ORIGINAL ARTICLE

Genomic Variant Surveillance of SARS-CoV-2 Positive Specimens Using a Direct PCR Product Sequencing Surveillance (DPPSS) Method

Nicole Ann L. Tuberon, RCh,[1] Francisco M. Heralde III, PhD,[1] Catherine C. Reportoso,[1,2] Arturo L. Gaitano III, MSc, RCh,[1,3] Wilmar Jun O. Elopre, MSc, RDN, RND[1,4] and Kim Claudette J. Fernandez[1]

[1]*Department of Biochemistry and Molecular Biology, College of Medicine, University of the Philippines Manila, Manila, Philippines*
[2]*Department of Biochemistry, University of the East Ramon Magsaysay Memorial Medical Center, Quezon City, Philippines*
[3]*Immunoinformatics and Novel Synthetic Peptide Immunogens Research (INSPIRe) Group, Chemistry Department, College of Science, Adamson University, Manila, Philippines*
[4]*Department of Nutrition and Dietetics, Central Mindanao University, Bukidnon, Philippines*

ABSTRACT

Background and Objective. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the causative agent of COVID-19 has significantly challenged the public health landscape in late 2019. After almost 3 years of the first ever SARS-CoV-2 case, the World Health Organization (WHO) declared the end of this global health emergency in May 2023. Although, despite the subsequent drop of COVID-19 cases, the SARS-CoV-2 infection still exhibited multiple waves of infection, primarily attributed to the appearance of new variants. Five of these variants have been classified as Variants of Concern (VOC): Alpha, Beta, Gamma, Delta, and the most recent, Omicron. Therefore, the development of methods for the timely and accurate detection of viral variants remains fundamental, ensuring an ongoing and effective response to the disease. This study aims to evaluate the feasibility of the application of an in-house approach in genomic surveillance for the detection of SARS-CoV-2 variants using in silico designed primers.

Methods. The primers used for the study were particularly designed based on conserved regions of certain genes in the virus, targeting distinct mutations found in known variants of SARS-CoV-2. Viral RNA extracts from nasopharyngeal samples (n=14) were subjected to quantitative and qualitative tests (Nanodrop and AGE). Selected samples were then analyzed by RT-PCR and amplicons were submitted for sequencing. Sequence alignment analysis was carried out to identify the prevailing COVID-19 variant present in the sample population.

Results. The study findings demonstrated that the in-house method was able to successfully amplify conserved sequences (spike, envelope, membrane, ORF1ab) and enabled identification of the circulating SARS-CoV-2 variant among the samples. Majority of the samples were identified as Omicron variant. Three out of four designed primers effectively bound into the conserved sequence of target genes present in the sample, revealing the specific SARS-CoV-2 variant. The detected mutations characterized for Omicron found in the identified lineages included K417N, S477N, and P681H which were also identified as mutations of interest. Furthermore, identification of the B.1.448 lineage which was not classified in any known variant also provided the potential of the developed in-house method in detecting unknown variants of COVID-19.

Conclusion. Among the five VOCs, Omicron is the most prevalent and dominant variant. The in-house direct PCR product sequencing surveillance (DPPSS) method provided an alternative platform for SAR-CoV-2 variant analysis which is accessible and affordable than the conventional diagnostic surveillance methods and the whole genome sequencing. Further evaluation and improvements on the oligonucleotide primers may offer significant contribution to the development of a specific and direct PCR-based detection of new emerging COVID-19 variants.

Keywords: SARS-CoV-2, Polymerase Chain Reaction (PCR), oligonucleotide primers, genomic surveillance, Omicron variant, B.1.448 lineage

Oral Research Presentation – 38th Philippine Chemistry Congress / 2024 International Conference on Chemistry at SMX Convention Center Davao, Lanang, Davao City, July 22-24, 2024.

Corresponding author: Nicole Ann L. Tuberon, RCh
Department of Biochemistry and Molecular Biology
College of Medicine, University of the Philippines Manila
547 Pedro Gil St., Ermita, Manila 1000, Philippines
Email: nltuberon@up.edu.ph
ORCiD: https://orcid.org/0009-0009-0076-9577

INTRODUCTION

The global public health landscape has been significantly challenged by the COVID-19 pandemic, since its causative agent, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first identified in December 2019.[1] As of August 2023, recorded cases worldwide have surpassed 760 million, with over 6.9 million reported deaths.[2] However, it is widely acknowledged that the actual figures are likely higher. In the Philippines, COVID-19 initially ranked as the third leading cause of death in 2021 but subsequently dropped to the eleventh position in 2022.[3] Despite the World Health Organization's (WHO) declaration of the end of the global health emergency related to COVID-19 in May 2023, the imperative for sustained monitoring and surveillance persists.[4] This is particularly crucial due to the persistent risk of new variants emerging, potentially leading to renewed spikes in cases and fatalities. Therefore, the development of methods for the timely and accurate detection of viral variants remains fundamental, ensuring an ongoing and effective response to the disease.

COVID-19 has exhibited multiple waves of infection, primarily attributed to the appearance of new variants.[5] Five of these variants have been classified as Variants of Concern (VOC): Alpha, Beta, Gamma, Delta, and the most recent, Omicron.[6] Each variant carries a distinct set of mutations in the viral genome, impacting pathogenicity, transmissibility, and morbidity.[7] Particularly, the Omicron variant has sparked global concern due to heightened transmissibility and increased resistance to vaccine-induced immunity.[8]

The SARS-CoV-2 proteome comprises four structural proteins: Spike (S) glycoprotein, Envelope (E) glycoprotein, Nucleocapsid (N) protein, and Membrane (M) glycoprotein.[9] Mutations primarily occur in the receptor binding domain within the spike glycoprotein, influencing the virus's behavior.[10] Notably, these genetic variations pose challenges to the efficacy of existing RT-PCR kits in detecting COVID-19 variants.[11,12] For example, mutations in annealing sites can impede primer attachment, reducing PCR test specificity and leading to false-negative results.[13] To address these challenges, strategic measures are recommended including conducting tests during the optimal time frame of viral replication and designing primers and probes that target conserved sequences in the genome.[14]

Evaluating the epidemiological characteristics of novel variants remains a formidable task. In the Philippines, VOCs are detected via whole genome sequencing; however, only a small number of samples can be processed at a time due to constraints in resources and equipment. As an alternative to whole genome sequencing, this study aims to develop PCR primers specifically designed to detect SARS-CoV-2 variants for use in genomic surveillance. A direct PCR product sequencing surveillance (DPPSS) method could serve as a sustainable and cost-effective approach for monitoring the frequency and spread of SARS-CoV-2 variants, as well as provide real-time information about genetic diversity and evolution of viruses. For these reasons, development of new genomic surveillance methods proves to be an invaluable asset in real-time monitoring and characterization of SARS-CoV-2 variants, contributing to the overall effective management of COVID-19. The present study aims to evaluate the effectiveness of the developed in-house approach, the DPPSS method in the detection of SARS-CoV-2 variants.

MATERIALS AND METHODS

Study Design

This is a descriptive study which aims to use a PCR-based method to determine the frequency of SARS-CoV-2 variants among SARS-CoV-2 positive nasopharyngeal / oropharyngeal specimens. The data obtained from this preliminary study will be used as a starting point for the validation of the primers in variant detection using larger sample size.

Primer Designing Process

Genomic data of SARS-CoV-2 wild type and variants were retrieved from NCBI GenBank and GISAID EpiCOvTM databases. SARS-CoV-2 Wuhan-Hu1 (NC_045512.2), a Wuhan isolate, was used as the reference genome. To design primers, NCBI Primer-BLAST was utilized and candidate primers sets from the ORF1ab, S, E, and M genes were obtained. In order to check whether the gene-specific primer sets are able to target the conserved regions and amplify the variant mutations, complete genome of SARS-CoV-2 genetic variants were aligned with selected nucleotide sequences (ORF1ab, S, E and M genes) from the reference genome using Clustal W codon-by-codon available in MEGA version X. The primer sets designed and used in this study were listed in Table 1. Primer synthesis was outsourced to a commercial service provider (Macrogen, Korea).

Sample Processing and Selection

Viral RNA samples (n=30) were extracted and purified from archived, inactivated SARS-CoV-2 positive nasopharyngeal swab specimens (UPMREB Code 2021-139-01) using PureLink™ Viral RNA/DNA Mini Kit following the manufacturer's instruction. The purified RNA isolates were then quantified using a Thermo Scientific NanoDrop® Spectrophotometer. The isolates with the highest purity, or the highest concentration, or a combination of both, were subjected to verification using Agarose Gel Electrophoresis (AGE). The inclusion of the Nanodrop and AGE analyses before subjecting the samples to the RT-PCR analysis is considered to be part of the quality assurance of the developed DPPSS method to ensure that RNA isolates are indeed present in the samples before the execution of sequence amplification. Among the initial 30 samples, only 14 samples were selected to be further analyzed.

Agarose Gel Electrophoresis

Qualitative analysis of the RNA isolates and DNA amplicons were carried out using Agarose Gel Electrophoresis (AGE) analysis. A 1% agarose gel was prepared using 1X Tris-acetate buffer (TAE) as diluent. The gel electrophoresis machine was set at 100 volts for the entire analysis.

RT-PCR Analysis

Using the in silico designed primers tabulated in Table 1, fourteen (14) selected samples listed in Table 2 were subjected to PCR using a QIAGEN OneStep RT-PCR Kit, following the manufacturer's instruction. Optimization was performed using a touchdown PCR protocol with annealing temperatures ranging from 55˚C to 58˚C to determine optimum annealing temperature of the primers.

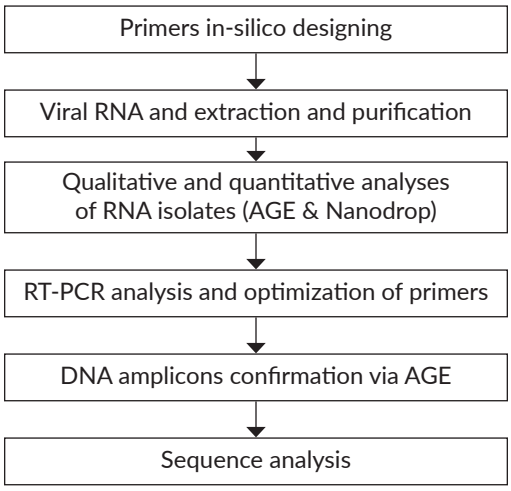

Figure 1. Overview of the methodology.

After the samples were amplified, DNA amplicons were further subjected to qualitative evaluation using AGE analysis to assess if the designed primers were able to successfully amplify conserved regions in the virus. Then the PCR products that matched the theoretical molecular weights and amplicon size were sent to Macrogen, Korea for product sequencing.

Sequence Analysis

Sequenced results were analyzed via BLAST (NCBI). The length used for the sequence alignment was derived based on the quality of the chromatogram profile of each sample both forward and reverse as a result of the product sequencing. The sequences were then aligned with existing SARS-CoV-2 sequences that were already available in the NCBI database, and their lineages were obtained. Finally, these lineages were used to identify the specific variants of the samples using the Pango classification. Overview of the overall methodology was summarized in Figure 1.

RESULTS

Qualitative and Quantitative Analysis

From October to December 2023, SARS-CoV positive samples were tested for SARS-CoV-2 variants using a direct PCR product sequencing surveillance (DPPSS) method. Viral RNA extraction was done in these samples followed by NanoDrop quantification using Thermo Scientific NanoDrop Spectrophotometer. The ratio of absorbance at 260nm and 280nm was used to assess the purity of RNA. A ratio of ~2.0 was accepted as "pure" RNA, but 1.8-2.1 was considered acceptable. In NanoDrop quantification of the samples, the A260/A280 ratio ranged from 1.74-1.97 (Table 3). As can be seen from Table 3, only 14 viral RNA samples from the

Table 1. In-silico Developed Oligonucleotide Primers

Target Gene	Primer Name	Primer Sequence (5' → 3')	Product Size (bp)
ORF1ab	SARS-CoV-2_Var_O1	Forward primer O1 ACCAATGTGCTATGAGGCCC	818
		Reverse primer O1 CATCACCCAACTAGCAGGCA	
S	SARS-CoV-2_Var_S1	Forward primer S1 CAAATCGCTCCAGGGCAAAC	1362
		Reverse primer S1 GTGGCAAAACAGTAAGGCCG	
S	SARS-CoV-2_Var_S2	Forward primer S2 GTCCTTCCCTCAGTCAGCAC	655
		Reverse primer S2 ACTCCTTTGAGCACTGGCT	
E, M	SARS-CoV-2_Var_EM	Forward primer EM1 CGATTGTGTGCGTACTGCTG	658
		Reverse primer EM1 AGGTCCTTGATGTCACAGCG	

Table 2. Summary of SARS-CoV-2 Positive Specimens

Sample	Primer Name	Sample Code
P	SARS-CoV-2_Var_O	P_O
A	SARS-CoV-2_Var_S1	A_S1
B		B_S1
D		D_S1
W		W_S1
F	SARS-CoV-2_Var_S2	F_S2
I		I_S2
AP		AP_S2
AU		AU_S2
AX		AX_S2
K	SARS-CoV-2_Var_EM	K_EM
N		N_EM
BL		BL_EM
BQ		BQ_EM

initially 30 samples were reported and selected to be further analyzed. Since, other samples were not pure enough and had little to no RNA concentration determined via Nanodrop analysis. Among these samples, the determined concentration of the RNA isolates ranged from 8.7 ng/UI to 54.2 ng/UI. This qualitative analysis was carried out as part of the quality assurance of the developed in-house method.

The 14 RNA isolates that passed both qualitative and quantitative analyses (Nanodrop and AGE analyses) were then subjected to RT-PCR process using the optimized touchdown method (Table 4), with a corresponding primer used for amplification for each sample. The chosen optimized annealing temperature used in the study was 58°C since the designed primers were mostly high in GC content. Thus, a higher temperature for primer annealing was used as the optimal annealing temperature.[15]

Before sending the samples to Macrogen for sequencing, each DNA amplicon was subjected to AGE analysis together with a molecular marker to qualitatively evaluate if the designed primers successfully amplified the RNA isolates using the touchdown method. AGE post-PCR showed presence of DNA fragments in primers SM1, SM2, EM, and O (Figure 2). As can be seen in Figure 2, primer O was successful in amplifying the corresponding ORF1ab gene present in sample P. On the other hand, samples A, B, D, and W1, were successfully amplified using the SM1 primer, which primarily targets the S1 subunit of the S protein. Primer SM2, on the other hand, effectively amplified S2 subunit

Table 3. NanoDrop Quantification of Samples (Pre-PCR)

Sample Code	RNA concentration (ng/UI)	Abs 260	Abs 280	260/280
P_O1	54.2	1.356	0.749	1.81
A_S1	54.2	1.356	0.749	1.81
B_S1	28.0	0.700	0.376	1.86
D_S1	10.6	0.265	0.151	1.75
W_S1	10.1	0.253	0.128	1.97
F_S2	54.2	1.356	0.749	1.81
I_S2	10.6	0.265	0.151	1.75
AP_S2	11.3	0.283	0.156	1.82
AU_S2	8.7	0.216	0.124	1.74
AX_S2	24.1	0.602	0.330	1.82
K_EM	54.2	1.356	0.749	1.81
N_EM	10.6	0.265	0.151	1.75
BL_EM	8.7	0.216	0.124	1.74
BQ_EM	24.1	0.602	0.330	1.82

Table 4. Optimized RT-PCR Profile

Process	Temperature (°C)	Time (minutes)
Reverse transcription	50	30
Initial denaturation	95	15
Denaturation	94	1
Annealing	55	1
Elongation	72	1
Repeat process 10X		
Denaturation	94	1
Annealing	58	1
Elongation	72	1
Repeat process 20X		
Final elongation	72	10
Cooling	4	∞
Cool down		

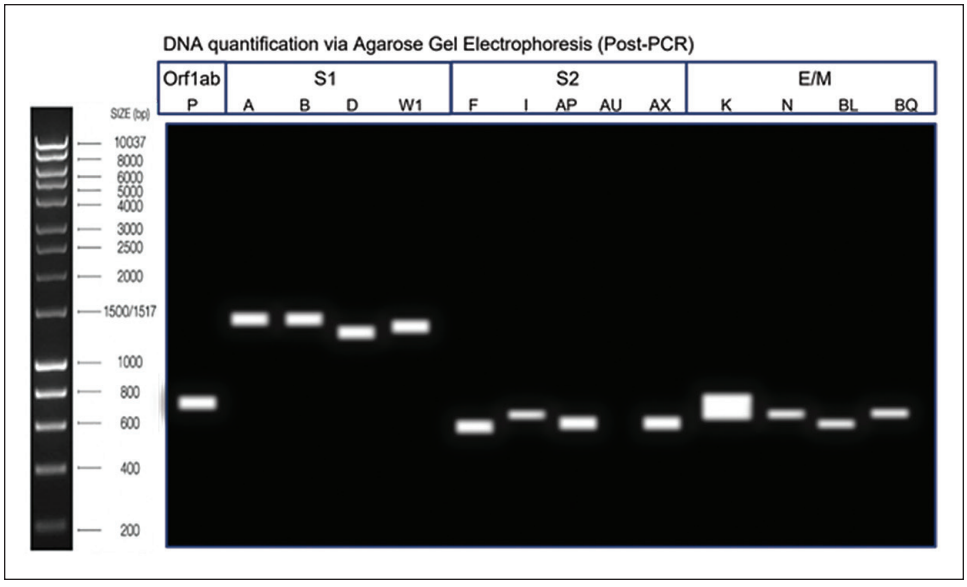

Figure 2. DNA confirmation via Agarose Gel Electrophoresis (AGE) (Post PCR).

present in sample F, I, AP, AU and AX. Finally, samples K, N, BL, and BQ were amplified using the EM primer, which specifically targets the E and M genes present in the samples. Furthermore, differences in the migration of samples using the same primer could be attributed to the differences in the concentration of each sample since concentration loading was not standardized in each sample to qualitatively evaluate whether the primers will be able to amplify samples regardless of its concentration. These results suggested that the designed primers were able to effectively amplify conserved regions of a certain target gene present in the RNA isolated regardless of its concentration and purity.

Genomic Sequence Surveillance Evaluation

Validation of the developed genomic surveillance method for the identification of the frequency and dominance of specific SARS-CoV-2 variant was carried out through the application of the four in silico-designed oligonucleotide primers namely: O1, S1, S2, and EM, where each primer corresponds to a certain target gene significant in the viral life cycle.

ORF1ab gene was used to evaluate the ability of designed primer (O1) to interact with the viral replication process of SARS-CoV-2. On the other hand, three of the in silico-designed oligonucleotide primers targeting various structural proteins' genes such as: spike (S) protein, envelope (E), and membrane (M) protein were utilized to determine its potential therapeutic activity by the application of the developed genomic surveillance method through touchdown polymerase chain reaction (PCR).

Results revealed that three of the in silico-designed oligonucleotide primers (S1, S2, and EM) successfully amplified sequence of interest leading to the classification of the SARS-CoV-2 variant dominating in the samples as Omicron (Table 5). From the results, 10 out of the 14 samples were further classified as Omicron variant. Among its subvariants, BA and its sub-lineage was the most dominant among the samples. This Omicron subvariant has been known to escalate and dominate the variant distribution in early 2023, based on the latest COVID-19 bio-surveillance report of the Department of Health (DOH) of the Republic of the Philippines.[16] Among the samples, P_O, A_S1, F_S2, and AP_S2 exhibited no significant similarity that can be used for its lineage identification. These results may be attributed to the low-quality profile obtained from the sequencing results of each sample as can be seen in the Supplementary Materials. On the other hand, various lineages were also obtained from the samples (e.g., XB, XBC, FL, and FY), where majority of which were the dominating Omicron subvariants during second and third quarter of 2023.[16]

Notably, most of the samples were able to have a significant blast identity score (NLT 83%), and an E value of NMT 1.0^{-55}, which is considered as a good hit for homology match. Since for nucleic acid-based search, the suggested threshold for the E value is ≤1e-6, and a sequence % identity of ≥70%,[17] proving the selectivity of each primer towards conserved sequences of various lineages that are further classified as Omicron variants.

Another distinguished finding was also observed in sample AU_S2. As presented in Table 5, AU_S2 sample matched with a certain sequence lineage (B.1.448), with an

Table 5. SARS-CoV-2 Positive Specimens

Sample Code	Sequence Length	COVID-19 Lineage	Variant	E Value	Blast Score (% Identity)
P_O[a]	N/A	N/A	N/A	N/A	N/A
A_S1[a]	N/A	N/A	N/A	N/A	N/A
B_S1[b]	270	BA.2.68[F]	Omicron	6.79575e-121	96.2%
D_S1[b]	1,248	BA.2.9[F]	Omicron	0	94.0%
W_S1[b]	1,179	BA.2.9[F]	Omicron	0	95.0%
F_S2[a]	N/A	N/A	N/A	N/A	N/A
I_S2	380	XBC.1.3[F]	Omicron	0	99.7%
	452	BA.2.3.1[R]	Omicron	2.40983e-162	99.7%
AP_S2[a]	N/A	N/A	N/A	N/A	N/A
AU_S2[a]	598	B.1.448	Unidentified	4.84745e-56	83.0%
AX_S2[b]	472	XBB.1.19[F]	Omicron	0	99.8%
K_EM[b]	455	CH.1.1.22[F]	Omicron	0	99.1%
N_EM[b]	401	FL.1.5.1[F]	Omicron	0	99.8%
BL_EM[b]	350	FY.3.1[F]	Omicron	8.34825e-176	98.9%
BQ_EM[b]	899	XBB.1.5.80[F]	Omicron	0	99.8%

[a] *No significant similarity found in both forward and reverse primers.*
[b] *No significant similarity found in reverse primer.*

F - forward, R - reverse

Table 6. Characteristic Mutation Profile of B.1.448

Gene	Amino Acid
ORF1a	T265I
ORF1b	P314L
S	D614G
ORF3a	Q57H
M	A2V
ORF8	S84L
N	S183Y

83% blast score, and based on lineage report[18], it has been identified that the B.1.448 lineage did not correspond to any of the five existing VOC (e.g., Alpha Beta, Delta, Gamma, Omicron). From this notable finding, it can be a plausible implication that the designed in silico oligonucleotide targeting S2 gene of the S-protein, was able to target specific mutated sequences that were already identified as a characteristic of the mutation profile of B.1.448 lineage (Table 6).[19]

As a lineage of interest, B.1.448 was further assessed to determine distinct mutations characterized in the lineage and in other VOC (Figures 3.1 and 3.2). It can be depicted from B and C of Figure 3.1, mutations P314L and D614G were both present in B.1.448, and in other VOC except Omicron.

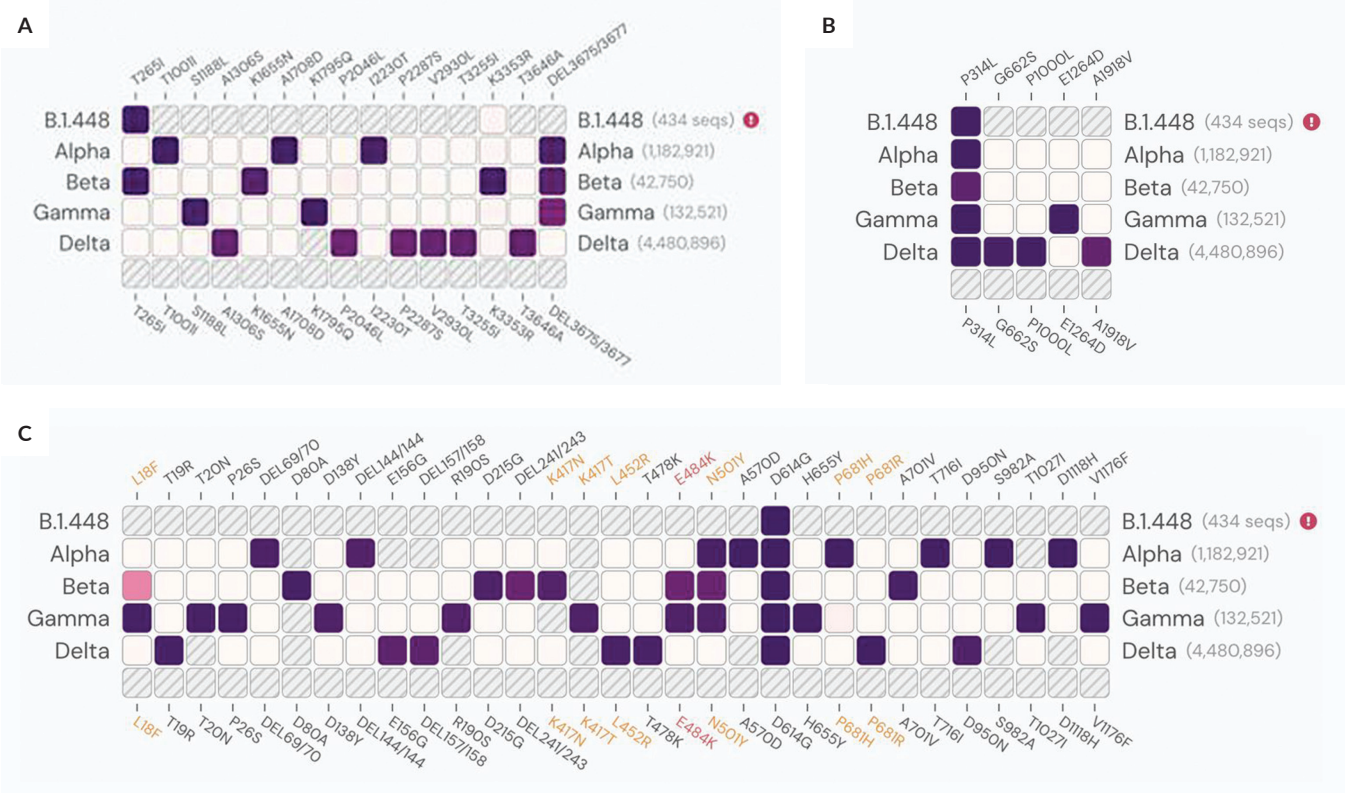

Figure 3.1. Mutation profile alignment **(A)** ORF1a gene, **(B)** ORF1b gene, and **(C)** S-Protein gene.

**Every last row corresponds to Omicron variant.*

(!) – Currently recorded 434 sequences.

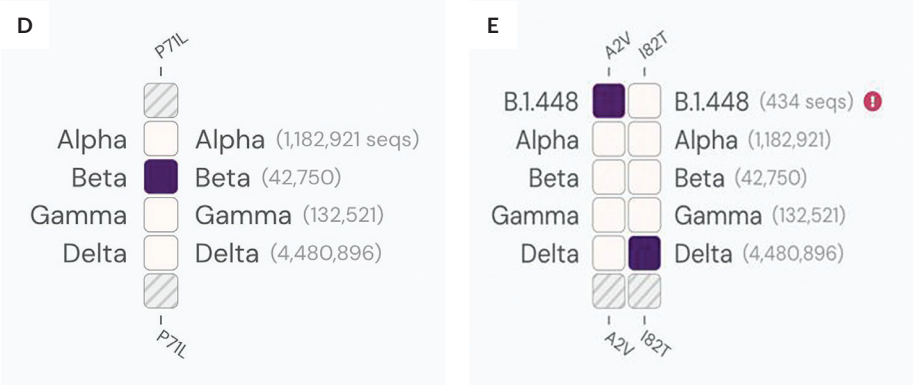

Figure 3.2. Mutation profile alignment **(D)** E-Protein gene and **(E)** M-Protein gene.

**Every last row corresponds to Omicron variant.*

(!) – Currently recorded 434 sequences.

Mutational Profile Characterizing Omicron Variant

Among other variants, it is noteworthy that the Omicron variant has a peculiar mutation profile characterized by high number of mutations, including 16 deletion and 3 insertions, in contrast to other VOC[20] where most of these mutations are found in the spike protein's receptor-binding domain (RBD). These mutations of concern play a vital role in viral infection and have a potential impact in its transmissibility and immune evasion capabilities of the virus.[20,21] Thus, most of developed COVID-19 vaccines target key mutations found at the S-protein of the virus. Another key characterization of Omicron variant among other VOC, it is characterized by 39 mutations in its S-protein, where 15 of these are located within the RBD, while the remaining mutations localize at the S1/S2 junction and the S2 domain[20,21] as shown in Figure 4.

In the present study, it can be depicted in Figures 5.1 to 5.3 the characteristic mutation profile of each lineage identified per sample. Evident from Figure 5.2 (C1 to C3),

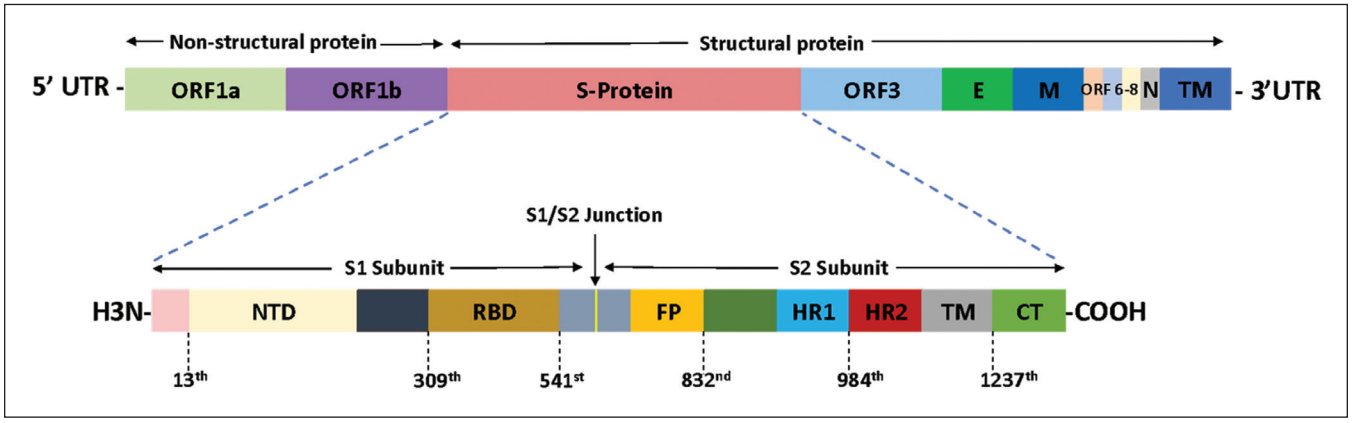

Figure 4. Representation of SARS-CoV-2 genome.

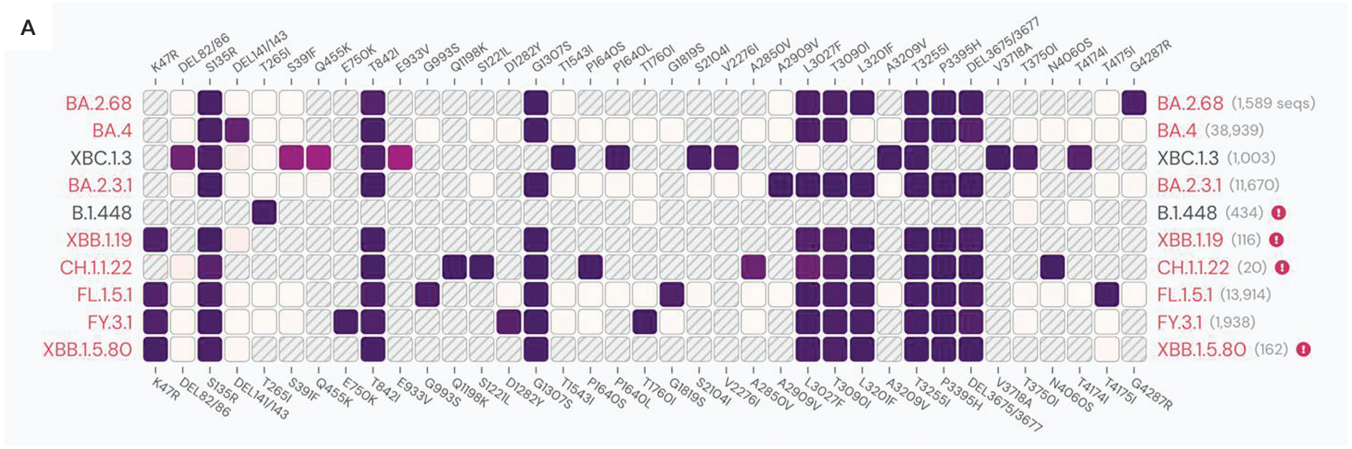

Figure 5.1. Mutation prevalence across lineage **(A)** ORF1a gene and **(B)** ORF1b gene.

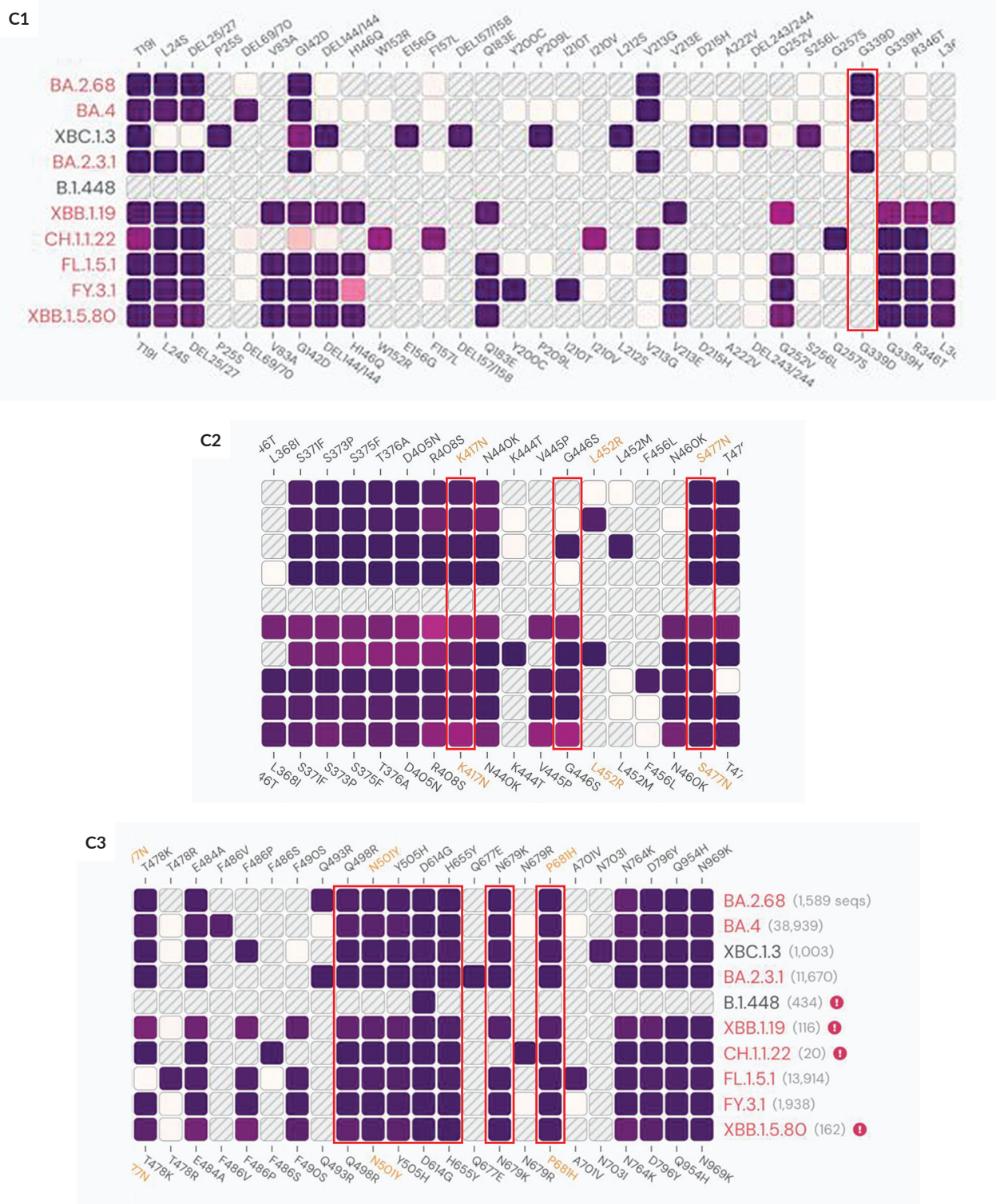

Figure 5.2. Mutation prevalence across lineage **(C1 to C3)** S-Protein gene.

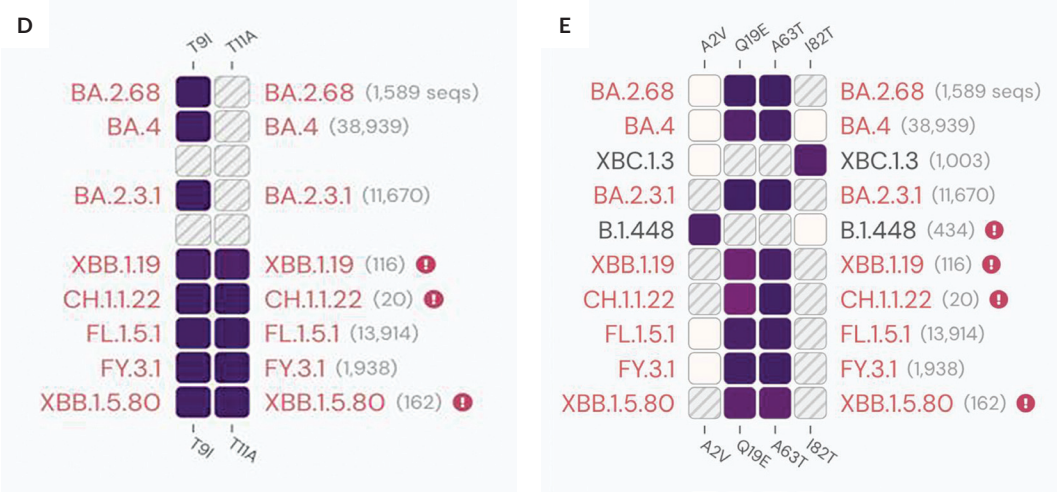

Figure 5.3. Mutation prevalence across lineage **(D)** E-Protein gene and **(E)** M-Protein gene.

Table 7. Distinct Mutations Observed in Identified Lineages at a Specific Target Gene as a Characteristic of Omicron Variant

COVID-19 Lineage	ORF1ab		S Protein												E	M	
			S1 Subunit								S2 Subunit						
	S135R	T842I	G339D	N440K	G446S	S477N	E484A	Q498R	Y505H	N679K	N764K	D796Y	Q954H	N969K	T9I	Q19E	A63T
BA.2.68	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BA.4	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
XBC.1.3	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
BA.2.3.1	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
XBB.1.19	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CH.1.1.22	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
FL.1.5.1	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
FY.3.1	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
XBB.1.5.80	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

** Unavailable data were left blank.*

lineages BA.2.68, BA.2.9, BA.2.3.1, CH.1.1.22, XBC.1.3, XBB.1.19, FL.1.5.1, FY.3.1, and XBB.1.5.80, exhibited mutations at H655Y, N679K, P681H, and Q498R, which is a distinct characteristic that is associated to Omicron variant.[20] Mutation in G444S was also observed in the following lineages: CH.1.1.22, XBC.1.3, XBB.1.19, FL.1.5.1, FY.3.1., while lineages BA.2.68, BA.2.9, and BA.2.3.1, displayed mutation at G339D. Among all mutations, D614G was observed in all lineages, while nine out of 10 determined lineages displayed the presences of various mutations of interest, including K417N, S477N, and P681H. The said mutations were considered as mutation of interest as it has been also identified to be present in other VOCs.[18,21] Thus, concluding that the identified lineages per sample were categorized as Omicron variant. Summarized in Table 7 the distinct amino acid mutations observed in the identified lineage at a specific target gene. The reported mutations in the table were known to be distinct key mutations for Omicron variant.[20] Particularly S477N, E484A, and Q498R were key mutations commonly observed at the receptor binding domain (RDB) of the S1 subunit of the spiked protein of the virus.[20,21]

DISCUSSION

Polyproteins as Target Genes

SARS-CoV-2_Var_O1

SARS-CoV-2 as the causative agent of COVID-19, is a positive-sense single stranded RNA virus [(+) ssRNA virus], with a genome like that of other *Coronaviridae* consisting of one large open-reading frame (ORF) encoding two overlapping polyproteins (ORF1a and ORF1ab). ORF1ab gene encodes two major polyproteins (pp) i.e., pp1a and pp1ab of the replicase polyprotein, in which replicase polyprotein (RPP) plays a vital role in SARS-CoV-2 pathogenesis. RPP is composed of three domains: the macro domain, papain-like protease (PLPRO) and the main protease (Mpro). Mpro

serves as a key component of the ORF1ab gene which is responsible for the processing of polyprotein necessary for virus assembly during viral replication.[22] The 2020 study of Naqvi et al., also suggested that targeting the ORF1ab gene can potentially disrupt the replication and transcription processes of the virus, thus inhibiting its ability to replicate and spread.[23] This is also supported by the study of Thomas in 2020, reiterating that targeting the ORF1ab gene may also aid in preventing the development of drug resistance, as ORF1ab gene is a conserved region of the whole SARS-CoV-2 viral genome.[24] Thus, the designing of oligonucleotide primer targeting the ORF1ab gene serves as an initial step in the development of a more specific and selective therapeutics against SARS-CoV-2.

Structural Proteins as Target Genes

SARS-CoV-2_Var_S1 & S2

SARS-CoV-2 has a highly glycosylated spike (S) protein that primarily belongs to the trimeric class I viral fusion glycoprotein.[22] This structural protein found on the surface of the SARS-CoV-2 virus is responsible for facilitating viral entry into host cell, by its initial attachment to the host cell membrane by interacting with the ACE2 receptor.[23] S protein is composed of three subunits: S1, S2, and S2'. Both studies of Shamsi et al., and Naqvi et al., suggests that the S1 and S2 domains of S protein are responsible for receptor binding and membrane fusion.[22,23] Where the process of viral infection is initiated through the binding and attachment of the virus to the human ACE2 receptor with the S1 domain of S protein.[22,23]

On the other hand, S2 domain functions as the fusion protein, which facilitates the fusion between the virions and the host cell membrane.[22,23] Fusion process causes the S2 domain to undergo three different conformational changes namely: pre-fusion native state, hairpin intermediate state, and post fusion hairpin state.[22,23] This alteration in the S protein conformation attributed to the S2 domain fusion activity can be directly linked to various variations in the mechanism of pathogenesis among different COVID viruses such as SARS-CoV-2.

SARS-CoV-2_Var_EM

Aside from Spike protein, SARS-CoV-2 contains two important structural proteins, the envelope (E) protein, and the membrane (M) protein. Both the E and M proteins have vital roles in viral assembly and structure.[23] Study of Naqvi et al., in 2020 elaborated the role of E protein as a viroporin, which forms protein-lipid pores involved in ion transport.[23] This ion channel activity of E protein promotes virus fitness and pathogenesis.[18] Thus, making E protein a potential therapeutic target can further inhibits viral assembly and reduce viral infectivity.[22,23]

The M protein is the most abundant envelope glycoprotein of SARS-CoV-2.[24] Although, various studies haven't completely understood the role of M protein in viral lifecycle, the study of Thomas in 2020, hypothesized that M protein plays a crucial role in stabilizing other structural protein including, nucleocapsid (N) protein, and promoting the completion of viral assembly by stabilizing the N protein-RNA complex inside the virion as part of the RNA packaging process.[24] Inhibition of M protein function makes it a potential therapeutic target, as it could also prevent viral replication and assembly.[23]

Omicron Variant Prevalence

As the Omicron variant continuously evolves, different and new subvariants are being discovered. The 2021 study of Shamsi et al., mentioned that the receptor-binding domain (RBD) within the S protein is the most unpredictable feature of SARS-CoV-2 with maximum variation in the receptor-binding motif.[22] This linked the findings of the study of Zhang et al. in 2021, which concluded that the S protein of the Omicron variant undergone multiple mutations; this includes 12 mutations in the RDB region, where half of these mutations are located around the N501 position at the C-terminus.[8]

Various studies also suggest that the rampant increase of COVID-19 cases in the latter days of 2021 is attributed to the emergence of the new SARS-CoV-2 variant namely, Omicron.[8,25] The uncontrolled viral evolution of the Omicron variant is attributed to its highly mutated nature. These several mutations have been found to be associated with its highly mutated S protein.[25] The large number of genetic changes in Omicron variant compared to previous variants, made Omicron a variant of concern (VOC) in late 2021.[8]

Moreover, the 2022 study of Poudel et al., emphasized that the high number of mutations in the structural protein, specifically S protein of Omicron variant makes it a potential target of neutralizing antibodies, as it's also plays a crucial role in viral entry into the host cells.[26] This claim is also supported by the study of VanBlargan et al. in 2022, which highlighted the importance of targeting the highly conserved residues of S protein of the SARS-CoV-2 to prevent resistance to the future variants with highly mutated spike sequences.[27]

Two studies of Chakraborty et al. in 2022 identified some of the key mutations that can be observed in the S-protein of Omicron variant, including S371L, D614G, E484A, N501Y, Q493K, K417N, S477N, Y505H, and G496S.[28,29] Among the said mutations, in the present study it has been justified and validated the existence of various mutations in the identified lineages. Aside from these mutations located at the structural protein region of the genome, Omicron variant also poses large number of mutations localized in its non-structural regions, including ORF1a and ORF1b.[28] Additionally, some of the key mutations observed in the RBD region are G339D, G446S, and Q498R mutations which results in the increase of molecular flexibility of S-protein.[20] On the other hand, common mutations including D614G, H655Y, N679K, P681H[20] are detected in the S1/S2 junctions

in which these mutations are commonly observed in other variants of COVID-19.

B.1.448 lineage

Unlike other identified lineages from our sample, B.1.448 lineage is found to be unidentified in terms of its Covid-19 variant classification. This lineage has been reported to be found at the Montefiore Health Systems (MHS) in the United States during the early wave of the pandemic.[30] As reported in the 2021 study of Fels et al., the B.1.448 represented the larger sub-lineage of B.1, which is also found to be the parent lineage of the identified variants of concern such as Omicron and Alpha.[30] This lineage first arises at Bronx, New York City, which was tagged as the early epicenter of COVID-19 pandemic in the United States of America (USA). Despite the efforts of the team of Fels in categorizing B.1.448 and their findings about the characteristics of its single nucleotide polymorphism (SNPs), which has 332 other genomes sharing lineage[30], B.1.448 in present time is still not categorized under specific known variant of COVID-19. And these may be due to the low number of sequences submitted that matched the said lineage.

Various studies focused on the descendants of B lineage since these were known to be the most predominant among SARS-CoV-2 virus. From the 2024 study of Tiwary, the genetic diversity of the B lineage and its sublineages, including B.1.448, show an increasing trend in genetic diversity overtime, suggesting that the B lineage is evolving rapidly. In which key feature of B.1 lineage is at the positive selection observed at the binding site 501, which leads to the N501Y mutation[31], which is a mutation of concern located at the S-protein of the genome. Despite of not exhibiting N501Y mutation, B.1.448 lineage still exhibit mutations at the S-protein and other regions of the viral genome that significantly influence its transmissibility, since S-protein of the viral genome is primarily involved in receptor recognition, viral attachment, and entry of the virus to the host cells.

CONCLUSION

From the results obtained, it has been identified that among the five VOCs, Omicron is the most prevalent and dominant variant within the sample population where various determined lineages per sample displayed the existence of key mutations distinct for the Omicron variant. These characteristic mutation profiles were justified to be observed in Omicron variant as these contribute mainly to the extremely high mutation nature of Omicron which can be associated and located primarily in its S-protein, particularly in the RBD region, and S1/S2 junction. Thus, the classification of different lineages of the said variant proves its highly mutated genome, as well as the feasibility of the application of the novel genomic surveillance method and application of the in silico-designed oligonucleotide primers in targeting conserved sequences in the viral genome. The ability of the primers to identify known and unknown variants was also proven provided by the identification of the B.1.448 lineage, which was broadly detected and reported at Bronx, New York City, the early epicenter of COVID-19 in the USA. The said lineage is known to be unidentified in terms of its variant classification due to its broadly shared genome like that to other lineages. Overall, the developed DPPSS method successfully amplified the genomic sequence of interest and effectively provided valuable information about the frequency and dominance of specific SARS-CoV-2 variants among samples. Further evaluation and sequence alteration of the oligonucleotide primers may offer significant contribution to the development of a specific and selective direct PCR-based detection of new emerging COVID-19 variants. It is also recommended that a validation study utilizing the DPPSS method using a larger sample size be conducted. Additional potential further study that can be made in our preliminary study, is the comparison of the developed method with an existing SARS-CoV-2 variant detection method. With this, robustness of the designed primers may be verified and validated altogether with the application of a larger sample size and additional in silico analyses comparison of the designed primers.

Data Availability Statement

Supplementary materials are available from the corresponding author upon reasonable request.

Statement of Authorship

All authors certified fulfillment of ICMJE authorship criteria.

Author Disclosure

All authors declared no conflicts of interest.

Funding Source

None.

REFERENCES

1. Liu YC, Kuo RL, Shih SR. COVID-19: The first documented coronavirus pandemic in history. Biomed J. 2020 Aug;43(4):328-333. doi: 10.1016/j.bj.2020.04.007. Epub 2020 May 5. PMID: 32387617; PMCID: PMC7199674.
2. World Health Organization, Coronavirus disease (COVID-19) [Internet]. 2023 [cited 2024 Jan]. Available from: https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19)
3. Philippine News Agency, Covid-19 out of 'top causes of death in PH' list: DOH [Internet]. 2023 [cited 2024 Jan]. Available from: https://www.pna.gov.ph/index.php/articles/1193578
4. Wise J. Covid-19: WHO declares end of global health emergency. BMJ. 2023 May 9; 381:1041. doi: 10.1136/bmj.p1041. PMID: 37160309.
5. Mohseni Afshar Z, Tavakoli Pirzaman A, Karim B, Rahimipour Anaraki S, Hosseinzadeh R, Sanjari Pireivatlou E, et al. SARS-CoV-2 Omicron (B.1.1.529) Variant: A Challenge with COVID-19. Diagnostics (Basel). 2023 Feb 2;13(3):559. doi: 10.3390/diagnostics13030559. PMID: 36766664; PMCID: PMC9913917.
6. Walker AS, Vihta KD, Gethings O, Pritchard E, Jones J, House T, et al. Tracking the Emergence of SARS-CoV-2 Alpha Variant in the

United Kingdom. N Engl J Med. 2021 Dec 30;385(27):2582-2585. doi: 10.1056/NEJMc2103227. Epub 2021 Dec 8. PMID: 34879193; PMCID: PMC8693687.
7. Mohseni Afshar Z, Babazadeh A, Janbakhsh A, Mansouri F, Sio TT, Sullman MJM. Coronavirus disease 2019 (Covid-19) vaccination recommendations in special populations and patients with existing comorbidities. Rev Med Virol. 2022 May;32(3): e2309. doi: 10.1002/rmv.2309. Epub 2021 Oct 22. PMID: 34677889; PMCID: PMC8646697.
8. Zhang X, Wu S, Wu B, Yang Q, Chen A, Li Y. SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance. Signal Transduct Target Ther. 2021 Dec 17;6(1):430. doi: 10.1038/s41392-021-00852-5. PMID: 34921135; PMCID: PMC8678971.
9. Astuti I, Ysrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes Metab Syndr. 2020 Jul-Aug;14(4):407-412. doi: 10.1016/j.dsx.2020.04.020. Epub 2020 Apr 18. PMID: 32335367; PMCID: PMC7165108.
10. Fernandes Q, Inchakalody VP, Merhi M, Mestiri S, Taib N, Moustafa Abo El-Ella D. Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics, and vaccines. Ann Med. 2022 Dec;54(1):524-540. doi: 10.1080/07853890.2022.2031274. PMID: 35132910; PMCID: PMC8843115.
11. Yüce M, Filiztekin E, Özkaya KG. COVID-19 diagnosis -A review of current methods. Biosens Bioelectron. 2021 Jan 15;172:112752. doi: 10.1016/j.bios.2020.112752. Epub 2020 Oct 24. PMID: 33126180; PMCID: PMC7584564.
12. Nayar G, Seabolt EE, Kunitomi M, Agarwal A, Beck KL, Mukherjee V, et al. Analysis and forecasting of global real time RT-PCR primers and probes for SARS-CoV-2. Sci Rep. 2021 Apr 26;11(1):8988. doi: 10.1038/s41598-021-88532-w. PMID: 33903676; PMCID: PMC8076216.
13. Gand M, Vanneste K, Thomas I, Van Gucht S, Capron A, Herman P, et al. Deepening of In Silico Evaluation of SARS-CoV-2 Detection RT-qPCR Assays in the Context of New Variants. Genes (Basel). 2021 Apr 13;12(4):565. doi: 10.3390/genes12040565. PMID: 33924636; PMCID: PMC8069896.
14. Böger B, Fachi MM, Vilhena RO, Cobre AF, Tonin FS, Pontarolo R. Systematic review with meta-analysis of the accuracy of diagnostic tests for COVID-19. Am J Infect Control. 2021 Jan;49(1):21-29. doi: 10.1016/j.ajic.2020.07.011. Epub 2020 Jul 10. PMID: 32659413; PMCID: PMC7350782.
15. Obradovic J, Jurisic V, Tosic N, Mrdjanovic J, Perin B, Pavlovic S, Djordjevic N. Optimization of PCR conditions for amplification of GC-Rich EGFR promoter sequence. J Clin Lab Anal. 2013 Nov;27(6):487-93. doi: 10.1002/jcla.21632. PMID: 24218132; PMCID: PMC6807403.
16. LATEST COVID-19 BIOSURVEILLANCE REPORT | Department of Health [Internet]. [cited 2024 Jan]. Available from: https://doh.gov.ph/covid19-variants
17. Choudhuri S. Sequence alignment and similarity searching in genomic databases. Elsevier eBooks [Internet]. 2014 [cited 2024 Dec]. p. 133–155. Available from: https://www.sciencedirect.com/science/article/pii/B9780124104716000062
18. Cov-Lineages.org Lineage Report [Internet]. [cited 2024 January]. Available from: https://cov-lineages.org/index.html
19. Gangavarapu K, Latif AA, Mullen JL, Alkuzweny M, Hufbauer E, Tsueng G, et al. Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. Nat Methods. 2023 Apr;20(4):512-522. doi: 10.1038/s41592-023-01769-3. Epub 2023 Feb 23. PMID: 36823332; PMCID: PMC10399614.
20. Alkhatib M, Salpini R, Carioti L, Ambrosio FA, D'Anna S, Duca L, et al. Update on SARS-CoV-2 Omicron Variant of Concern and Its Peculiar Mutational Profile. Microbiol Spectr. 2022 Apr 27;10(2):e0273221. doi: 10.1128/spectrum.02732-21. Epub 2022 Mar 30. PMID: 35352942; PMCID: PMC9045195.
21. Chakraborty C, Bhattacharya M, Sharma AR, Dhama K, Agoramoorthy G. A comprehensive analysis of the mutational landscape of the newly emerging Omicron (B.1.1.529) variant and comparison of mutations with VOCs and VOIs. Geroscience. 2022 Oct;44(5):2393-2425. doi: 10.1007/s11357-022-00631-2. Epub 2022 Aug 22. PMID: 35989365; PMCID: PMC9393103.
22. Shamsi A, Mohammad T, Anwar S, Amani S, Khan MS, Husain FM, et al. Potential drug targets of SARS-CoV-2: From genomics to therapeutics. Int J Biol Macromol. 2021 Apr 30;177:1-9. doi: 10.1016/j.ijbiomac.2021.02.071. Epub 2021 Feb 9. PMID: 33577820; PMCID: PMC7871800.
23. Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, et al. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis, and therapies: Structural genomics approach. Biochim Biophys Acta Mol Basis Dis. 2020 Oct 1;1866(10):165878. doi: 10.1016/j.bbadis.2020.165878. Epub 2020 Jun 13. PMID: 32544429; PMCID: PMC7293463.
24. Thomas S. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. Pathog Immun. 2020 Oct 19;5(1):342-363. doi: 10.20411/pai.v5i1.377. PMID: 33154981; PMCID: PMC7608487.
25. Syed AM, Ciling A, Khalid MM, Sreekumar B, Chen PY, Kumar GR, et al. Omicron mutations enhance infectivity and reduce antibody neutralization of SARS-CoV-2 virus-like particles. medRxiv [Preprint]. 2022 Jan 2:2021.12.20.21268048. doi: 10.1101/2021.12.20.21268048. Update in: Proc Natl Acad Sci U S A. 2022 Aug 2;119(31): e2200592119. PMID: 34981067; PMCID: PMC8722610.
26. Poudel S, Ishak A, Perez-Fernandez J, Garcia E, León-Figueroa DA, Romaní L, et al. Highly mutated SARS-CoV-2 Omicron variant sparks significant concern among global experts - What is known so far? Travel Med Infect Dis. 2022 Jan-Feb;45:102234. doi: 10.1016/j.tmaid.2021.102234. Epub 2021 Dec 8. PMID: 34896326; PMCID: PMC8666662.
27. VanBlargan LA, Errico JM, Halfmann PJ, Zost SJ, Crowe JE Jr, Purcell LA, et al. An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. Nat Med. 2022 Mar;28(3):490-495. doi: 10.1038/s41591-021-01678-y. Epub 2022 Jan 19. PMID: 35046573; PMCID: PMC8767531.
28. Chakraborty C, Bhattacharya M, Sharma AR, Dhama K, Agoramoorthy G. A comprehensive analysis of the mutational landscape of the newly emerging Omicron (B.1.1.529) variant and comparison of mutations with VOCs and VOIs. Geroscience. 2022 Oct;44(5):2393-2425. doi: 10.1007/s11357-022-00631-2. Epub 2022 Aug 22. PMID: 35989365; PMCID: PMC9393103.
29. Chakraborty C, Bhattacharya M, Sharma AR, Mallik B. Omicron (B.1.1.529) - A new heavily mutated variant: Mapped location and probable properties of its mutations with an emphasis on S-glycoprotein. Int J Biol Macromol. 2022 Oct 31; 219:980-997. doi: 10.1016/j.ijbiomac.2022.07.254. Epub 2022 Aug 8. PMID: 35952818; PMCID: PMC9359758.
30. Fels JM, Khan S, Forster R, Skalina KA, Sirichand S, Fox AS, et al. Genomic surveillance of SARS-CoV-2 during the first year of the pandemic in the Bronx enabled clinical and epidemiological inference. medRxiv [Preprint]. 2022 Mar 17:2021.02.08.21250641. doi: 10.1101/2021.02.08.21250641. Update in: Cold Spring Harb Mol Case Stud. 2022 Jul 13:mcs.a006211. doi: 10.1101/mcs.a006211. PMID: 33594384; PMCID: PMC7885943.
31. Tiwary BK. A positive selection at binding site 501 in the B.1 lineage might have triggered the highly infectious sub-lineages of SARS-CoV-2. Gene. 2024 Jul 15;915:148427. doi: 10.1016/j.gene.2024.148427. Epub 2024 Apr 2. PMID: 38575097.