RESEARCH ARTICLE

Temporal changes in the accessory protein mutations of SARS-CoV-2 variants and their predicted structural and functional effects

Christian Alfredo K. Cruz 🅖 | Paul Mark B. Medina 🅖

Biological Models Laboratory, Department of Biochemistry and Molecular Biology, College of Medicine, University of the Philippines Manila, Manila, Metro Manila, Philippines

Correspondence
Paul Mark B. Medina, Biological Models Laboratory, Department of Biochemistry and Molecular Biology, College of Medicine, University of the Philippines Manila, Manila, Metro Manila 1000, Philippines.
Email: pmbmedina@post.upm.edu.ph

Abstract

Emerging variants enable the continuous spread of SARS-CoV-2 in humans. The factors contributing to behavioral differences in variants remain elusive despite associations with several Spike protein mutations. Exploring accessory proteins may provide a wider understanding of these differences since these proteins may affect viral processes that occur beyond infection. Various bioinformatics tools were utilized to identify significant accessory protein mutations and determine their structural and functional effects over time. The ViruClust web application was used to retrieve accessory protein amino acid sequences and determine mutation frequencies in these sequences across time. The structural and functional effects of the mutations were determined using Missense3D and PROVEAN, respectively. The accessory and Spike protein mutations were compared using mutation densities. Q57H and T151I of ORF3a; T21I and W27L of ORF6; G38V, V82A, and T120I of ORF7a; S31P and T40I of ORF7b; and R52I, C61F, and I121L of ORF8 were highly frequent in most variants of concern and were within known functional domains. Thus, these are good candidates for further experimental evaluation. Among the accessory proteins, ORF6 and ORF8 were highlighted because of their strong and weak correlation with Spike protein mutations, respectively.

KEYWORDS
accessory proteins, bioinformatics, COVID-19, mutations, SARS-CoV-2 variants

1 | INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, the causative agent of coronavirus disease 2019 (COVID-19), has been causing a pandemic since 2020. Emerging SARS-CoV-2 variants have been fueling the continuous spread of this virus in humans despite widespread vaccination. So far, the origin of these variants is still unknown and the factors that drive the mutations in these variants remain elusive.

As of January 2022, the World Health Organization (WHO) has recognized and monitored 12 SARS-CoV-2 variants that are classified into variants of interest (VOI) and variants of concern (VOC). The WHO defines a VOI as a "SARS-CoV-2 isolate with mutations that lead to amino acid changes associated with established or suspected phenotypic implications." Moreover, these variants have been shown to "cause community transmission, multiple COVID-19 case clusters, or has been detected in multiple countries." A VOC also has similar characteristics to a VOI. However, a VOC is "associated with an increase in transmissibility, change in the epidemiology, increase in virulence, change in disease outcome, or decrease in the effectiveness of diagnostics, vaccines, therapeutics, or public health and social measures" in a global scale.[1] So far, there are five VOC (Alpha, Beta, Gamma, Delta, and Omicron) and two VOI (Lambda and Mu).

The genetic basis for this classification has mostly been associated with the Spike protein, a viral membrane protein that facilitates entry of SARS-CoV-2 in susceptible cells by binding to the human ACE2 receptor (hACE2). The earliest known Spike protein mutation is the D614G mutation, which enables increased deposition of Spike protein in viral particles, increased binding affinity to hACE2, and increased infectivity of SARS-CoV-2 in human bronchial epithelial cells.[2] K417N/T, N439K, E484K, L452R, and N501Y mutations are also significant mutations since these also lead to increased infectivity and transmission of SARS-CoV-2 variants[3-5] and resistance to human antiviral immune response.[6] Aside from decreased hACE2 receptor binding affinity, Spike protein mutations such as P681R may also lead to increased infectivity through efficient Spike protein cleavage.[4]

Despite associations between Spike protein mutations and SARS-CoV-2 variants, there are still aspects of these variants that cannot be explained solely by the mutations in the Spike protein. There seems to be a difference in the intrinsic virulence of the Alpha, Beta, Gamma, and Delta variants in humans even if they contain similar amounts of Spike protein mutations.[7] The severity of the disease caused by the Alpha and Beta variants did not differ significantly in SARS-CoV-2 infected Syrian hamsters.[8] However, Delta variant infection leads to more severe pathology in infected Syrian hamsters compared to the Alpha variant.[9] On the other hand, the recently emerging Omicron variant may have decreased virulence despite having a significantly higher number of mutations in the Spike protein than the Delta variant.[10] Therefore, there is a need to explore the inherent effects of SARS-CoV-2 mutations in the other proteins expressed by SARS-CoV-2.

Non-Spike proteins may provide a wider landscape for understanding the contribution of genetic changes in the behavior of SARS-CoV-2. The potential of non-Spike proteins in inducing behavioral changes in SARS-CoV-2 is exemplified by mutations in the nucleocapsid protein, a viral protein encapsulating the viral genetic material, such as the R203K and G204R mutations of the Alpha variant. These changes have been associated with increased production of viral particles leading to increased infectivity and virulence.[11]

Aside from structural proteins (Spike, Nucleocapsid, Membrane, and Envelope proteins), SARS-CoV-2 can express a set of accessory proteins. This subset of viral proteins may not be involved in in vitro viral replication, but they may play a role in mediating different virus-host interactions during infection.[12] There are at least nine accessory proteins that can be expressed by SARS-CoV-2: ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF10, and ORF14.[13] Different SARS-CoV-2 accessory proteins contribute to several viral processes. ORF9b interacts with TOM70 in transfected cells to facilitate immune modulation,[14] while ORF3a and ORF7b induce apoptosis in infected cells.[15,16] Thus, mutations in accessory proteins may give rise to possible mechanisms contributing to changes in the virulence of SARS-CoV-2 variants.

Using different bioinformatic platforms, this study aimed to identify significant amino acid mutations in the accessory proteins of SARS-CoV-2 variants and determine their likely effect on the structure and function of this set of proteins over time in silico. Furthermore, the study also aimed to determine how these mutations correlate with changes in the Spike protein.

2 | METHODOLOGY

Amino acid sequences, via the ViruClust web application,[17] were collected and divided into two data sets, which represent the date of collection of the samples: early (March 01, 2021–April 30, 2021) and late (July 01, 2021–August 31, 2021). These data sets represent the emergence of the Alpha and Delta variants. As of writing, these are the most recent waves of COVID-19 variants available in the ViruClust database. The geographical characteristics and lineage distribution of sequences in each group were identified. The amino acid sequences were also analyzed according to known variants: all the VOC (e.g., Alpha, Beta, Gamma, and Delta) and two representative VOI (e.g., Lambda and Mu). The ViruClust web application is available at http://gmql.eu/viruclust/.
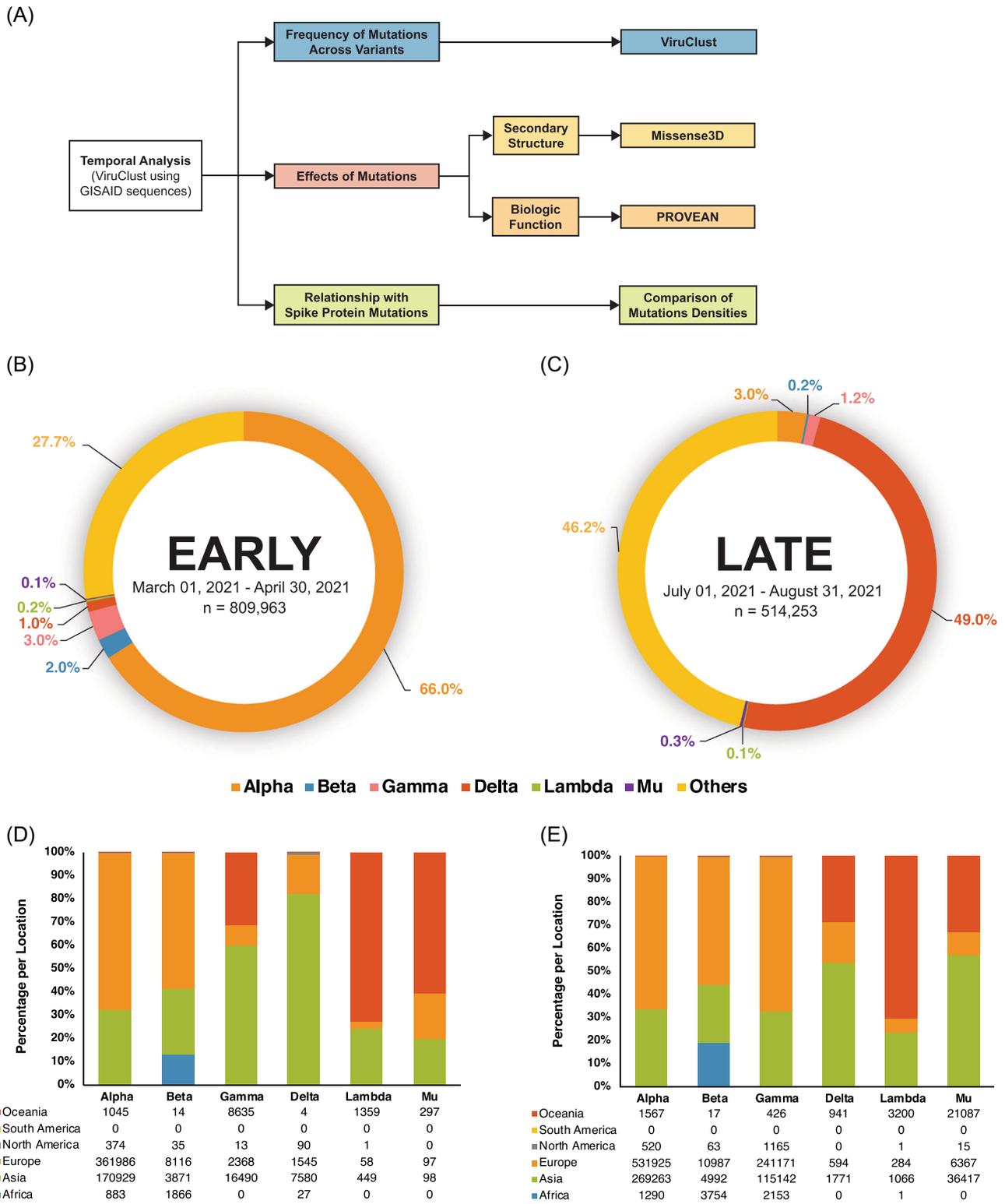
The amino acid changes were characterized and compared across time using different bioinformatic tools and quantitative measurements of mutations (Figure 1A). The frequency of mutations was determined using ViruClust. The effects of the most frequent mutations in each variant on the secondary structure and biologic function of each accessory protein were determined using Missense3D[18] and PROVEAN,[19] respectively. Lastly, the mutations in the accessory proteins were compared to Spike protein mutations by calculating the mutation densities for each protein. The mutation density is the number of mutations divided by the amino acid length of the corresponding accessory protein. Then, the correlation coefficient of the mutation densities across the time points was compared for each accessory protein across the different variants. The Missense3D and PROVEAN web servers are available at http://www.sbg.bio.ic.ac.uk/%7Emissense3d and http://provean.jcvi.org, respectively.

For further information, details of the methodology can be found in the supplementary documents of this article.

3 | RESULTS

3.1 | Characteristics of sequences obtained from GISAID

A total of 1,324,216 amino acid sequences were retrieved using ViruClust (Figure 1B,C). The number of sequences obtained for the early and late groups was 809,963 and 514,253, respectively. Most of the collected sequences were from Europe and Asia. There were no sequences obtained from South America, which is common to both data sets. Furthermore, the Alpha variant was the predominant variant in Group A at 66%, while the Delta variant was more frequently observed in Group B at 49%.

(A)

(B)

EARLY
March 01, 2021 - April 30, 2021
n = 809,963

27.7%
0.1%
0.2%
1.0%
3.0%
2.0%
66.0%

(C)

LATE
July 01, 2021 - August 31, 2021
n = 514,253

3.0% 0.2% 1.2%
46.2%
49.0%
0.3% 0.1%

■ Alpha ■ Beta ■ Gamma ■ Delta ■ Lambda ■ Mu ■ Others

(D)

	Alpha	Beta	Gamma	Delta	Lambda	Mu
■Oceania	1045	14	8635	4	1359	297
■South America	0	0	0	0	0	0
■North America	374	35	13	90	1	0
■Europe	361986	8116	2368	1545	58	97
■Asia	170929	3871	16490	7580	449	98
■Africa	883	1866	0	27	0	0

(E)

	Alpha	Beta	Gamma	Delta	Lambda	Mu
■Oceania	1567	17	426	941	3200	21087
■South America	0	0	0	0	0	0
■North America	520	63	1165	0	1	15
■Europe	531925	10987	241171	594	284	6367
■Asia	269263	4992	115142	1771	1066	36417
■Africa	1290	3754	2153	0	1	0

FIGURE 1 Methodology and preliminary results. Different bioinformatic tools were used to characterize the amino acid sequences of accessory proteins ORF3a, ORF6, ORF7a, ORF7b, and ORF8 via the ViruClust web application (A). The amino acid sequences were grouped into two time points: Early (March 2021–April 2021) and Late (July 2021–August 2021). The sequences were characterized based on the presence of SARS-CoV-2 variants (early B, late C) and the geographic distrib.

3.2 | Frequency of mutations in accessory protein genes across different variants

The ViruClust web application was used to identify missense and deletion mutations, which represent changes from the SARS-CoV-2 ancestral strain that was first identified in Wuhan, China. Moreover, the frequency of mutations was quantified as the percentage of the collected amino acid sequences for the specific variant with the specific point mutation plotted against the amino acid position of each accessory protein mutation. Only accessory proteins ORF3a, ORF6, ORF7a, ORF7b, and ORF8 were included in the study due to the availability of data in ViruClust. Therefore, the other SARS-CoV-2 accessory proteins such as ORF3b, ORF9b, ORF10, and ORF14 were not analyzed in this study.
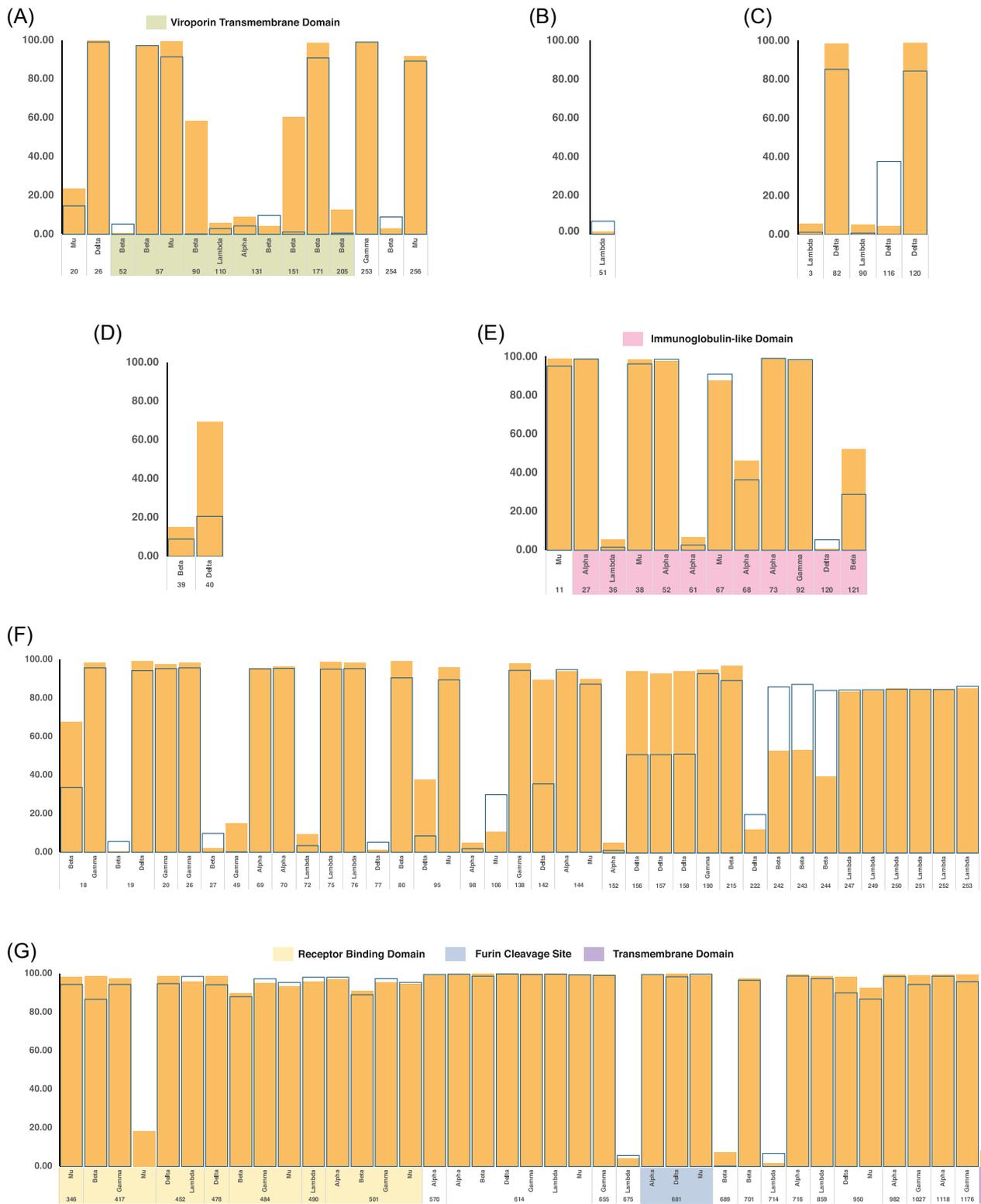
ORF3a is a 275 amino acid multi-pass transmembrane protein that forms an ion channel in the host cell membrane.[20] The Alpha and the Delta variant had the greatest number of ORF3a mutations in the early and late groups, respectively (Supporting Infomation: Figure 1). However, all ORF3a mutations in the Alpha variant occurred at less than 10% frequency (Figure 2A). The Delta variant had the highest increase (sevenfold) in the number of mutations between the two time points. The most frequent mutations in all variants are mostly within the ion channel domain of the protein, which includes the Q57H mutation (Figure 3A). This mutation is also present in both a VOC (Beta) and a VOI (Mu).

(A)

Variant	Mutation Frequency (%)			
	Early		Late	
Alpha	W131C	4	W131C	8
	L15F	3	T151I	3
	T151I	2	T223I	2
	P240S	1	L85F	1
	G100C	1	L15F	1
Beta	Q57H	100	S171L	100
	S171L	90	Q57H	100
	W131L	10	T151I	60
	G254*	9	V90I	60
	L52F	5	S205T	10
Gamma	S253P	100	S253P	100
	I118M	4	T223I	2
	Q57H	2	G251C	1
	G172C	1	S216L	1
	T269M	1	W193R	1
Delta	S26L	100	S26L	100
	Q57H	2	E239Q	3
	D238Y	2	P104S	1
	G172R	0.4	G224C	1
	G172C	0.3	D27Y	1
Lambda	A110S	3	A110S	6
	D155Y	1	R68K	4
	R68K	0.6	E242V	4
	T175I	0.5	D155Y	3
	I118V	0.5	G172R	2
Mu	Q57H	90	Q57H	100
	V256I	90	V256I	90
	I20M	20	I20M	20
	D222Y	2	D222Y	2
	K235R	1	K235R	2

(B)

Variant	Mutation Frequency (%)			
	Early		Late	
Alpha	T21I	0.2	L52V	0.2
	R20S	0.2	T21I	0.2
	Y31H	0.1	R20S	0.2
	W27L	0.1	W27L	0.2
	F7L	0.09	F7Y	0.1
Beta	K48E	0.1	P57L	2
	P57L	0.03	K48E	1
	T21I	0.01	T21I	0.4
	E54D	0.01	E54D	0.2
			I36T	0.1
Gamma	P57L	0.8	T21I	0.2
	F2S	0.2	P57L	0.1
	W27L	0.1	W27L	0.1
	K42N	0.04	I36T	0.1
	R20S	0.03	S25-	0.1
Delta	K48N	0.4	K48N	2
	E54*	0.3	E46V	1
	W27L	0.2	Q56R	0.2
	R20K	0.1	W27L	0.2
	E13D	0.08	P57L	0.1
Lambda	A51S	7	E54*	4
	T21I	1	A51S	2
	E54*	1	T21I	1
	L127F	0.4	V24I	1
	V24I	0.3	K48N	0.5
Mu			Y31F	0.2
			L4H	0.1
			V5I	0.1

(C)

Variant	Mutation Frequency (%)			
	Early		Late	
Alpha	Q62*	2	Q94L	3
	P84S	0.4	Q62*	3
	P84L	0.3	K119N	1
	S83L	0.3	P34S	1
	P34S	0.3	P34L	1
Beta	F63-	1.0	V93F	0.4
	V93F	0.9	T115I	0.3
	T115I	0.2	F63-	0.2
	Q62H	0.1	T61I	0.2
	L96F	0.1	H73Y	0.2
Gamma	V93F	5	P99L	0.9
	L96F	0.5	T11I	0.4
	T28I	0.3	T28I	0.4
	P99S	0.3	P99S	0.4
	I103-	0.2	S83L	0.3
Delta	V82A	90	T120I	100
	T120I	80	V82A	100
	L116F	40	L116F	4
	G38V	0.4	P45L	2
	P45L	0.4	V71I	1
Lambda	I3T	1	I3T	6
	Q90*	0.8	Q90*	5
	P34S	0.4	H47Y	3
	H47Y	0.2	T115I	3
	V24F	0.2	G38V	1
Mu	V62L	1	P45L	1
	L18F	0.8	L12F	1
	P45L	0.6	P84S	0.9
	P84S	0.2	G38V	0.8
			T120I	0.7

(D)

Variant	Mutation Frequency (%)			
	Early		Late	
Alpha	H42Y	0.3	A15S	0.4
	L32F	0.2	L4F	0.3
	L34M	0.1	L32F	0.3
	S31L	0.1	S31L	0.3
	W29C	0.1	I2N	0.2
Beta	E39*	9	E39*	20
	W29L	0.1	W29L	0.2
	T40I	0.02	T40I	0.2
	L4F	0.01	L4F	0.1
Gamma	L32F	0.07	L11F	0.6
	S31L	0.1	L25F	0.4
	M1I	0.05	L4F	0.2
	T40I	0.05	T40I	0.2
	H37N	0.03	S31L	0.2
Delta	T40I	21	T40I	70
	E39*	0.3	E33V	0.4
	H37Y	0.1	F13L	0.1
	A15S	0.1	F19L	0.1
	L32F	0.1	H42Y	0.0
Lambda			C41F	1
			L18V	0.2
			S5A	0.2
			S5L	0.2
			S31P	0.2
Mu	K21N	1	T40I	1.0
	E106*	0.4	E106*	0.4
	T229I	0.2	K21N	0.1
			T229I	0.1
			I27T	0.1

(E)

Variant	Mutation Frequency (%)			
	Early		Late	
Alpha	Y73C	100	Y73C	100
	Q27*	100	Q27*	100
	R52I	100	R52I	100
	K68*	40	K68*	43
	C61F	3	C61F	7
Beta	I121L	30	I121L	50
	E106*	4	P93H	3
	F120V	2	W45L	3
	P38S	0.8	F120V	1.0
	K68E	0.8	Q27*	0.9
Gamma	E92K	100	E92K	100
	R52I	0.4	G8*	0.6
	S67F	0.3	A65V	0.4
	A65V	0.2	S67F	0.4
	Y73C	0.1	Q72H	0.3
Delta	F120L	5	A65S	2
	D119Y	3	L60F	2
	G8V	2	F86-	0.5
	I121L	0.3	T87S	0.5
	E106*	0.2	S67F	0.5
Lambda	I121L	3	P36L	5
	P36L	2	I121L	4
	L242F	1	H17Y	2
	F120V	1	G66D	2
	A65V	0.3	F120V	1
Mu	P38S	100	T11K	100
	T11K	100	P38S	100
	S67F	90	S67F	90
	L85F	1	R52I	4
	I9T	0.2	E106Q	1

(F)

Variant	Mutation Frequency (%)			
	Early		Late	
Alpha	D614G	100	S982A	100
	A570D	100	D614G	100
	P681H	100	D1118H	100
	T716I	100	A570D	100
	S982A	100	T716I	100
Beta	D614G	100	D614G	100
	A701V	100	D80A	100
	D80A	90	K417N	100
	D215G	90	A701V	100
	N501Y	90	D215G	100
Gamma	D614G	100	D614G	100
	H655Y	100	V1176F	100
	N501Y	100	H655Y	100
	E484K	100	T1027I	100
	V1176F	100	L18F	100
Delta	D614G	100	D614G	100
	P681R	100	P681R	100
	L452R	90	T19R	100
	T478K	90	L452R	100
	T19R	90	T478K	100
Lambda	D614G	100	D614G	100
	L452Q	100	T859N	100
	F490S	100	L452Q	100
	T859N	100	F490S	100
	T76I	90	G75V	90
Mu	P681H	100	D614G	100
	D614G	100	P681H	100
	E484K	100	R346K	100
	N501Y	100	T95I	100
	R346K	90	N501Y	90

FIGURE 2 Selected accessory and spike protein mutations. The most frequent ORF3a (A), ORF6 (B), ORF7a (C), ORF7b (D), ORF8 (E), and Spike (F) mutations were selected to determine the effects of these mutations on secondary structure and biologic function.

FIGURE 3 Frequency of accessory and spike protein mutations. Point mutations of the different SARS-CoV-2 variants with frequencies greater than or equal to 5% and less than or equal to 100% were plotted based on the location of each mutation across the amino acid sequences of ORF3a (A), ORF6 (B), ORF7a (C), ORF7b (D), ORF8 (E) and the Spike protein (F and G). Highlighted regions correspond to the functional domains of each protein. Data shown are both from the early (March 01, 2021–April 30, 2021) and late (July 01, 2021–August 31, 2021) time points.

ORF6 is a nuclear protein with 61 amino acid residues. This protein may interact with the nuclear pore of SARS-CoV-2 infected cells via a putative binding site.[21] The mutations of this accessory protein occurred at less than 10% frequency in all the variants (Figures 2B and 3B and Supporting Information: Figure 2). The Delta variant accumulated the most ORF6 mutations between the time points with a fivefold increase. The most common mutations are T21I and W27L. The T21I mutation can be found in both a VOC (Alpha and Beta) and VOI (Lambda), while W27L is only present in VOCs (Alpha, Gamma, and Delta).

ORF7a is a single-pass transmembrane protein with 121 amino acid residues. This protein contains a putative binding site for lymphocyte antigen an endoplasmic localization signal.[22] The Alpha variant had the highest number of ORF7a mutations in the early time point, while the Delta variant has the highest number of mutations in the late time point (Supporting Information: Figure 3). Overall, the Delta variant had the highest number of ORF7a mutations, with a fourfold increase, in between the two time points. Only the Delta variant had ORF7a mutations that have frequencies greater than 50% in the amino acid sequences in both time points (Figure 2C). The most consistent mutations in this variant are T120I and V82A. The T120I is within the di-lysine motif of ORF7a, which is an important structure for endoplasmic reticulum localization of proteins[23] (Figure 3C). Mutations were also seen in the lymphocyte antigen-binding site of ORF7a in both the Alpha and Delta variants. However, the most frequent mutations in the Delta variant were not within this functional domain.

ORF7b is predicted to be a 43 amino acid single-pass transmembrane protein in the endoplasmic reticulum.[24] The data show that there is mostly low ORF7b mutation frequency of <1% in all variants except in the Delta variant (Figure 2D). Interestingly, the Lambda variant had no detectable mutations in the early time point (Supporting Information: Figure 4). Only the Delta variant had mutation frequencies that reached greater than 50% in both time points (Figure 3D). The most frequent mutation is T40I, which is only present in VOCs (Beta, Gamma, and Delta). This mutation has also increased frequency in Group B by threefold. The majority of the variations occur beyond the transmembrane domain.

ORF8 is a 121 amino acid single-pass transmembrane protein in the endoplasmic reticulum[25] with an Immunoglobulin-like (Ig-like) domain.[26] The Alpha variant has ORF8 mutations that are present in all sequences in both time points such as Y73C, Q27* (* = no amino acid equivalent due to formation of a stop codon), and R52I. Moreover, the Alpha variant was able to retain the same set of mutations in both time points. The most frequent mutations were not maintained in the Delta variant through time (Figure 2 and Supporting Information: Figure 5). The most common ORF8 mutations across the variants were R52I and I121L. The bulk of the mutations across the variants encompasses the Ig-like domain of ORF8 (Figure 3E).

All in all, the data show that point mutations occur in the functional domains of the majority of the accessory proteins. The results also reveal that ORF8 was the only accessory protein that did not accumulate high-frequency mutations between the two time

points in the Delta variant. This is in contrast with other variants, especially the Alpha variant, which had high-frequency ORF8 mutations. Interestingly, ORF6 was the only accessory protein that did not contain any high-frequency mutations across the variants.

3.3 | Effect of mutations on the secondary structure of SARS-CoV-2 accessory proteins

The 10 most frequent mutations per accessory protein were selected for the determination of the possible structural and functional effects of these mutations. The effects of the mutations on secondary structure were determined using the Missense3D online platform.

Most of the selected mutations in ORF3a were predicted to have no structural damage to the protein (Figure 4A). Only the T151I mutation of ORF3a was found to lead to structural damage, which is a buried H-bond breakage in both the Alpha and Beta variants.

All the selected mutations in ORF6 were predicted to have no structural damage to the protein (Figure 4B).

Most of the selected mutations in ORF7a were predicted to have no structural damage to the protein (Figure 4C). Only the G38V mutation was found to influence the secondary structure of ORF7a by introducing a steric clash in the Delta, Lambda, and Mu variants.

Most of the selected mutations in ORF7b were predicted to have no structural damage to the protein (Figure 4D). Only the S31P mutation was found to influence the secondary structure of ORF7b by introducing a steric clash in the Lambda variant.

Most of the selected mutations in ORF8 were predicted to have no structural damage to the protein (Figure 4E). Only the C61F mutation was found to influence the secondary structure of ORF8 by introducing a steric clash and disulfide breakage in the Alpha variant.

In summary, most of the point mutations in the accessory proteins did not contribute to a change in the secondary structure of these proteins. However, there were four mutations that can contribute to a change in the secondary structure such as T151I of ORF3a, G38V of ORF7a, S31P of ORF7b, and C61F of ORF8.

3.4 | Effect of mutations on the biologic function of SARS-CoV-2 accessory proteins

The PROVEAN software was used to determine if the selected mutations can independently lead to either a possible loss-of-function or neutral outcome. A loss-of-function outcome results when the mutation leads to a nonfunctional protein. Meanwhile, a neutral outcome implies that the protein has retained its function despite the presence of the mutation. Like the Missense3D platform, only the individual effects of each mutation can be analyzed by PROVEAN.

In ORF3a, the number of loss-of-function and neutral mutations were similar in number across the VOC. (Figure 5A). However, more neutral ORF3a mutations were identified in the VOIs. The T151I mutation of ORF3a is a loss-of-function mutation.

(A)

ORF3a				
Variant	Mutation	Effect on Secondary Structure	Mutation	Effect on Secondary Structure
Alpha	W131C	No structural damage	T223I	-
	L15F		L85F	-
	T151I	Buried H-bond breakage		
	P240S	-		
	G100C	-		
Beta	Q57H	No structural damage	T151I	Buried H-bond breakage
	S171L	No structural damage	V90I	-
	W131L	-		
	S205T	-		
	L52F	-		
Gamma	S253P	-	W193R	-
	I118M	-	T223I	-
	Q57H	No structural damage	G251C	-
	G172C	-	S216L	-
	T269M	-		
Delta	S26L	-	D27Y	-
	Q57R	No structural damage	E239Q	-
	D238Y	No structural damage	P104S	-
	G172R	-	G224C	No structural damage
	G172C	-		
Lambda	A110S	No structural damage	E242V	-
	D155Y	No structural damage	G172R	No structural damage
	R68K	No structural damage		
	T175I	No structural damage		
	I118V	No structural damage		
Mu	Q57H	No structural damage		
	V256I	-		
	I20M	-		
	D222Y	-		
	K235R	-		

(B)

ORF6				
Variant	Mutation	Effect on Secondary Structure	Mutation	Effect on Secondary Structure
Alpha	T21I	-	L52V	-
	R20S	-	F7Y	-
	Y31H	No structural damage		
	W27L	No structural damage		
	F7L	-		
Beta	K48E	-		
	P57L	-		
	T21I	-		
	E54D	-		
	I36T	No structural damage		
Gamma	P57L	-	T21I	-
	F2S	-	I36T	No structural damage
	W27L	No structural damage		
	K42N	-		
	R20S	-		
Delta	K48N	-	P57L	-
	Q56R	-	E46V	-
	W27L	No structural damage		
	R20K	-		
	E13D	-		
Lambda	A51S	-		
	T21I	-		
	K48N	-		
	L127F	-		
	V24I	No structural damage		
Mu	Y31F	No structural damage		
	L4H	-		
	V5I	-		

(C)

ORF7a				
Variant	Mutation	Effect on Secondary Structure	Mutation	Effect on Secondary Structure
Alpha	K119N	-	Q94L	No structural damage
	P84S	No structural damage	P34L	No structural damage
	P84L	No structural damage		
	S83L	-		
	P34S	No structural damage		
Beta	H73Y	No structural damage	T61I	No structural damage
	V93F	No structural damage		
	T115I	-		
	Q62H	No structural damage		
	L96F	No structural damage		
Gamma	V93F	No structural damage	P99L	-
	L96F	No structural damage	T11I	-
	T28I	No structural damage		
	P99S	-		
	S83L	-		
Delta	V82A	No structural damage	V71I	No structural damage
	T120I	-		
	L116F	-		
	G38V	Steric clash		
	P45L	No structural damage		
Lambda	I3T	-	T115I	-
	G38V	Steric clash		
	P34S	No structural damage		
	H47Y	No structural damage		
	V24F	No structural damage		
Mu	V62L	-	P45L	No structural damage
	L18F	-	L12F	-
	P45L	No structural damage	T120I	-
	P84S	No structural damage		
	G38V	Steric clash		

(D)

ORF7b				
Variant	Mutation	Effect on Secondary Structure	Mutation	Effect on Secondary Structure
Alpha	H42Y	-	A15S	No structural damage
	L32F	No structural damage	L4F	-
	L34M	-	I2N	
	S31L	No structural damage		
	W29C	No structural damage		
Beta	L4F	-		
	W29L	No structural damage		
	T40I	-		
Gamma	L32F	No structural damage	L11F	No structural damage
	S31L	No structural damage	L25F	No structural damage
	M1I	-	L4F	-
	T40I	-	S31L	No structural damage
	H37N	-		
Delta	T40I	-	H42Y	-
	F19L	No structural damage	E33V	-
	H37Y	-	F13L	No structural damage
	A15S	No structural damage		
	L32F	No structural damage		
Lambda	C41F	-		
	L18V	No structural damage		
	S5A	No structural damage		
	S5L	No structural damage		
	S31P	Steric clash		
Mu	K21N	-		
	T40I	-		
	T229I	-		
	I27T	No structural damage		

(E)

ORF8				
Variant	Mutation	Effect on Secondary Structure	Mutation	Effect on Secondary Structure
Alpha	Y73C	No structural damage		
	R52I	No structural damage		
	C61F	Disulphide breakage, Steric clash		
Beta	I121L	No structural damage	P93H	No structural damage
	W45L	No structural damage		
	F120V	No structural damage		
	P38S	No structural damage		
	K68E	No structural damage		
Gamma	E92K	No structural damage	Q72H	No structural damage
	R52I	No structural damage		
	S67F	No structural damage		
	A65V	No structural damage		
	Y73C	No structural damage		
Delta	F120L	No structural damage	A65S	No structural damage
	D119V	No structural damage	L60F	No structural damage
	G8V	-	T87S	No structural damage
	I121L	No structural damage		
	S67F	No structural damage		
Lambda	I121L	No structural damage	H17Y	-
	P36L	No structural damage	G66D	No structural damage
	L242F	-		
	F120V	No structural damage		
	A65V	No structural damage		
Mu	S67F	No structural damage	R52I	No structural damage
	L85F	-		
	K117N	-		
	I9T	-		
	E106Q	No structural damage		

(F)

Spike				
Variant	Mutation	Effect on Secondary Structure	Mutation	Effect on Secondary Structure
Alpha	D614G	No structural damage	D1118H	No structural damage
	A570D	No structural damage		
	T716I	No structural damage		
	S982A	No structural damage		
	P681H	No structural damage		
Beta	D614G	No structural damage	K417N	No structural damage
	A701V	No structural damage		
	D80A	Buried charge replaced, buried salt bridge breakage		
	D215G	No structural damage		
	N501Y	No structural damage		
Gamma	D614G	No structural damage	T1027I	No structural damage
	V1176F	-	L18F	No structural damage
	H655Y	No structural damage		
	N501Y	No structural damage		
	E484K	No structural damage		
Delta	D614G	No structural damage		
	P681R	No structural damage		
	L452R	No structural damage		
	T478K	No structural damage		
	T19R	No structural damage		
Lambda	D614G	No structural damage	G75V	No structural damage
	L452Q	No structural damage		
	F490S	No structural damage		
	T859N	No structural damage		
	T716I	No structural damage		
Mu	D614G	No structural damage	T95I	Buried H-bond breakage
	P681H	No structural damage		
	R346K	No structural damage		
	N501Y	No structural damage		
	E484K	No structural damage		

FIGURE 4 Effects of accessory and spike protein mutations on secondary structure. The structural consequences of each point mutation of the corresponding accessory or spike protein were determined using the Missense3D software. Each table represents an accessory protein and the selected point mutations for each variant.

(A)

ORF3a

Variant	Frequent Mutations	
Alpha	W131C, T223I, T151I, L85F, G100C	L15F, P240S
Beta	Q57H, T151I, W131L	S171L, V90I, S205T, L52F
Gamma	S253P, W193R, Q57H, G172C, T223I, G251C, S216L	T269M, I118M
Delta	P104S, Q57R, G224C, G172R, G172C	D27Y, E239Q, S26L, D238Y
Lambda	G172R, D155Y	E242V, A110S, R68K, T175I, I118V
Mu	Q57H, D222Y	I20M, V256I, K235R

(B)

ORF6

Variant	Frequent Mutations	
Alpha	T21I, R20S, L52V, W27L, F7Y, F7L	Y31H
Beta	P57L, T21I, I36T	K48E, E54D
Gamma	P57L, F2S, W27L, T21I, I36T, S25del, R20S	K42N
Delta	P57L, E46V, W27L, E13D	K48N, Q56R, R20K
Lambda	T21I	K48N, V24I
Mu	L4H	Y31F, V5I

(C)

ORF7a

Variant	Frequent Mutations	
Alpha	P34L, P84S, P84L, S83L, P34S	Q94L, K119N
Beta	F63del, H73Y, T61I	V93F, T115I, Q62H, L96F
Gamma	T28I, P99S, P99L, S83L, I103del	T11I, V93F, L96F
Delta	V82A, P45L, G38V, P45L	V71I, T120I, L116F
Lambda	I3T, G38V, P34S, H47Y, V24I	T115I
Mu	G38V, P45L, P84S	T120I, L12F

(D)

ORF7b

Variant	Frequent Mutations	
Alpha	H42Y, L32F, A15S, S31L, L4F, W29C	I2N, L34M
Beta	L4F, W29L	T40I
Gamma	L32F, S31L, M1I, L11F, L25F, L4F, S31L	T40I, H37N
Delta	H42Y, F19L, E33V, A15S, F13L, L32F	T40I, H37Y
Lambda	C41F, L18V, S31P	S5A, S5L
Mu	I27T	T40I

(E)

ORF8

Variant	Frequent Mutations	
Alpha	Y73C, C61F, R52I	
Beta	P93H, W45L, F120V, P38S	I121L, K68E
Gamma	Y73C, R52I	E92K, Q72H, S67F, A65V
Delta	F120L, D119V, G8V, L60F, F86del	A65S, T87S, S67F, I121L
Lambda	P36L, G66D, F120V, H17Y	I121L, A65V
Mu	R52I	S67F, E106Q, I9T

(F)

Spike

Variant	Frequent Mutations	
Alpha	T716I	D614G, A570D, D1118H, S982A, P681H
Beta		D614G, A701V, D80A, D215G, K417N, N501Y
Gamma		D614G, V1176F, H655Y, N501Y, T1027I, L18F, E484K
Delta		D614G, P681R, L452R, T478K, T19R
Lambda		D614G, L452Q, F490S, T859N, G75V, T76I
Mu		D614G, P681H, R346K, N501Y, T95I, E484K

(G)

SARS-CoV-2 Protein	LOF Mutations (%)	N Mutations (%)	LOF/N
ORF3a	54.55	45.45	1.20
ORF6	62.86	37.14	1.69
ORF7a	62.50	37.50	1.67
ORF7b	71.43	28.57	2.50
ORF8	55.88	44.12	1.27
Spike	2.78	97.22	0.03

FIGURE 5 Effects of accessory and Spike protein mutations on biologic function. The functional consequences of each point mutation to the corresponding accessory or Spike protein were determined using the PROVEAN software. Each table represents an accessory (A–E) or Spike (F) protein and the selected point mutations for each variant. The effect on biologic function was classified into loss-of-function () and neutral () mutations. Meanwhile, a summary table (G) was created to show the difference between the number of loss-of-function (LOF) and neutral (N) mutations in each accessory or Spike protein

In ORF6, the number of loss-of-function mutations were greater than neutral mutations (Figure 5B). Moreover, the VOCs have more loss of function mutations than neutral mutations.

In ORF7a and ORF7b, the number of loss-of-function mutations was greater than neutral mutations (Figure 5C,D). The G38V mutation of ORF7a and the S31P mutation of ORF7b are loss-of-function mutations.

In ORF8, the number of loss-of-function and neutral mutations were similar in number across all variants (Figure 5E). All the selected ORF8 mutations of the alpha variant were

identified as loss-of-function mutations including the C61F mutation.

All in all, the proportion of loss-of-function and neutral mutations are the same in ORF3a and ORF8. Meanwhile, ORF6, ORF7a, and ORF7b have more loss-of-function mutations than neutral mutations (Figure 5G). Interestingly, despite not having any effect on the secondary structure of ORF6, some ORF6 mutations may still lead to a loss of function. T151I of ORF3a, G38V of ORF7a, S31P of ORF7b, and C61F of ORF8 were shown to be loss-of-function mutations.

3.5 | Correlation of accessory protein mutations with Spike protein mutations

The same analyses on the frequency of mutations and the effect of the mutations on secondary structure and biologic function were done using the Spike protein sequences of each variant.

The D614G mutation is the only mutation that is present in all variants (Figure 2F and Supporting Information: Figure 6). The Delta variant is the only variant that was able to preserve its most frequent mutations (D614G, P681R, L452R, T478K, T19R) across time.

In terms of location, most of the mutations occur in the different structural and functional domains of the Spike protein (Figure 3F). All variants contain mutations in the receptor-binding domain. All variants except the Beta and Lambda variants contain mutations in the transmembrane domain. All variants except the Beta variant contain mutations in the furin cleavage site of the Spike protein.

The effect of most of the selected mutations in the Spike protein was predicted to have no structural damage to the protein

(Figure 4F). The D80A and T95I are mutations that may influence the secondary structure of the Spike protein. The former mutation may lead to the replacement of a buried charge and the breakage of a salt bridge in the Beta variant. Meanwhile, the latter mutation may lead to the breakage of a buried hydrogen bond in the Mu variant.

Most of the selected mutations of the Spike protein are neutral mutations except for the T716I mutation in the Alpha variant (Figure 5F). The D80A and T95I mutations are neutral mutations.

To determine the possible correlations between the mutations of accessory proteins to the Spike protein, the mutation densities of these proteins were calculated and compared across the different variants (Figure 6A,B). Correlation of the mutation densities of the accessory proteins to the Spike protein revealed that ORF7b was highly correlated ($r > 0.9$) with the Spike protein ($r = 0.95$) in the early group (Figure 6C). Meanwhile, all the accessory proteins were highly correlated with the Spike protein in the late group wherein ORF6 had the highest correlation ($r = 0.99$) (Figure 6D). ORF8 was the least correlated accessory protein with the Spike protein in both time points across the different variants.

4 | DISCUSSION

This study identified several amino acid mutations in the accessory proteins of SARS-CoV-2 variants. Despite the limited contribution of these mutations to the secondary structure of these proteins, some are still good candidates for further investigation such as the Q57H and T151I mutations of ORF3a; T21I and W27L mutations of ORF6; G38V, V82A, and T120I mutations of ORF7a; S31P and T40I

(A)

(B)

ORF3a — ORF6 — ORF7a — ORF7b — ORF8 — Spike

(C)

Correlation Coefficients						
	Spike	ORF3a	ORF6	ORF7a	ORF7b	ORF8
Spike	1					
ORF3a	0.882526	1				
ORF6	0.821189	0.528546	1			
ORF7a	0.818072	0.657131	0.867585	1		
ORF7b	0.945581	0.754711	0.772775	0.80851	1	
ORF8	0.007731	0.276042	0.03582	-0.006632	-0.310982	1

(D)

Correlation Coefficients						
	Spike	ORF3a	ORF6	ORF7a	ORF7b	ORF8
Spike	1					
ORF3a	0.979849	1				
ORF6	0.993783	0.989627	1			
ORF7a	0.974299	0.998319	0.989693	1		
ORF7b	0.97152	0.991724	0.991206	0.996957	1	
ORF8	0.947599	0.98544	0.965259	0.987819	0.981592	1

FIGURE 6 Comparison of mutation densities between SARS-CoV-2 accessory proteins and the Spike protein. The mutation densities of each accessory protein per variant were compared across the early (A) and late (B) time points. The correlation coefficient between the accessory and Spike proteins was calculated in both the early (C) and late (D) time points.

mutations of ORF7b; and R52I, C61F, and I121L mutations of ORF8. These may influence the function of these accessory proteins because these are found within known functional domains and are shown to be highly frequent in most VOC. It should be noted that there are still no experimental studies that can provide evidence on the effects of these mutations. Thus, the effects of these mutations should be verified through experimental studies.

This study also determined the correlation between the SARS-CoV-2 accessory and Spike protein mutations. Most Spike protein mutations occur in the different structural and functional domains of the Spike protein in different relative frequencies across the different variants. The DG14G mutation is the only Spike protein mutation present in all SARS-CoV-2 variants, which is consistent with previous reports.[27] The effect of most of the selected Spike protein mutations, along with D614G, were predicted to be neutral mutations that may not produce any structural damage to the protein. These findings are in contrast with the observed experimental effects of D614G of increasing infectivity, transmission, and replication kinetics of SARS-CoV-2 in vitro and in vivo.[28] These structural and functional damages were not apparent since only the effects of single mutations were examined and investigated. Therefore, these results should also be verified experimentally.

The correlations with the Spike protein mutations highlighted ORF6 and ORF8. ORF8 is not well conserved through time, which suggests that there may be no selection pressure to retain the structure and function of ORF8. Furthermore, the mutation density of this protein did not correlate well with the Spike protein relative to the other accessory proteins, which is also observed in the ORF8 of other beta coronavirus species.[29] The lack of consistency with the Spike protein may be because ORF8 may not contribute to the increased virulence in the Delta variant. Even if initial studies reveal that it may contribute to immune evasion[28-30] cytokine storm,[31] the contribution of ORF8 on the pathogenesis may not be reflected in real life. This observation is consistent with previous studies showing the presence of genomes with or without truncated ORF8[32] where these deletions were associated with mild disease.[33] Lastly, the highly random changes of ORF8 also implies that more studies are needed if ORF8 is to be considered as a therapeutic target.

On the other hand, ORF6 was the only accessory protein that did not have any high-frequency mutations across the SARS-CoV-2 variants. Thus, the structure of the protein may be retained as SARS-CoV-2 evolves over time. The putative role of ORF6 in the pathogenesis of SARS-CoV-2 may be in mediating immune evasion by disrupting the activity of transcription factors associated with the innate immune response.[21] Like ORF8, so far, there is no evidence linking ORF6 mutations with increased infection or severity of COVID-19. Thus, experimental studies should be done to verify the effects of the mutations. However, the strong correlation between the mutation of Spike protein and ORF6 suggests that ORF6 can be used as a possible surrogate marker for the Spike protein.

The findings of this study are consistent results with other in silico investigations on SARS-CoV-2 mutations. The highly variable mutations in ORF8 support that the gene is a mutational hotspot.[34] The Q57H and T151I mutations identified in ORF3a were also observed by another study utilizing a different bioinformatic approach.[35] Additionally, the Q57H mutation in ORF3a is also consistently observed in a mutational analysis done using sequences obtained from the early part of the COVID-19 pandemic.[36]

Despite the numerous findings on accessory protein mutations in this study, interpretation of these data should be done with caution due to some limitations. First, the time points used in the study reflect the transition when the Delta variant overtook the Alpha variant in terms of the number of sequences, which explains the discrepancy in the number of samples collected per variant between the two time points. Despite the uneven number of sequences, there were still retained amino acid sequences in ORF6. Secondly, there are effects of mutations that may have been missed by the bioinformatic tools since they only predict the effects of individual mutations and do not account for the possible synergistic effects of multiple mutations in influencing the structure and function of the accessory proteins. Lastly, it should be emphasized that these tools can only make predictions on the effects of mutations. Therefore, experimental validation must be done to ascertain if these mutations lead to changes in function.

Nonetheless, this study was still able to identify significant amino acid substitutions in the accessory proteins (Q57H and T151I in ORF3a; T21I and W27L in ORF6; G38V, V82A, and T120I in ORF7a; S31P and T40I in ORF7b; and R52I, C61F, and I121L in ORF8) that may alter their functions. Given the preliminary evidence that these accessory proteins largely contribute to immune evasion of SARS-CoV-2 in humans, these accessory proteins may play a role in facilitating viral pathogenesis. Thus, these mutations can be candidates for experimental studies to further broaden the understanding of the effects of these changes in viral processes downstream of infection.

5 | CONCLUSION

SARS-CoV-2 continues to evolve into different variants across time. The Spike protein is the most implicated SARS-CoV-2 protein for the infectivity and pathogenesis of the virus, which may provide a skewed perspective because past and current research are mostly focused on this protein. Accessory proteins may also play a vital role in facilitating viral processes. This study was able to monitor the temporal changes in accessory proteins ORF3a, ORF6, ORF7a, ORF7b, and ORF8 using amino acid sequences from two-time points during the transition from the Alpha variant to Delta variant dominance. All accessory proteins obtained several mutations that can be good candidates for further experimental evaluation. Knowing the effects of these mutations may facilitate the discovery of potential therapeutic targets, disease mechanisms, and factors that drive mutations. Furthermore, as some accessory proteins are correlated with the Spike protein, the mutations in future SARS-CoV-2 variants may be predicted. The succeeding SARS-CoV-2

variant would likely have mutations in ORF8 but may remain free from ORF6 mutations.

AUTHOR CONTRIBUTIONS

Christian Alfredo K. Cruz and Paul Mark B. Medina contributed in the conceptualization, data analysis, writing, and revision of this study article. Christian Alfredo K. Cruz performed the data collection and produced the figures of this study. Paul Mark B. Medina contributed to the critical evaluation of the data and figures of this manuscript.

CONFLICT OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY STATEMENT

The data generated by this study are available in the main text and supplementary materials. Raw data can be retrieved from ViruClust, Missense3D, and PROVEAN web applications at http://gmql.eu/viruclust/http://www.sbg.bio.ic.ac.uk/%7Emissense3d and http://provean.jcvi.org, respectively.

ORCID

Christian Alfredo K. Cruz [ORCID] http://orcid.org/0000-0003-2849-8683
Paul Mark B. Medina [ORCID] http://orcid.org/0000-0001-6116-1818

REFERENCES

1. Global Consultation on a Decision Framework for Assessing the Impact of SARS-CoV-2 Variants of Concern on Public Health Interventions, 29 March 2021: Meeting Report: World Health Organization; 2021.
2. Hou YJ, Chiba S, Halfmann P, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science*. 2020;370:1464-1468.
3. Barton MI, MacGowan SA, Kutuzov MA, Dushek O, Barton GJ, van der Merwe PA. Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *eLife*. 2021;10:e70658.
4. Liu Y, Liu J, Johnson BA, et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *Cell Reports*. 2022;39:110829.
5. Liu Y, Liu J, Plante KS, et al. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*. 2021;602:294-299.
6. Garcia-Beltran WF, Lam EC, Denis KS, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*. 2021;184(9):2372-2383.
7. Ong SWX, Chiew CJ, Ang LW, et al. Clinical and virological features of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern: A retrospective cohort study comparing B. 1.1. 7 (Alpha), B. 1.351 (Beta), and B. 1.617. 2 (Delta). *Clin Infect Dis*. 2021:1-9.
8. Abdelnabi R, Boudewijns R, Foo CS, et al. Comparing infectivity and virulence of emerging SARS-CoV-2 variants in Syrian hamsters. *EBioMedicine*. 2021;68:103403.
9. Saito A, Irie T, Suzuki R, et al. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature*. 2021;602:1-10.
10. Wolter N, Jassat W, Walaza S, et al. Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: a data linkage study. *The Lancet*. 2022;399:437-446.
11. Wu J, Shi Y, Pan X, et al. SARS-CoV-2 ORF9b inhibits RIG-I-MAVS antiviral signaling by interrupting K63-linked ubiquitination of NEMO. *Cell Rep*. 2021;34(7):108761.
12. Laguette N, Benkirane M. Shaping of the host cell by viral accessory proteins. *Front Microbiol*. 2015;6:142.
13. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. *Cell*. 2020;181(4):914-921.
14. Thorne LG, Bouhaddou M, Reuschl AK, et al. Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature*. 2022;602:487-495.
15. Ren Y, Shu T, Wu D, et al. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cell Mol Immunol*. 2020;17(8):881-883.
16. Yang R, Zhao Q, Rao J, et al. SARS-CoV-2 accessory protein ORF7b mediates tumor necrosis factor-α-induced apoptosis in cells. *Front Microbiol*. 2021;12:654709.
17. Cilibrasi L, Pinoli P, Bernasconi A, Canakoglu A, Chiara M, Ceri S. ViruClust: direct comparison of SARS-CoV-2 genomes and genetic variants in space and time. *Bioinformatics*. 2022;38:1988-1994.
18. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJ. Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? *J Mol Biol*. 2019;431(11):2197-2212.
19. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745-2747.
20. Kern DM, Sorum B, Mali SS, et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat Struct Mol Biol*. 2021;28:573-582.
21. Miorin L, Kehrer T, Sanchez-Aparicio MT, et al. SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *Proc Natl Acad Sci*. 2020;117(45):28344-28354.
22. Zhou B, Thao TTN, Hoffmann D, et al. SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature*. 2021;592(7852):122-127.
23. Bazan B, Wiktor M, Maszczak-Seneczko D, Olczak T, Kaczmarek B, Olczak M. Lysine at position 329 within a C-terminal dilysine motif is crucial for the ER localization of human SLC35B4. *PLoS One*. 2018;13(11):e0207521.
24. Schaecher SR, Mackenzie JM, Pekosz A. The ORF7b protein of severe acute respiratory syndrome coronavirus (SARS-CoV) is expressed in virus-infected cells and incorporated into SARS-CoV particles. *J Virol*. 2007;81(2):718-731.
25. Flower TG, Buffalo CZ, Hooy RM, Allaire M, Ren X, Hurley JH. Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *Proc Natl Acad Sci*. 2021;118:2.
26. Tan Y, Schneider T, Leong M, Aravind L, Zhang D. Novel immunoglobulin domain proteins provide insights into evolution and pathogenesis of SARS-CoV-2-related viruses. *mBio*. 2020;11(3):e00760-20.
27. Chakraborty C, Saha A, Sharma AR, Bhattacharya M, Lee SS, Agoramoorthy G. D614G mutation eventuates in all VOI and VOC in SARS-CoV-2: is it part of the positive selection pioneered by Darwin? *Mol Ther Nucleic Acids*. 2021;26:237-241.
28. Zhang J, Xiao T, Cai Y, et al. Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science*. 2021;374(6573):1353-1360.
29. Pereira F. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infect Genet Evol*. 2020;85:104525.

30. Zhang Y, Chen Y, Li Y, et al. The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. *Proc Natl Acad Sci*. 2021;118(23):1-12.

31. Lin X, Fu B, Yin S, et al. ORF8 contributes to cytokine storm during SARS-CoV-2 infection by activating IL-17 pathway. *Iscience*. 2021;24(4):102293.

32. Hassan SS, Kodakandla V, Redwan EM, et al. An issue of concern: unique truncated ORF8 protein variants of SARS-CoV-2. *PeerJ*. 2022;10:e13136.

33. Young BE, Fong SW, Chan YH, et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet*. 2020;396(10251):603-611.

34. Badua CLDC, Baldo KAT, Medina PMB. Genomic and proteomic mutation landscapes of SARS-CoV-2. *J Med Virol*. 2021;93(3):1702-1721.

35. Azad GK, Khan PK. Variations in Orf3a protein of SARS-CoV-2 alter its structure and function. *Biochem Biophys Rep*. 2021;26:100933.

36. Chaudhuri D, Majumder S, Datta J, Giri K. In silico study of mutational stability of SARS-CoV-2 proteins. *Protein J*. 2021;40(3):328-340.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.