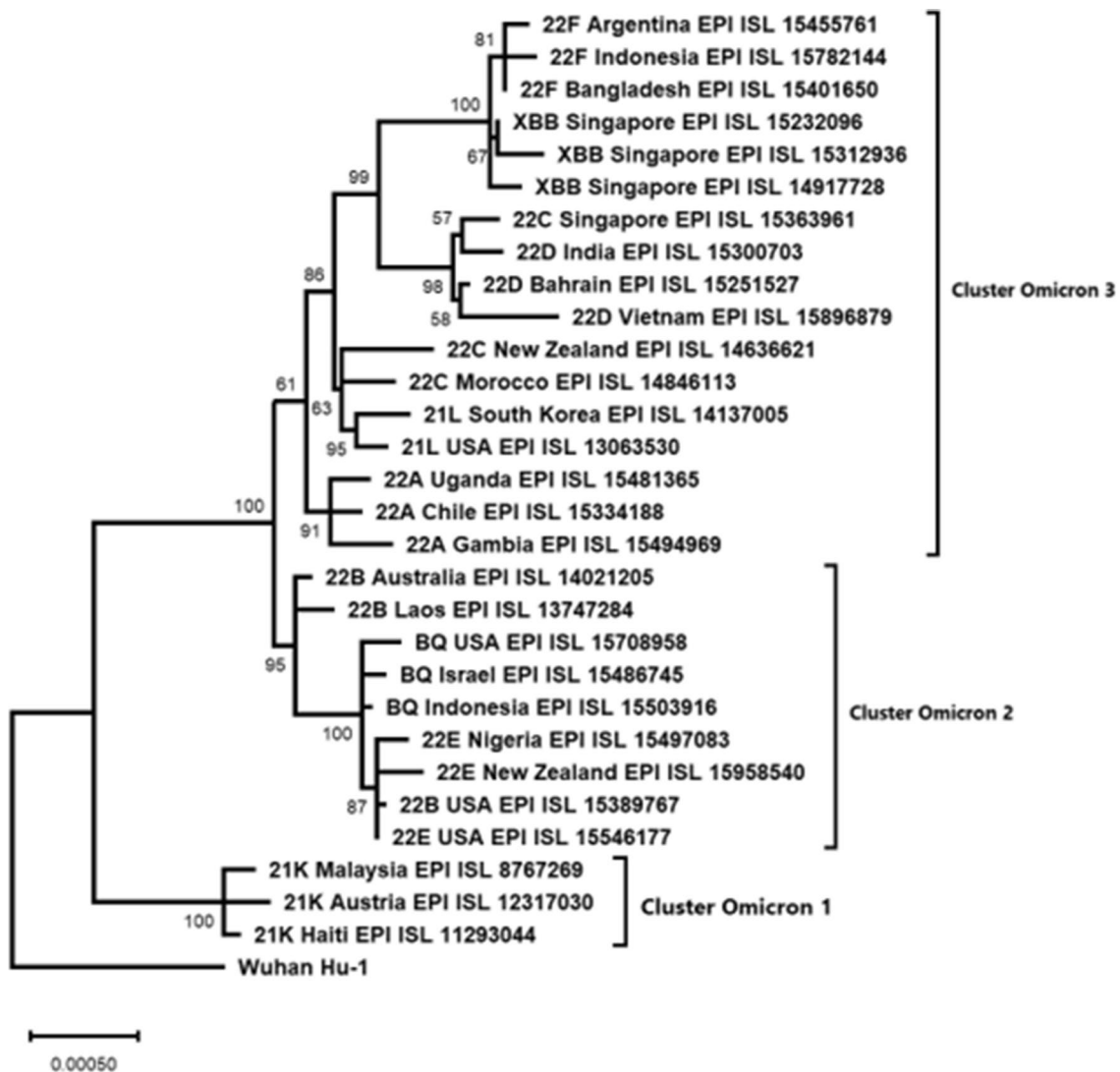
Fig. 1 Phylogenetic tree, based on full genome sequences of three randomly selected representatives of the Omicron clades and lineages of SARS-CoV-2 rooted to the Wuhan-Hu-1 strain. The clade, country of origin, and EPI-ID are indicated in the isolate names. The evolutionary history was inferred using the maximum-likelihood method and the Kimura 2-parameter model in MEGA 11 software [13]. The phylogeny was tested by the bootstrap method with 100 replications. 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. There were a total of 28934 positions in the final dataset.

for viral genome packaging [16]. The presence of del-28346-28354 in the NC can potentially alter the biology of the virus. In the coronavirus mouse hepatitis virus, a deletion in the NC resulted in a small-plaque phenotype in tissue culture [17]. Plaque size is also an indicator of dengue virus attenuation [18], in which the molecular determinants of small plaque size are mutations/substitutions in NS1, NS3, and the 3'-UTR [19]. It is therefore plausible that the deletion of the NC in SARS-CoV-2 is an indirect indicator of its attenuation.

The deletion in the 3'-UTR is another notable characteristic of the Omicron variant. This deletion is dominant in all clades/lineages, with the exception of clade 21K. This region might be important for recognition by the SARS-CoV-2 RNA-dependent RNA polymerase and cellular components for the initiation of anti-genomic (negative strand) RNA synthesis [12]. The deletion in the 3'-UTR is another probable indirect indicator of SARS-CoV-2 subvariant attenuation.

Although Omicron SARS-CoV-2 spread faster than other variants and became the dominant variant globally, it was reported to cause milder clinical signs [5]. The intensive care unit admission rates for Omicron-infected patients were much lower than those of Delta- and Delta-/Omicron-infected patients [20], suggesting that this variant has reduced virulence.

Since many dominant amino acid residues in the spike protein are uniformly divergent from Wuhan-Hu-1, individuals who have recovered from an Omicron infection might be expected to have protective immunity to all Omicron clades/lineages. Applying the template of spike protein residues and their possible functions as published previously [4], the consensus amino acid changes in Omicron clades/lineages relative to Wuhan-Hu-1 are located in the receptor-binding domain/receptor binding site (RBD/RBS) (S376P, S378F, K420N, N443K, S480N, T481K, E487A, Q501R, N504Y, Y508H), linear epitopes (S378F, K420N, E487A, Q501R, D617G, N767K), possible conformation-dependent epitopes (N682K, P684H), the S1/S2 cleavage site ((N682K, P684H), the fusion peptide (D799Y), and heptad repeat 1 (Q957H and N972K). With this pattern, it is expected that the Omicron variant has different biological characteristics than the original SARS-CoV-2 strain.

It was observed in this study that reversion of indels and mutations might have occurred in SARS-CoV-2. In this case, the term "reversions" or "reverse mutations", refers to any mutational processes or mutations that restores the wild-type phenotype to an organism already carrying a phenotype-altering forward mutation [21]. This phenomenon has been described for many viruses [22]. Our data show that the 21K clade has unique indels and substitutions, while the remaining sequence is homologous to that of Wuhan-Hu-1. The revertant virus evolved further with deletions and substitutions in various genome segments.

The indels and substitutions in 21K seem to be unstable or generate lower virus fitness.

In this study, we confirm the clade separation reported by Nextstrain. XBB is close to the 22F clade, while BQ1 is close to the 22E clade. Clade 22F shares many amino acid substitutions with the XBB lineage, while clade 22E shares many with the BQ.1 lineage. This manuscript was drafted to provide valid data on the position of both lineages in SARS-CoV-2 phylogeny. This should suppress public speculation that XBB and BQ.1 are *de novo* subvariants and should demonstrate that the genetic make-up of these subvariants is similar to that of other members of the clade. The phylogenetic analysis (Fig. 1) also confirmed that the BQ.1 lineage is close to the 22E clades, while the XBB lineage is close to 22F clade.

Another note from this analysis is that the accessory proteins might not be critical for SARS-CoV-2 integrity. Without those proteins, SARS-CoV-2 remains viable. ORF8 is truncated in the 22F clade and XBB lineage. A stop codon at position 8 of ORF8 is dominant in that clade/lineage. Stop codons were also present in ORF6 and ORF7A of some strains (not shown). The accessory proteins have been described to contribute to the pathogenesis of SARS-CoV-2 [10, 11]. Deletions in ORF7 and 8 have been associated with milder symptoms [23]. ORF8 interferes with host immune responses in various ways, including downregulating MHC class I molecules [24], antagonizing interferon [25], activating interleukin 17, and cytokine storms [26]. It is therefore reasonable to suggest that the truncated ORF8 is additional indirect evidence of lower virulence or attenuation, especially in clade 22E.

This study does not provide information for prediction of the outcomes of SARS-CoV-2 infection. A scientific task force should be formed in each country to study the association of Omicron with the patient's clinical status so that the public can be aware of whether any emerging variant or subvariant warrants a change in COVID-19 prevention protocols.

In conclusion, the indels and polymorphic amino acids across the whole genome of SARS-CoV-2 Omicron clades are either clade-specific or shared among clades. Del-28346-28354 in NC is unique to Omicron. Variation in the 3'-UTR are common to all clades/lineages, except clade 21K. Clade 21K has four unique indels and substitutions in ORF1AB and 14 in the spike protein, while the remaining sequence is homologous to that of Wuhan-Hu-1, which probably represents reverted indels/substitutions. ORF8 is truncated at amino acid 8 in the 22F clade and in the XBB lineage. Three indirect lines of evidence of SARS-CoV-2 attenuation in Omicron clades were identified, namely, the deletion in NC, the deletion in the 3'-UTR, and the truncation of ORF8 in the 22F clade and XBB lineage.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Boekel L, Besten YR, Hooijberg F, Wartena R, Steenhuis M, Vogelzang E, Leeuw M, Atiqi S, Tas SW, Lems WF et al (2022) SARS-CoV-2 breakthrough infections in patients with immune-mediated inflammatory diseases during the omicron dominant period. Lancet Rheumatol 4(11):e747–e750

2. Johnson AG, Amin AB, Ali AR, Hoots B, Cadwell BL, Arora S, Avoundjian T, Awofeso AO, Barnes J, Bayoumi NS et al (2022) COVID-19 incidence and death rates among unvaccinated and fully vaccinated adults with and without booster doses during periods of delta and omicron variant emergence—25 US jurisdictions, April 4-December 25, 2021. MMWR Morb Mortal Wkly Rep 71(4):132–138

3. Kared H, Wolf AS, Alirezaylavasani A, Ravussin A, Solum G, Tran TT, Lund-Johansen F, Vaage JT, Nissen-Meyer LS, Nygaard UC et al (2022) Immune responses in Omicron SARS-CoV-2 breakthrough infection in vaccinated adults. Nat Commun 13(1):4165

4. Mahardika GN, Mahendra NB, Mahardika BK, Suardana IBK, Pharmawati M (2022) Annotating spike protein polymorphic amino acids of variants of SARS-CoV-2 Including Omicron. Biochem Res Int 2022:2164749

5. Meo SA, Meo AS, Al-Jassir FF, Klonoff DC (2021) Omicron SARS-CoV-2 new variant: global prevalence and biological and clinical characteristics. Eur Rev Med Pharmacol Sci 25(24):8012–8018

6. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579(7798):265–269

7. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, Duan G (2020) Virology, epidemiology, pathogenesis, and control of COVID-19. Viruses 12:4

8. Jungreis I, Sealfon R, Kellis M (2021) SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat Commun 12(1):2642

9. Yang D, Leibowitz JL (2015) The structure and functions of coronavirus genomic 3' and 5' ends. Virus Res 206:120–133

10. Redondo N, Zaldivar-Lopez S, Garrido JJ, Montoya M (2021) SARS-CoV-2 accessory proteins in viral pathogenesis: knowns and unknowns. Front Immunol 12:708264

11. Silvas JA, Vasquez DM, Park JG, Chiem K, Allue-Guardia A, Garcia-Vilanova A, Platt RN, Miorin L, Kehrer T, Cupic A et al (2021) Contribution of SARS-CoV-2 accessory proteins to viral pathogenicity in K18 human ACE2 transgenic mice. J Virol 95(17):e0040221

12. Verma R, Saha S, Kumar S, Mani S, Maiti TK, Surjit M (2021) RNA-Protein interaction analysis of SARS-CoV-2 5' and 3' untranslated regions reveals a role of lysosome-associated membrane protein-2a during viral infection. MSystems 6(4):e0064321

13. Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38(7):3022–3027

14. Chatterjee S, Kim CM, Lee YM, Seo JW, Kim DY, Yun NR, Kim DM (2022) Whole-genome analysis and mutation pattern of SARS-CoV-2 during first and second wave outbreak in Gwangju, Republic of Korea. Sci Rep 12(1):11354

15. Suardana IBK, Mahardika BK, Pharmawati M, Sudipa PH, Sari TK, Mahendra NB, Mahardika GN (2022) Whole-genome comparison of representatives of all variants of SARS-CoV-2, including subvariant BA.2 and the GKA clade. Research Square

16. Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Brereton MD, Ward MD, Zimmerman MI, Vithani N, Griffith D, Wagoner JA et al (2020) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. BioRxiv 2:2

17. Koetzner CA, Parker MM, Ricard CS, Sturman LS, Masters PS (1992) Repair and mutagenesis of the genome of a deletion mutant of the coronavirus mouse hepatitis virus by targeted RNA recombination. J Virol 66(4):1841–1848

18. Goh KC, Tang CK, Norton DC, Gan ES, Tan HC, Sun B, Syenina A, Yousuf A, Ong XM, Kamaraj US et al (2016) Molecular determinants of plaque size as an indicator of dengue virus attenuation. Sci Rep 6:26100

19. Blaney JE Jr, Johnson DH, Manipon GG, Firestone CY, Hanson CT, Murphy BR, Whitehead SS (2002) Genetic basis of attenuation of dengue virus type 4 small plaque mutants with restricted replication in suckling mice and in SCID mice transplanted with human liver cells. Virology 300(1):125–139

20. Bal A, Simon B, Destras G, Chalvignac R, Semanas Q, Oblette A, Queromes G, Fanget R, Regue H, Morfin F et al (2022) Detection and prevalence of SARS-CoV-2 co-infections during the Omicron variant circulation in France. Nat Commun 13(1):6316

21. Rosenberg SM (2013) Reverse mutation. In: Maloy S, Hughes K (eds) Brenner's encyclopedia of genetics, 2nd edn. Academic Press, New York, pp 220–221

22. Cann AJ (2012) Chapter 3—genomes. In: Cann AJ (ed) Principles of molecular virology, 5th edn. Academic Press, New York, pp 55–101

23. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, Zhuang Y, Kalimuddin S, Low JGH, Tan CW et al (2020) Discovery and genomic characterization of a 382-nucleotide deletion in ORF7b and ORF8 during the early evolution of SARS-CoV-2. MBio 11:4

24. Vinjamuri S, Li L, Bouvier M (2022) SARS-CoV-2 ORF8: one protein, seemingly one structure, and many functions. Front Immunol 13:1035559

25. Li JY, Liao CH, Wang Q, Tan YJ, Luo R, Qiu Y, Ge XY (2020) The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Res 286:198074

26. Lin X, Fu B, Yin S, Li Z, Liu H, Zhang H, Xing N, Wang Y, Xue W, Xiong Y et al (2021) ORF8 contributes to cytokine storm during SARS-CoV-2 infection by activating IL-17 pathway. Science 24(4):102293