

Article

Not peer-reviewed version

Genetic Variations of Three Kazakhstan Strains of the SARS-COV-2 Virus

[Bekbolat Ussebayev](#)^{*}, Kulyaisan Sultankulova, Yerbol Burashev, Aibarys Melisbek, Meiirzhan Shirinbekov, [Balzhan Myrzakhmetova](#), [Izat Smekenov](#), Asankadir Zhunushov, [Aslan Kerimbaev](#), Sergazy Nurabaev, [Olga Chervyakova](#), Nurlan Kozhabergenov, Lesbek Kutumbetov

Posted Date: 18 February 2025

doi: 10.20944/preprints202502.1398.v1

Keywords: SARS-CoV-2; whole genome sequencing; primers; mutation; phylogenetic analysis; COVID-19



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Genetic Variations of Three Kazakhstan Strains of the SARS-CoV-2 Virus

Bekbolat S. Ussebayev ^{1,2,*}, Kulyaisan T. Sultankulova ¹, Yerbol D. Burashev ^{1,3}, Aibarys M. Melisbek ^{1,3}, Meirzhan Z. Shirinbekov ¹, Balzhan S. Myrzakhmetova ¹, Izat Smekenov ¹, Asankadir Zhunushov ², Aslan A. Kerimbaev ¹, Sergazy S. Nurabaev ¹, Olga V. Chervyakova ¹, Nurlan S. Kozhabergenov ¹ and Lesbek B. Kutumbetov ^{1,*}

¹ Al-Farabi Kazakh National University, Almaty, Kazakhstan

² Research Institute for Biological Safety Problems, «National holding QazBioPharm», Gvardeysky 080409, Kazakhstan

³ Institute of Biotechnology of the National Academy of Sciences of the Kyrgyz Republic, Bishkek, Kyrgyzstan

* Correspondence: b.ussebayev@biosafety.kz

Abstract: Prompt determination of the etiological agent is important in an outbreak of pathogens with pandemic potential particularly dangerous infectious diseases. Molecular genetic methods allow for arriving at an accurate diagnosis, employing timely preventive measures and controlling the spread of the disease causing agent. In this study, whole-genome sequencing of three SARS-CoV-2 strains was performed using the Sanger method, which provides high accuracy in determining nucleotide sequences and avoids errors associated with multiple DNA amplification. Complete nucleotide sequences of samples KAZ/Britain/2021, KAZ/B.1.1/2021 and KAZ/Delta020/2021 were obtained with sizes of 29,751 bp, 29,815 bp and 29,840 bp, respectively. According to the COVID-19 Genome Annotator, 127 mutations were detected in the studied samples compared to the reference strain. The strain KAZ/Britain/2021 contained 3 deletions, 7 synonymous mutations and 27 non-synonymous mutations, the second strain KAZ/B.1.1/2021 contained 1 deletion, 5 synonymous mutations and 31 non-synonymous mutations and the third strain KAZ/Delta020/2021 contained 1 deletion, 5 synonymous mutations and 37 non-synonymous mutations, respectively. Variations C241T, F106F, P314L and D614G found in the 5' UTR, ORF1ab and S region were common to all three studied samples, respectively. According to PROVEAN data, the loss-of-function mutations identified in strains KAZ/Britain/2021, KAZ/B.1.1/2021 and KAZ/Delta020/2021 include 5 mutations (P218L, T716I, W149L, R52I and Y73C), 2 mutations (S813I and Q992H) and 8 mutations (P77L, L452R, I82T, P45L, V82A, F120L, F120L and R203M), respectively. Phylogenetic analysis showed that the strains studied (KAZ/Britain/2021, KAZ/B.1.1/2021 and KAZ/Delta020/2021) belong to different SARS-CoV-2 lineages, which are closely related to samples from Germany (OU141323.1 and OU365922.1), Mexico (OK432605.1) and again, Germany (OV375251.1 and OU375174.1), respectively. The nucleotide sequences of the studied SARS-CoV-2 virus strains were registered in the Genbank database with the accession numbers: ON692539.1, OP684305, and OQ561548.1.

Keywords: SARS-CoV-2; whole genome sequencing; primers; mutation; phylogenetic analysis; COVID-19

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), was first reported in December 2019 in Wuhan, Hubei Province, China [1]. According to WHO COVID-19 data, as of January 12, 2025, a total of 777,315,739 confirmed cases have been reported, of which 7,083,869 have resulted in death (data.who.int). The first cases of COVID-19 in the Republic of Kazakhstan were reported on March 13, 2020 [2,3].

The SARS-CoV-2 virion is spherical or ellipsoidal, with an average diameter ranging from 60 to 140 nanometers [4]. The SARS-CoV-2 virus genome consists of ~29.9 kb and is organized in the following order from 5' → 3': open reading frame (ORF) 1ab (replicase), structural spike glycoprotein (S), ORF3a protein, structural envelope protein (E), structural membrane glycoprotein (M), ORF6 protein, ORF7a protein, ORF7b protein, ORF8 protein, structural nucleocapsid phosphoprotein (N), and ORF10 protein [5].

Over time, all viruses, including SARS-CoV-2, undergo molecular genetic changes. Most of these changes have little effect on the properties of the virus. However, some mutations can affect various aspects, such as its infectivity, transmissibility, the effectiveness of treatment and vaccines, as well as virulence [6]. In addition, since its emergence in 2019, the SARS-CoV-2 virus has undergone continuous changes, which contributed to the emergence of multiple lineages and variants (Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Omicron (B.1.1.529), etc. [7,8], having differences in transmission characteristics, ability to cause severe disease, and the ability to evade the immune response [9].

The expansion of the complete genomic sequences of the SARS-CoV-2 virus in the information databases (GISAID and GenBank NCBI) was made possible by rapid genome sequencing using Sanger or NGS methods [10]. These SARS-CoV-2 genomes in the context of the COVID-19 pandemic can provide invaluable information on the evolution of the virus and allow tracking of the geographic distribution of individual mutations, as well as monitoring of the spread of the virus in the human population [11]. In addition, the evolution of the virus is facilitated by the adaptation of the virus in different conditions and results from a balance between its genetic information and genome variability [12]. Studying the evolution and genetic changes in the genomes of various variants of SARS-CoV-2 is extremely important in developing clinical and political strategies within geographical regions [13], as well as for the creation of diagnostic tests and vaccines against this virus [14].

The aim of this work is to sequence the complete genome of three isolates of the SARS-CoV-2 virus, determine their genetic variations and identify various types of mutations present in the different strains.

2. Materials and Methods

2.1. Sample Collection

Three strains of the SARS-CoV-2 virus were received for molecular genetic studies at the Research Institute for Biological Safety Problems in 2022 from the Scientific and Practical Center for Sanitary and Epidemiological Expertise and Monitoring branch of the National Center for Public Health, a republican state enterprise on the right of economic use of the Ministry of Health of the Republic of Kazakhstan.

2.2. RNA Extraction

Total RNA was extracted from virus-containing fluid using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. RNA concentrations were estimated using Qubit RNA HS assay kits (Life Technologies, Carlsbad, California, USA) on a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, California, USA) according to the manufacturer's protocol.

2.3. cDNA Synthesis

Reverse transcription (RT) was performed using the SuperScript VILO cDNA Synthesis Kit (Invitrogen, Thermo Fisher Scientific, Carlsbad, California, USA) in a Mastercycler X50s thermal cycler at the following conditions: 25°C for 10 min; 42°C for 60 min; 85°C for 5 min. The reaction composition and temperature-time conditions were followed according to the manufacturer's instructions.

2.4. Primer Design and Synthesis

Specific overlapping primers for amplification and sequencing of all SARS-CoV-2 virus genes were manually searched and designed on the NCBI website using the GenBank database. The nucleotide sequence of the sequencing primers was designed based on the SARS-CoV-2 isolate Wuhan-Hu-1 reference strain ([NC_045512.2](https://www.ncbi.nlm.nih.gov/nuclref/NC_045512.2)). The specificity of the primers was checked using NCBI Primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast). The primers were designed so that each pair overlapped each other and their sequences were conserved in all SARS-CoV-2 virus variants. As a result, 65 pairs of sequencing primers were selected to amplify the complete genome of SARS-CoV-2 virus variants with an overlap of about 100 nucleotide base pairs (bp). The estimated length of the amplicons ranged from 600 to 772 bp (zenodo.org). Oligonucleotides were synthesized on an automatic DNA/RNA Synthesizer H-16 oligonucleotide synthesizer (K&A Labs GmbH, Schaaflheim, Germany) using the phosphoramidite method and was performed according to the manufacturer's protocol. The synthesized primers were eluted from the columns with a concentrated ammonia solution. The primers were then dried on a rotary evaporator and purified by alcohol precipitate.

2.5. Polymerase Chain Reaction (PCR) Setup

Amplification was performed on a Mastercycler X50s thermal cycler using the Platinum SuperFi PCR Master Mix kit (Invitrogen, Thermo Fisher Scientific, Vilnius, Lithuania) according to the manufacturer's instructions. PCR was performed in a total volume of 25 µl, composed of: 12.5 µl of 2X Platinum SuperFi PCR Master Mix, 1.25 µl of each of 10 µM forward and reverse primers, 3 µl of cDNA template, 5 µl of 5X SuperFi GC Enhancer, and PCR-grade water to bring the volume to 25 µl. PCR products were amplified using the following conditions: initial denaturation 95°C – 0.5 min; with subsequent 35 amplification cycles with denaturation at 95°C for 0.1 min, annealing at 57°C for 0.5 min, elongation at 72°C for 0.5 min; final elongation at 72°C for 5 min.

Horizontal gel electrophoresis was performed in 1.5% agarose gel (Sigma, USA) stained with ethidium bromide in Tris-acetate buffer at a voltage of 100 volts/cm of gel length for 30 minutes. The gel was subsequently viewed using a MiniBIS Pro transilluminator (DNR Bio Imaging Systems, Ltd, Jerusalem, Israel). Visualization and documentation of gel electrophoresis results were performed using the GelCapture program (DNR Bio-Imaging Systems Ltd, Israel). A 100 bp DNA Ladder (New England Biolabs, Ipswich, MA, USA) was used as a molecular mass marker. The PCR product was purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, Carlsbad, California, USA) according to the manufacturer's instructions.

2.6. Determination of Nucleotide Sequences

Sequencing of the whole SARS-CoV-2 virus genome after purification of the PCR product was carried out using termination dideoxynucleotides (Sanger method) with the AB BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Inc., Austin, TX, USA) and specific overlapping primers designed from different viral genes used in the amplification step. The products were purified using the BigDye Xterminator kit (Applied Biosystems, Foster City, California, USA) and sequenced using a 3130 XL Genetic Analyzer (HITACHI, Tokyo, Japan). After sequencing, the obtained nucleotide sequence data were processed using the Sequencher v.5.4 program (Gene Codes Corporation, Ann Arbor, MI, USA).

2.7. Lineage Determination and Mutation Identification of the Studied Isolates

The SARS-CoV-2 virus strain lineage determination was performed using the Pangolin COVID-19 database (<https://pangolin.cog-uk.io>). The alignment of the SARS-CoV-2 virus nucleotide sequences with the reference strain and the identification of mutations were performed using the COVID-19 Genome Annotator Tool and Annotator (<http://giorgilab.unibo.it/coronannotator>)

2.8. Analysis of Non-Synonymous Mutation Function

PROVEAN software (<http://provean.jcvi.org/index.php>) was used to determine whether the selected mutations independently resulted in potential loss of function or a neutral effect. Mutation scores above the default threshold of -2.5 imply a neutral effect, while scores below this threshold indicate a deleterious effect [15].

2.9. Phylogenetic Analysis of Nucleotide Sequences

Evolutionary analysis was performed in MEGA 11 [16]. A phylogenetic tree including three samples, a reference genome, and genomes of different SARS-CoV-2 lineages was constructed using the Neighbor-Joining method [17]. The percentage of replicates in which related taxa were grouped together in a bootstrap test (1000 replicates) was shown next to the branches [18]. The tree was drawn to scale, with branch lengths (next to the branches) in the same units as the evolutionary distances used to construct the phylogenetic tree. Evolutionary distance was calculated using the maximum composite likelihood method [19] and expressed in units of base substitutions per site. To construct the phylogenetic tree, sequences were first aligned in the GenBank database and viral nucleotide sequences that were similar to the sequence of the strain under study were selected. The most suitable substitution model was selected for tree construction. A preliminary tree was constructed using the appropriate model, then the tree was pruned, typical strains were selected by year and territory, and the lineage of the strain under study was determined. When constructing a new phylogenetic tree, strains of another lineage were selected.

3. Results

3.1. PCR Amplification of SARS-CoV-2 Virus Strains

After RNA extraction and cDNA synthesis, amplification was performed using specific sequencing primers (zenodo.org) for the complete SARS-CoV-2 virus genome by PCR according to the manufacturer's protocol described above. Figures 1, 2 and 3 show the results of the electropherogram using the developed 65 pairs of sequencing primers.

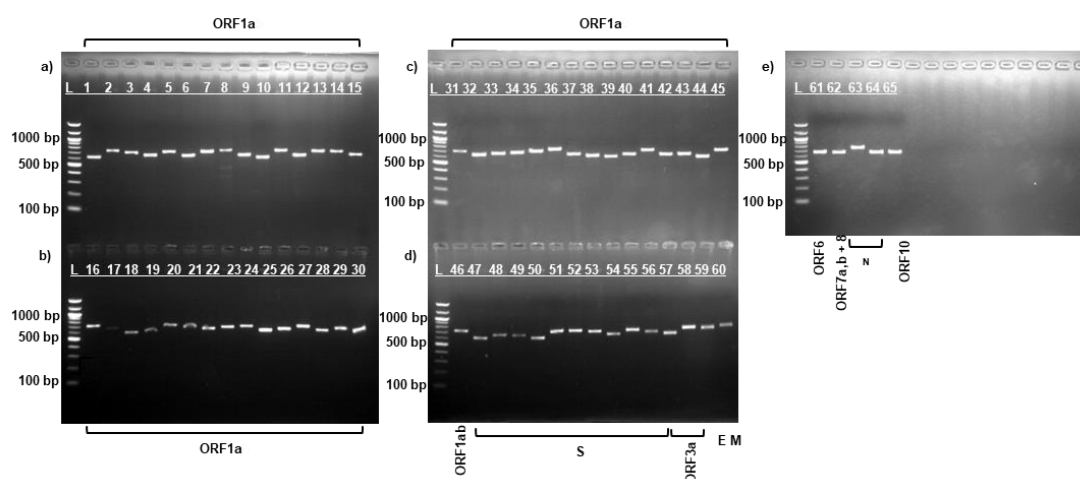


Figure 1. - Electrophoresis of amplified fragments of all genes of the SARS-CoV-2/human/KAZ/Britain/2021 strain. Upper and lower gel (a to d: lanes 1 to 46 – ORF1ab gene; lower gel d: lanes 47 to 56 – S gene; lanes 57 to 58 – ORF3a gene; lane 59 – E gene; lane 60 – M gene; upper gel e: lane 61 – ORF6 gene; lane 62 – ORF7a, ORF7b and ORF8; lanes 63 to 64 – N gene; lane 65 – ORF 10 gene; lane M – 100 bp DNA marker.

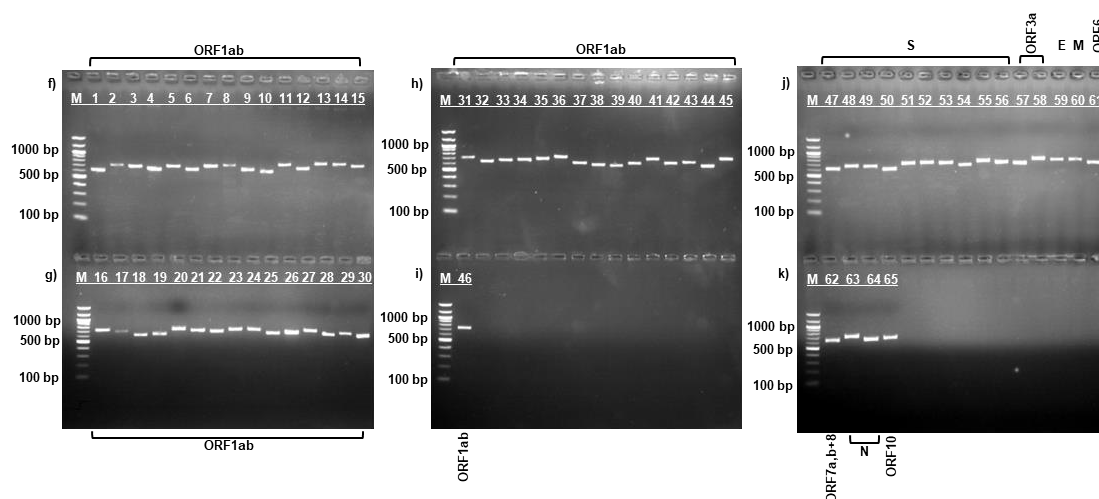


Figure 2. - Electrophoresis of amplified fragments of the entire gene of the SARS-CoV-2/human/KAZ/B1.1/2021 strain. Upper and lower gel (f to i: lanes 1 to 46 – ORF1ab gene; upper gel d: lanes 47 to 56 – S gene; lanes 57 to 58 – ORF3a gene; lane 59 – E gene; lane 60 – M gene; lane 61 – ORF6 gene; lower gel k: lane 62 – ORF7a, ORF7b and ORF8 genes; lanes 63 to 64 – N gene; lane 65 – ORF 10 gene; lane M – 100 bp DNA marker.

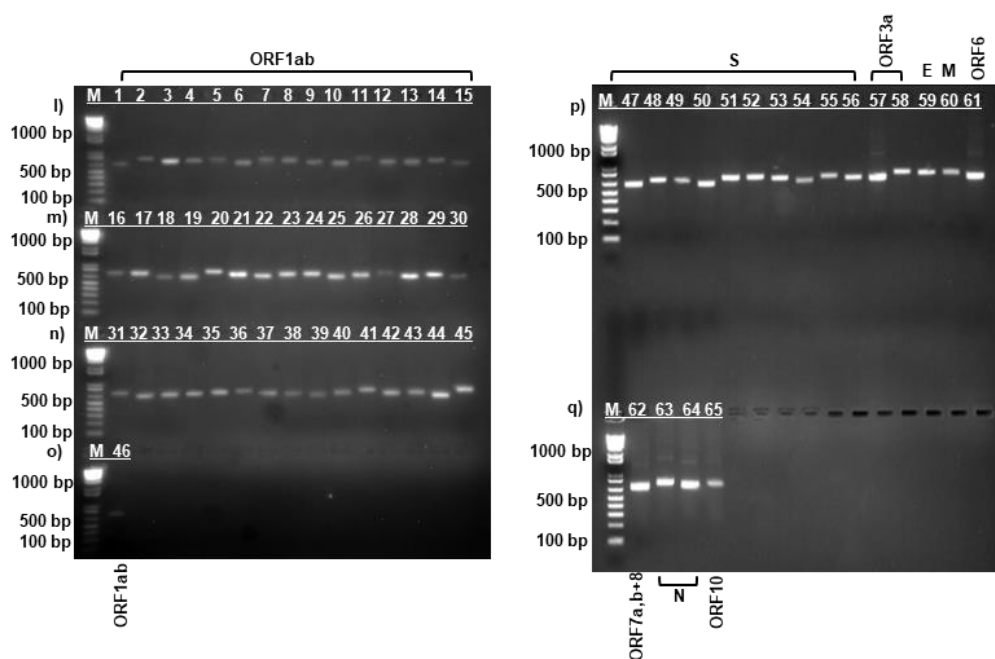


Figure 3. - Electrophoresis of amplified fragments of the entire gene of the SARS-CoV-2/human/KAZ/Delta020/2021 strain. Upper and lower gel (l to o: lanes 1 to 46 – ORF1ab gene; upper gel p: lanes 47 to 56 – S gene; lanes 57 to 58 – ORF3a gene; lane 59 – E gene; lane 60 – M gene; lane 61 – ORF6 gene; lower gel q: lane 62 – ORF7a, ORF7b and ORF8 genes; lanes 63 to 64 – N gene; lane 65 – ORF 10 gene; lane M – 100 bp DNA marker.

As can be seen in Figures 1, 2 and 3, fragments of the complete genome of the SARS-CoV-2 virus samples were obtained using the developed sequencing primers (zenodo.org). Electrophoretic analysis yielded products with a molecular weight ranging between 612-732 bp. The length of the obtained amplicons corresponds to the length of the synthesized sequencing primers.

3.2. Characteristics of the Genomes of the Studied SARS-CoV-2 Virus Strains

The size of the genomes of the studied samples SARS-CoV-2/human/KAZ/Britain/2021 (KAZ/Britain/2021), SARS-CoV-2/human/KAZ/B1.1/2021 (KAZ/B1.1/2021) and SARS-CoV-2/human/KAZ/Delta-020/2021 (KAZ/Delta020/2021) were 29.751 bp, 29.815 bp and 29.840 bp, respectively, and the GC content was 38%, 37.95% and 38%, respectively [3,20,21].

The nucleotide sequences of Kazakhstan SARS-CoV-2 virus strains were analyzed using the Pangolin COVID-19 database (<https://pangolin.cog-uk.io>). According to Pangolin COVID-19 data, the studied strains KAZ/Britain/2021, KAZ/B1.1/2021 and KAZ/Delta020/2021 belong to the B.1.1.7, B.1.1 and AY.122 lineages of the SARS-CoV-2 virus, respectively.

The COVID-19 Genome Annotator tool ([giorgilab](https://giorgilab.com)) was used to detect mutations in the obtained nucleotide sequences. According to the COVID-19 Genome Annotator data, a total of 127 mutations were detected in the studied strains compared to the reference strain. The most variable regions in the analysis of the genomic distribution of SNP (Single Nucleotide Polymorphism) and amino acid substitutions were the ORF1ab protein, which makes up 2/3 of the SARS-CoV-2 virus (Table 1), and the S protein (Table 2 and Figure 4).

Table 1. Variations in the 5'UTR, ORF1ab, and 3'UTR regions of the studied SARS-CoV-2 virus strains compared to the Wuhan-Hu-1 reference sequence ([NC 045512](https://www.ncbi.nlm.nih.gov/nuclref/NC_045512)).

Protein	Data for strain:									Amino acid change
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B1.1/2021		KAZ/Delta020/2021		Type of mutation	
	Position	Variant	Variant	Position	Variant	Position	Variant	Position		
5' UTR ^b	106	C	-	-	T	29	-	-	-	106
	210	G	-	-	-	-	T	184	-	210
	241	C	T	215	T	164	T	215	-	241
ORF1a ^b	344	C	-	-	T	267	-	-	SNP	L27F
	913	C	T	887	-	-	-	-	SNP_silent	S36S
	1048	G	-	-	-	-	T	1022	SNP	K81N ^c
	1688	A	C	1662	-	-	-	-	SNP	I295L
	1899	G	-	-	-	-	T	1873	SNP	R365L
	2110	C	T	2084	-	-	-	-	SNP_silent	N435N
	2530	A	-	-	G	2453	-	-	SNP_silent	E575E
	3037	C	T	3011	T	2960	T	3011	SNP_silent	F106F
	3267	C	T	3241	-	-	-	-	SNP	T183I
	4181	G	-	-	-	-	T	4155	SNP	A488S
	4449	C	-	-	A	4372	-	-	SNP	T577N
	4455	C	-	-	T	4378	-	-	SNP	A579V
	4475	C	-	-	T	4398	-	-	SNP	R586C
	5388	C	A	5362	-	-	-	-	SNP	A890D
	5829	A	-	-	C	5752	-	-	SNP	K1037T

Protein	Data for strain:									Amino acid change
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B.1.1/2021		KAZ/Delta020/2021		Type of mutation	
	Position	Variant	Variant	Position	Variant	Position	Variant	Position		
	5986	C	T	5960	-	-	-	-	SNP_silent	F1089F
	6402	C	-	-	-	-	T	6376	SNP	P1228L
	6954	T	C	6928	-	-	-	-	SNP	I1412T
	7042	G	T	7016	-	-	-	-	SNP	M1441I
	7124	C	-	-	-	-	T	7098	SNP	P1469S
	8986	C	-	-	-	-	T	8960	SNP_silent	D144D
	9053	G	-	-	-	-	T	9027	SNP	V167L
	9749	A	-	-	G	9672	-	-	SNP	K399E
	9867	T	-	-	G	9790	-	-	SNP	L438R
	10029	C	-	-	-	-	T	10003	SNP	T492I
	10198	C	-	-	T	10121	-	-	SNP_silent	D48D
	11195	C	T	11169	-	-	-	-	SNP	L75F
	11201	A	-	-	-	-	G	11175	SNP	T77A
	11288	TCTGGTTT T	del	11261	del	11210	-	-	SNP_stop	S106
	11332	A	-	-			G	11306	SNP_silent	V120V
	14120	C	T	14085			-	-	SNP	P218L
	14408	C	T	14373	T	14322	T	14382	SNP	P314L
	14676	C	T	14641	-	-	-	-	SNP_silent	P403P
	15017	C	-	-	T	14931	-	-	SNP	A517V
	15279	C	T	15244	-	-	-	-	SNP_silent	H604H
	15451	G	-	-	-	-	A	15425	SNP	G662S
	16176	T	C	16141	-	-	-	-	SNP	T903T
	16466	C	-	-	-	-	T	16440	SNP	P77L
	18271	G	-	-	-	-	A	18245	SNP	E78K
	18337	G	-	-	-	-	T	18311	SNP	A100S
	19220	C	-	-	-	-	T	19194	SNP	A394V
	20405	C	T	20370	-	-	-	-	SNP	P262L
	20759	C	-	-	T	20673	-	-	SNP	A34V
	21080	A	-	-	G	20994	-	-	SNP	K141R
	21215	A	G	21180	-	-	-	-	SNP	H186R

Protein	Data for strain:								Amino acid change	
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B.1.1/2021		KAZ/Delta020/2021			Type of mutation
	Position	Variant	Variant	Position	Variant	Position	Variant	Position		
	21446	A	-	-	G	21360	-	-	SNP	K263R
3' UTR	27389	C	-	-	T	27303	-	-	-	27389
	29733	-	-	-	TA	29648	-	-	-	29733
	29742	G	-	-	-	-	T	29716	-	29742
	29755	-	-	-	C	29672	-	-	-	29755
	29790	-	-	-	T	29708	-	-	-	29790

^a Severe acute respiratory syndrome coronavirus 2 strain Wuhan-Hu-1, complete genome sequence (GenBank accession number NC_045512). ^b UTR, untranslated region. ^c K81N, the K-to-N change at position 81.

The data presented in Table 1 show that the analysis of mutations in the 5'UTR (untranslated region), ORF1ab and 3'UTR regions of the studied isolates revealed a total of 61 variations at the nucleotide level. Among them, 18 nucleotide substitutions and 1 deletion were found in the strain KAZ/Britain/2021, 21 nucleotide substitutions and 1 deletion in the strains KAZ/B.1.1/2021, and 20 mutations in the strain KAZ/Delta020/2021. A mutation at position 241 in the 5'UTR region of the virus was detected in all three strains studied and resulted in a C to T nucleotide substitution. A C to T nucleotide substitution at positions 3037 and 14408 in the ORF1ab region was detected in all strains studied and resulted in one silent substitution (SNP silent) at position 106 (F106F) and one missense mutation at position 314 (P314L), respectively. A deletion of one amino acid residue S106 (serine → deletion) was observed in two samples studied (KAZ/Britain/2021 and KAZ/B.1.1/2021) in the NSP6 region.

Table 2. Mutations in the S protein of the studied SARS-CoV-2 virus strains compared to the Wuhan-Hu-1 reference sequence ([NC_045512](#)).

Protein	Data for strain:							
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B.1.1/2021		KAZ/Delta020/2021	
	Position	Variant	Variant	Position	Variant	Position	Variant	Position
S	21618	C	-	-	-	-	G	21592
	21646	C	-	-	T	21560	-	-
	21648	C	-	-	T	21562	-	-
	21765	TACATG	del	21729	-	-	-	-
	21766	A	-	-	-	-	del	21739
	21784	T	-	-	A	21698	-	-
	21789	C	-	-	T	21703	-	-
	21846	C	-	-	T	21760	-	-
	21987	G	-	-	-	-	A	21961
	21993	ATT	del	21951	-	-	-	-
	22185	C	-	-	-	-	T	22159
	22407	A	-	-	-	-	T	22381
	22917	T	-	-	-	-	G	22891
	22995	C	-	-	-	-	A	22969

Protein	Data for strain:							
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B1.1/2021		KAZ/Delta020/2021	
	Position	Variant	Variant	Position	Variant	Position	Variant	Position
	23014	A	-	-	C	22928	-	-
	23063	A	T	23019	-	-	-	-
	23271	C	A	32227	-	-	-	-
	23403	A	G	23359	G	23317	G	23377
	23520	C	-	-	T	23434	-	-
	23604	C	A	23560	-	-	G	23578
	23709	C	T	23665	-	-	-	-
	23751	C	-	-	T	23665	-	-
	23997	C	-	-	T	23911	-	-
	24000	G	-	-	T	23914	-	-
	24410	G	-	-	-	-	A	24384
	24506	T	G	24462	-	-	-	-
	24538	A	-	-	T	24452	-	-
	24914	G	C	24870	-	-	-	-

^a Severe acute respiratory syndrome coronavirus 2 strain Wuhan-Hu-1, complete genome sequence (GenBank accession number [NC_045512](#)).

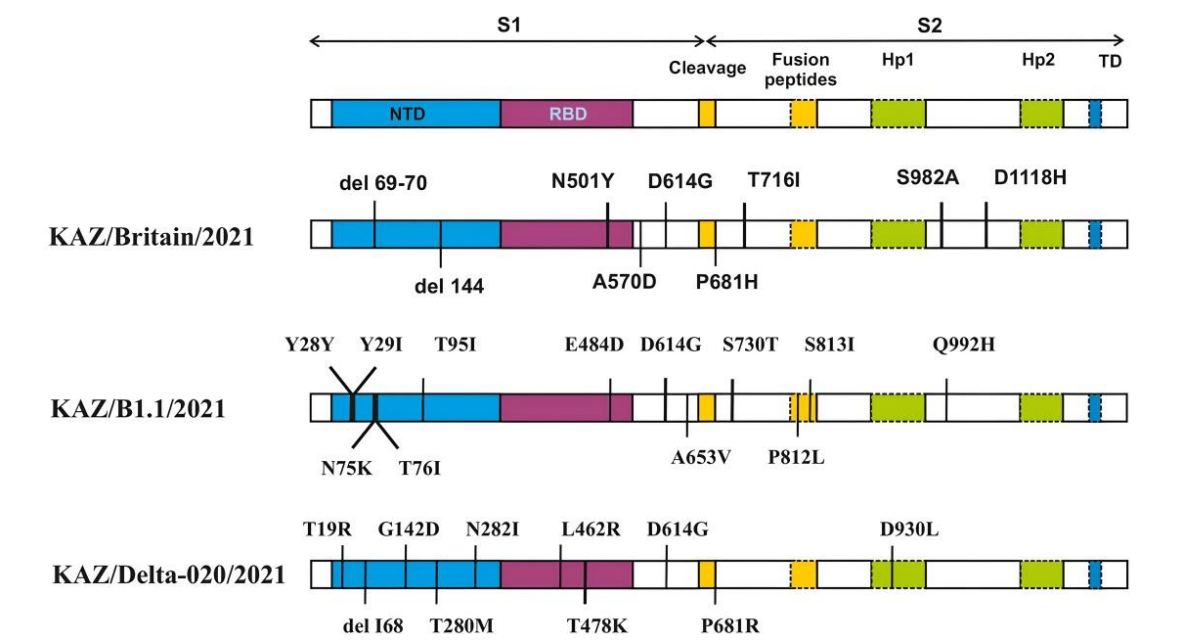


Figure 4. - Amino acid changes in the S protein of the studied strains. Description of the SARS-CoV-2 spike mutation in different virus isolates: KAZ/Britain/2021, KAZ/B1.1/2021 and KAZ/Delta020/2021. Colored bars describe the structural domain of the spike protein. NTD - N-terminal domain; RBD - receptor-binding domain; Hp1 - thermal protein 1; Hp2 - thermal protein 2. TD - transmembrane domain.

Table 2 and Figure 4 shows the distribution of SNP and amino acid substitutions identified in the S protein of the studied strains. A total of 31 mutations were found in the S protein of the studied strains. Among them, 7 amino acid substitutions and 2 deletions were detected in the KAZ/Britain/2021 strain, 12 amino acid substitutions in the KAZ/B1.1/2021 strain, and 9 amino acid substitutions and 1 deletion in the KAZ/Delta020/2021 strain.

Mutational changes in the virus occurred more often in the S1 region than in the S2 region. In the S1 region of the KAZ/Britain/2021 strain, 2 deletions in the NTD region and 4 amino acid changes were detected, of which 1 mutation belongs to the RBD region. In the S2 region, 3 amino acid changes were detected compared to the reference strain. In the strain KAZ/B1.1/2021, 12 mutations were detected compared to the original strain, such as Y28Y, T29I, N74K, T76I, T95I, E484D, D614G, A653V, S730T, P812L, S813I and Q992H. In the S protein of the strain KAZ/Delta020/2021, sets of mutations (L452R, T478K and P681R) were found, which are unique only to the Delta variant.

The distribution of SNPs and amino acid substitutions, in addition to the ORF1ab and S proteins, was observed in the ORF8, N, ORF7a, ORF3a, M, ORF6, and ORF7b proteins of the studied strains and amounted to 9, 9, 6, 4, 3, 2, and 1 amino acid substitution, respectively (Table 3).

Table 3. Mutations in the remaining proteins of the studied SARS-CoV-2 virus strains compared to the Wuhan-Hu-1 reference sequence ([NC 045512](#)).

Protein	Data for strain:									Amino acid change
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B1.1/2021		KAZ/Delta020/2021		Type of mutation	
	Position	Variant	Variant	Position	Variant	Position	Variant	Position		
ORF3a	25459	C	-	-	-	-	T	25443	SNP	S26L ^b
	25688	C	-	-	T	25602	-	-	SNP	A99V
	25838	G	T	25794	-	-	-	-	SNP	W149L
	26110	C	-	-	T	26024	-	-	SNP	P240S
M	26767	T	-	-	-	-	C	26741	SNP	I82T
	26895	C	-	-	T	26809	-	-	SNP	H125Y
	27008	G	-	-	T	26922	-	-	SNP	K162N
ORF6	27281	GG	AA	27237	-	-	-	-	SNP_stop	W27
	27285	TC	AT	27241	-	-	-	-	SNP	NL28KF
ORF7a	27527	C	-	-	-	-	T	27501	SNP	P45L
	27638	T	-	-	-	-	C	27612	SNP	V82A
	27630	C	-	-	T	27544	-	-	SNP_silent	A79A
	27667	G	-	-	A	27581	-	-	SNP	E92K
	27739	C	-	-	T	27653	-	-	SNP	L116F
	27752	C	-	-	-	-	T	27726	SNP	T120I
ORF7b	27874	C	-	-	-	-	T	27848	SNP	T40I
ORF8	27919	T	-	-	-	-	C	27893	SNP	I9T
	27972	C	T	27928	-	-	-	-	SNP_stop	Q27
	28048	G	T	28004	-	-	-	-	SNP	R52I
	28095	A	T	28051	-	-	-	-	SNP_stop	K68
	28111	A	G	28067	-	-	-	-	SNP	Y73C
	28251	T	-	-	-	-	C	28225	SNP	F120L
	28253	C	-	-	-	-	A	28227	SNP	F120L
	28255	T	-	-	-	-	A	28229	SNP	I121N
	28258	A	-	-	-	-	G	28232	SNP_silent	*122*
N	28280	GAT	CTA	28236	-	-	-	-	SNP	D3L

Protein	Data for strain:									Amino acid change
	Wuhan-Hu-1 ^a		KAZ/Britain/2021		KAZ/B.1.1/2021		KAZ/Delta020/2021		Type of mutation	
	Position	Variation	Variation	Position	Variation	Position	Variation	Position		
	28461	A	-		-	-	G	28435	SNP	D63G
	28881	GGG	AAC	28837	AAC	28837	-	-	SNP	RG203KR
	28881	G	-	-	-	-	T	28855	SNP	R203M
	28916	G	-	-	-	-	T	28890	SNP	G215C
	28977	C	T	28933	-	-	-	-	SNP	S235F
	29236	C	-	-	-	-	T	29210	SNP_silent	G312G
	29402	G	-	-	-	-	T	29376	SNP	D377Y
	29436	A	-	-	T	29350	-	-	SNP	K388I

^a Severe acute respiratory syndrome coronavirus 2 strain Wuhan-Hu-1, complete genome sequence (GenBank accession number [NC_045512](#)). ^b S26L, the S-to-L change at position 26.

As shown in Table 3, a total of 34 variations were detected in the three isolates compared to the original strain. In the current study, four mutations were detected in the ORF3a protein across different samples: one mutation (W149L) in the KAZ/Britain/2021 sample, two mutations (A99V, P240S) in the KAZ/B.1.1/2021 strain, and one mutation (S26L) in the KAZ/Delta020/2021 strain. Three mutations were detected in the M protein across the strains: two mutations (H125Y and K162N) in the KAZ/B.1.1/2021 strain and one mutation (I82T) in the KAZ/Delta020/2021 strain. Two mutations (W27 and NL28KF) were detected in the ORF6 protein, which were found only in the KAZ/Britain/2021 strain. Six mutations were detected in the ORF7a protein: three mutations (A79A, E92K and L116F) were detected in the KAZ/B.1.1/2021 strain, and three mutations (P45L, V82A and T120I) were detected in the KAZ/Delta020/2021 strain. Only one mutation (T40I) was detected in the ORF7b protein in the KAZ/Delta020/2021 strains. Eight mutations were detected in the ORF8 protein: four mutations (Q27, R52I, K68 and Y73C) were found in the KAZ/B.1.1/2021 strain, and four mutations (F120L, F120L, I212N and 122) were found in the KAZ/Delta020/2021 strain. Nine mutations were detected in the N protein compared to the original virus strain: three mutations in the KAZ/Britain/2021 strain (D3L, RG203KR and S235F), two mutations in the KAZ/B.1.1/2021 strain (RG203K and K388I) and four mutations in the KAZ/Delta020/2021 strain (D63G, R203M, G215C, G312G and D377Y).

3.3. Impact of Mutations on Biological Function of Proteins in the Studied SARS-CoV-2 Samples

The PROVEAN web server was used to assess whether the selected mutations could lead to a potential loss of function or remain neutral. Loss of function occurs when a mutation leads to the formation of a non-functional protein. At the same time, a neutral result means that the protein function is preserved despite the presence of a mutation. The PROVEAN platform is focused only on the analysis of the individual effects of each of the mutations identified in the studied virus isolates (Table 4) [22].

Table 4. Estimates of various mutations in the genomes of the studied SARS-CoV-2 strains.

Protein	KAZ/Britain/2021			KAZ/B1.1/2021			KAZ/Delta020/2021		
	Amino acid change	Оценка PROVEA N	The effect of variation on protein	Amino acid change	Оценка PROVEA N	The effect of variation on protein	Amino acid change	Оценка PROVEA N	The effect of variation on protein
ORF1ab	I295L	0.232	Neutral	L27F	-0.047	Neutral	K81N	-0.070	Neutral
	T183I	0.216	Neutral	A579V	0.011	Neutral	R365L	-0.939	Neutral
	T577N	0.240	Neutral	R586C	-0.727	Neutral	A488S	-0.061	Neutral
	A890D	-1.749	Neutral	K1037 T	-1.196	Neutral	P1228 L	-1.038	Neutral
	I1412T	-0.370	Neutral	K399E	-1.877	Neutral	P1469S	0.338	Neutral
	M1441I	0.263	Neutral	L438R	0.659	Neutral	V167L	-0.696	Neutral
	L75F	-2.290	Neutral	P314L	-0.446	Neutral	T492I	1.435	Neutral
	P218L	-5.021	Deleterious	A517V	-1.291	Neutral	T77A	-0.878	Neutral
	P314L	-0.446	Neutral	A34V	1.158	Neutral	P314L	-0.446	Neutral
	P262L	-0.014	Neutral	K141R	-0.221	Neutral	G662S	-2.475	Neutral
	H186R	-0.267	Neutral	K263R	-1.344	Neutral	P77L	-6.845	Deleterious
	-	-	-				E78K	-1.123	Neutral
	-	-	-	-	-	-	A100S	1.338	Neutral
	-	-	-	-	-	-	A394V	-1.523	Neutral
S	H69del	0.260	Neutral	Y28Y	0.000	Neutral	T19R	-0.839	Neutral
	Y145del	0.853	Neutral	T29I	-1.538	Neutral	I68del	-0.821	Neutral
	N501Y	-0.090	Neutral	N74K	-1.309	Neutral	G142D	-0.277	Neutral
	A570D	-0.682	Neutral	T76I	-0.115	Neutral	T208M	-0.314	Neutral
	D614G	0.598	Neutral	T95I	-1.214	Neutral	N282I	-3.717	Deleterious
	P681H	0.060	Neutral	E484D	-0.210	Neutral	L452R	0.559	Neutral
	T716I	-3.293	Deleterious	D614G	0.598	Neutral	T478K	-0.524	Neutral
	S982A	-1.505	Neutral	A653V	-0.715	Neutral	D614G	0.598	Neutral
	D1118 H	-1.142	Neutral	S730T	-0.040	Neutral	P681R	0.741	Neutral
	-	-	-	P812L	-0.868	Neutral	D950N	-1.631	Neutral
	-	-	-	S813I	-2.867	Deleterious	-	-	-
	-	-	-	Q992H	-4.059	Deleterious	-	-	-
ORF3	-	-	-	-	-	-	S26L	-2.314	Neutral
				A99V	-1.962	Neutral	-	-	-

Protein	KAZ/Britain/2021			KAZ/B.1/2021			KAZ/Delta020/2021		
	Amino acid change	Оценка PROVEA N	The effect of variation on protein	Amino acid change	Оценка PROVEA N	The effect of variation on protein	Amino acid change	Оценка PROVEA N	The effect of variation on protein
	W149L	-9.419	Deleterious	-	-	-	-	-	-
M	-	-	-	P240S	-1.495	Neutral	-	-	-
	-	-	-	H125Y	0.799	Neutral	I82T	-3.853	Deleterious
	-	-	-	K162N	0.501	Neutral	-	-	-
ORF7a	-	-	-	-	-	-	P45L	-10.000	Deleterious
	-	-	-	-	-	-	V82A	-2.667	Deleterious
	-	-	-	A79A	0.000	Neutral	-	-	-
	-	-	-	E92K	-1.842	Neutral	-	-	-
	-	-	-	L116F	-1.263	Neutral	-	-	-
	-	-	-	-	-	-	T120I	-1.789	Neutral
ORF	-	-	-	-	-	-	T40I	-2.000	Neutral
ORF8	R52I	-6.417	Deleterious	-	-	-	I9T	-1.333	Neutral
	Y73C	-4.500	Deleterious	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	F120L	-2.667	Deleterious
	-	-	-	-	-	-	F120L	-2.667	Deleterious
	-	-	-	-	-	-	I121N	-0.667	Neutral
N	D3L	-0.230	Neutral	-	-	-	D63G	-0.929	Neutral
	-	-	-	-	-	-	R203M	-3.304	Deleterious
	-	-	-	-	-	-	G215C	-0.953	Neutral
	-	-	-	-	-	-	S235F	-1.738	Neutral
	-	-	-	-	-	-	D377Y	-1.779	Neutral
	-	-	-	K388I	-1.204	Neutral	-	-	-
	-	-	-	-	-	-	-	-	-

^aThe threshold value was set at -2.5.

Table 4 shows that the proportion of loss-of-function mutations detected in the three studied genomes (KAZ/Britain/2021, KAZ/B.1/2021, and KAZ/Delta020/2021) of SARS-CoV-2 was studied, and 5 (P218L, T716I, W149L, R52I, and Y73C), 2 (S813I, and Q992H), and 8 (P77L, L452R, I82T, P45L,

V82A, F120L, F120L, and R203M) loss-of-function mutations were identified, respectively. Among the genes in the studied samples, the proportion of loss-of-function mutations was higher in the S and ORF8 genes than in other genes.

3.4. Phylogenetic Analysis

Phylogenetic analysis between the studied isolates and other isolates belonging to different lineages of the SARS-CoV-2 virus from the international GenBank NCBI database are presented in Figure 5.

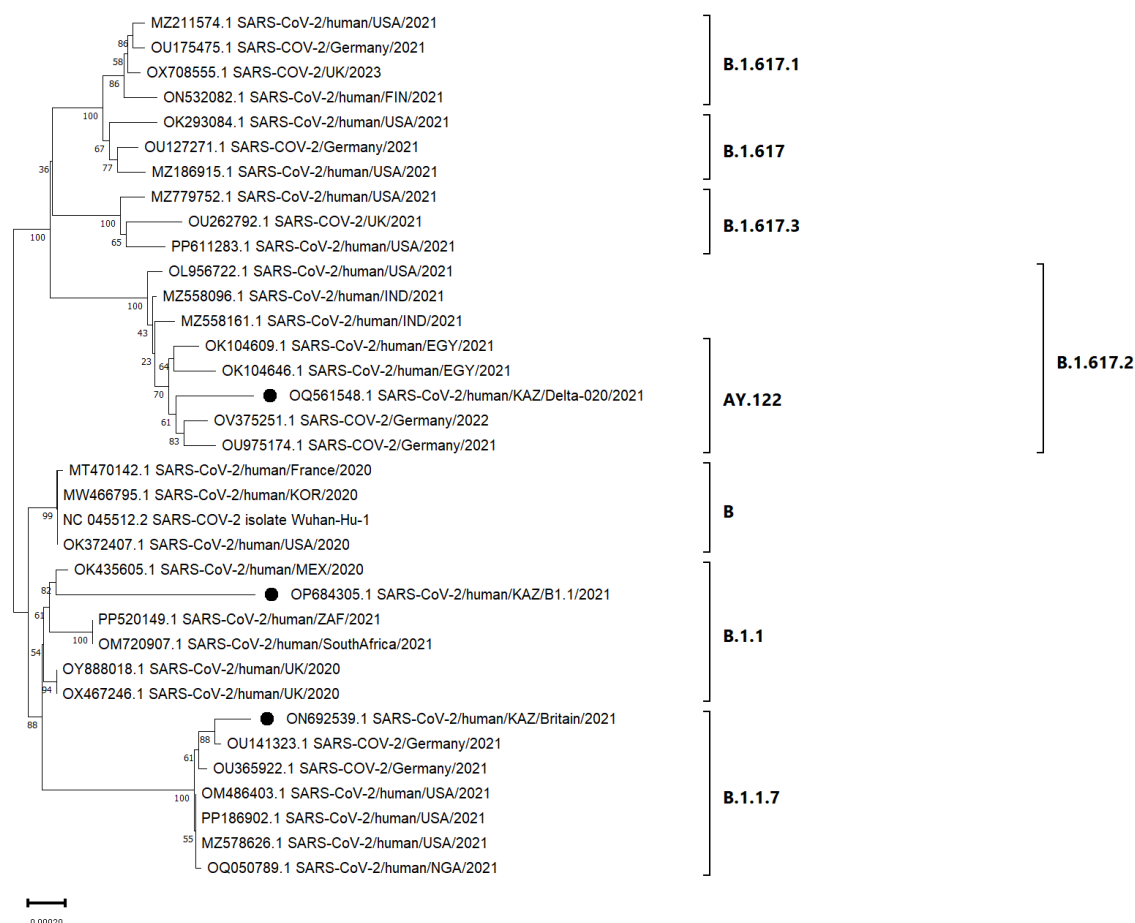


Figure 5. – Phylogenetic analysis of the studied SARS-CoV-2 isolates and 35 global strains belonging to different virus lineages such as B.1.617.1, B.1.617.2, B.1.617, B.1.617.3, AY.122, B.1.1.7, B.1.1 and B, which were obtained from the NCBI GenBank database. Here, the x-axis represents the scale of the tree. The studied SARS-CoV-2 virus samples belonging to different lineages are indicated by circles.

Based on the phylogenetic analysis, the studied strains KAZ/Britain/2021, KAZ/B1.1/2021 and KAZ/Delta020/2021 belong to different SARS-CoV-2 lineages. KAZ/Britain/2021 formed a group (bootstrap (BS) = 100%) with isolates belonging to the B.1.1.7 SARS-CoV-2 lineage. The nucleotide identity between them ranged from 99.96 to 99.97 percent. Within the monophylogenetic group, [OU141323.1](#) SARS-CoV-2/Germany/2021 was the most similar to KAZ/Britain/2021, with a nucleotide similarity of 99.97%. KAZ/B1.1/2021 groups (bootstrap (BS) = 100%) with various samples belonging to the B.1.1 SARS-CoV-2 lineage. KAZ/B1.1/2021 closely matched the samples from Mexico ([OK435605.1](#)), showing a nucleotide identity of 99.84%. KAZ/Delta020/2021 formed a monophyletic group (bootstrap (BS) = 61% and (BS) = 100%) with samples that belong to the AY.122 and B.1.617.2

lineages, respectively. However, our KAZ/Delta020/2021 showed high similarity to isolates from Germany ([OV375251.1](#) and [OU975174.1](#)), which had a nucleotide identity of 99.94%.

4. Discussion

Cleaveland S. et al. reported that most viruses infecting humans are zoonotic. Zoonotic viruses, after entering a cell, adapt inefficiently to a new host and replicate and transmit slowly [23]. Their transmission from animal to human and from human to human depends on many factors, including potential adaptive evolution to virulent strains [24].

RNA viruses are characterized by higher replication fidelity ($\sim 10^{-4}$ error/site/cycle) and genetically diverse RNA polymerase compared to DNA viruses [25]. However, when RNA viruses circulate in the community, genetic changes continuously occur due to copying errors of RNA polymerase. This, in turn, leads to mutations in the genome [26]. Lee et al. analyzed the rate of genome evolution of several SARS-CoV-2 virus strains over one month and found that the average evolution rate ranged from 1.7926×10^{-3} to 1.8266×10^{-3} substitutions per site per year [27,28], but four months after the pandemic, the mutation rate of the whole SARS-CoV-2 virus genome was 3.95×10^{-4} per nucleotide per year [29].

The rapid evolution of the SARS-CoV-2 genome highlights the need to develop antiviral drugs against the virus [30]. To develop effective antiviral drugs, it is necessary to determine which variant is most actively circulating in society during the pandemic. This depends on the data collected on COVID-19 infection, the epidemiological features among different population groups, as well as the patterns of viral spread in different areas. The modern approach to the use of genomic and information technologies in epidemiological surveillance of SARS-CoV-2 pathogens occupies an important place in measures to prevent and control the virus [31].

Sanger sequencing is considered the most optimal method for sequencing short fragments ($<1,000$ bp), and is useful for filling gaps in partial whole genomes [32,33]. An important step for the successful implementation of the Sanger method is the production of a PCR amplicon from the samples under study and the development of sequencing oligonucleotide primers for the amplification of this PCR amplicon [26].

It is impossible to obtain the complete genomic nucleotide sequence of SARS-CoV-2 virus in a single reaction using the Sanger method. Therefore, in our current study, we designed a set of sequencing primers targeting SARS-CoV-2 virus to obtain the complete genomic nucleotide sequence ([zenodo.org](#)). The specific designed sequencing primers were selected based on the Wuhan-Hu-1 reference strain, and each designed primer pair overlapped with each other and their sequence was conserved among all SARS-CoV-2 virus variants. The length of the sequencing primers ranged from 600 bp to 772 bp with a GC content of 38% to 50%, and the melting temperature was in the range of 55–57 °C. The PCR products of the studied samples were obtained using the designed sequencing primers.

The length of the amplicons obtained (Figure 1, Figure 2 and Figure 3) corresponds to the length of the synthesized sequencing primers. The developed specific primers covered 100% and amplified the entire genome of the studied samples. After sequencing, the nucleotide sequences of the studied samples were obtained and analyzed in the Pangolin COVID-19 database (<https://pangolin.cog-uk.io>). According to the Pangolin COVID-19 data, the KAZ/Britain/2021 strain belongs to the B.1.1.7 lineage (Alpha variant). Alpha differs from other variants of the virus by the presence of mutations in the S protein, such as deletion 69-70, deletion 144, N501Y, A570D, D614G, P681H, T716I, S982A and D1118H [34,35]. The mutations identified in the S gene of the studied isolate KAZ/Britain/2021 are 100% consistent with the mutations found in the S gene of the alpha variant (Table 2 and Figure 4).

Bo Meng et al. suggest that the H69-V70 deletion in the NTD region of the S1 subunit of the spike protein, found in the studied SARS-CoV-2 virus sample, is associated with increased infectivity and evasion of the host immune response [36–38]. Weng S. et al. confirmed that the $\Delta 144/145$ deletion blocks the binding sites of neutralizing antibodies, which is important in preventing the virus from

entering the cell and possibly interfering with its replication [38,39]. Some studies describe that H69-V70, N501Y and P681H, which may affect viral infectivity [36,40]. N501Y increases viral infectivity by 70-80% and enhances the binding affinity of the viral S protein to human ACE2 [36,40–42]. According to some studies, mutations A570D, T716I, S982A, and D1118H are the result of accumulated mutations of the virus in the community environment, which together increase the lethality and transmissibility of the SARS-CoV-2 virus [42,43]. D614G was found in all three isolates and has become the most common mutation among SARS-CoV-2 variants during the global pandemic [44]. Lubinski B. et al. showed that P681H can increase its cleavage by furin-like proteases, although this process does not lead to viral entry [45]. According to Pangolin COVID-19 data, strain KAZ/B.1.1/2021 belongs to the B.1.1 lineage. Currently, the SARS-CoV-2 virus is divided into two lineages: A and B. Lineage B includes 47 different lineages, and lineage B.1.1 is part of this lineage (observablehq.com).

The strain KAZ/Delta020/2021, according to Pangolin COVID-19, belongs to the AY.122 lineage (Delta variant, B.1.617). As indicated by the source observablehq.com, B.1.617 is divided into three sublineages: B.1.617.1, B.1.617.2 and B.1.617.3. Dhawan M. et al. note that the B.1.617.2 lineage emerged during the second wave of coronavirus infection in India. The B.1.617.2 lineage includes 134 different lineages, one of which is AY.122. Some literature sources emphasize that B.1.617.2 is characterized by differences from other viral variants due to a unique set of mutations, such as L452R, T478K and P681R. These mutations make it particularly infectious and resistant to neutralizing antibodies in previously infected or vaccinated individuals [46–49]. Other studies have also shown that the T19R, T478K, P681R, and D950N mutations found in the S gene enhance viral replication and help it evade the host's immune response [50,51].

The resulting nucleotide sequences were tested using the COVID-19 Genome Annotator ([giorgilab](https://giorgilab.com)) to detect mutations. According to the COVID-19 Genome Annotator, a total of 127 mutations were detected in the isolates tested compared to the reference strain. In these studied isolates, the following types of mutations were encountered: SNP, silent mutations, stop codon, deletion and mutations occurring in the 5' and 3' untranslated region in the genome compared to the original strain, and their quantities in the studied genomes were 92, 17, 3, 5, 5 and 5, respectively (Table 1, Table 2 and Figure 4 and Table 3). Analysis of the distribution of SNP in the studied genomes showed that the most common in the ORF1ab gene (n = 36) and S (n = 27). The main silent mutations were found in the ORF1ab gene (n = 13), in the remaining genes S, ORF7b, ORF8 and N only one was detected, respectively. Deletions were found only in the S gene (n = 3) and ORF1ab (n = 2).

The study identified mutations in the 5'UTR (C241T), ORF1ab (F106F and P314L) and S (D614G) regions that were common to all three isolates studied. The study by Periwal N et al. showed that the C nucleotide at position 241 in the 5'UTR region was replaced by a T nucleotide as early as the summer of 2020 [52]. Kim et al. reported that this mutation in the 5'UTR region may affect the rate of transcription and replication of the SARS-CoV-2 virus [11,53,54]. Some studies predicted that the synonymous F106F mutation identified in the NSP3 region of the ORF1ab gene may play a role in mRNA processing, altering the properties of the viral protein [11,54]. The missense mutation P314L, found in the NSP12 region of the ORF1b gene, is considered to be part of the core replication/transcription complex and is a conserved protein in coronaviruses [5,55]. Thus, the P314L mutation affects SARS-CoV-2 RNA replication by participating in the activity of RdRp (RNA-dependent RNA polymerase) [56,57]. In addition, RdRp plays an important role in the process of SARS-CoV-2 viral replication and transcription [58]. D614G was detected in all three isolates and has become the most common mutation among SARS-CoV-2 variants during the global pandemic [44].

It is important to evaluate the change in function of the mutations identified in the study, which may have effects on viral circulation. Aside from this, further studies on these mutations can contribute to the development of various antiviral drugs against SARS-CoV-2. This study revealed significant changes in amino acids in structural and accessory proteins (P218L and P77L in ORF1ab; T716I, S813I, Q992H and N282I in S; W149L in ORF3a; I82T in M; P45L and V82A in ORF7a; R52I,

Y73C and F120L in ORF8; R203M in N), which may cause functional alterations and affect functional characteristics of the virus.

Phylogenetic analysis of SARS-CoV-2 virus isolates showed that the studied samples belong to different virus lineages. In the study, the KAZ/Britain/2021 strain showed significant similarity to the [OU141323.1](#) SARS-CoV-2/Germany/2021 isolate and formed a group with strains that belong to the B.1.1.7 lineage. The nucleotide similarity between these isolates was 99.97%, indicating their very close genetic relationship. KAZ/B1.1/2021 grouped with various isolates that belong to the B.1.1 lineage. In addition, it showed close similarity to samples obtained from Mexico ([OK435605.1](#)) SARS-CoV-2/human/MEX/CMX-51/2020), demonstrating a nucleotide identity of 99.84%. KAZ/Delta020/2021 clustered with isolates belonging to the AY.122 and B.1.617.2 lineages. However, KAZ/Delta020/2021 showed high similarity to isolates from Germany [OV375251.1](#) and [OU975174.1](#), with 99.94% nucleotide identity. According to Pangolin COVID-19 data, the AY.122 lineage is one of the sublineages of B.1.617. B.1.617 emerged in late 2020 in Maharashtra, India [59]. In mid-June 2021, a mutated Delta variant (B.1.617.2), known as the Delta plus, was identified in India [60].

Therefore, the SARS-CoV-2 samples were fully amplified and sequenced using the developed primers, which allowed us to identify mutations compared to the reference strain Wuhan-Hu-1 ([NC_045512.2](#)). Sanger-based whole-genome sequencing of the studied SARS-CoV-2 isolates was successfully demonstrated. The data obtained using molecular genetic methods during the pandemic are of great importance for understanding the biology of the virus, developing new diagnostic and therapeutic methods, and making informed public health decisions. Continued research in this area will allow us to be better prepared for future pandemics.

Author Contributions. Conceptualization, B.S.U., K.T.S., and Y.B.D., methodology, B.S.U., A.M.M., A.T.Z., and formal analysis, B.S.U., investigation, A.M.M., I.S., M.Z.S., and N.S.K., resources, B.S.M., Y.B.D., and L.B.K., data curation, B.S.U., writing - original draft, B.S.U., writing - review and editing, B.S.U. and K.T.S., visualization, B.S.U., supervision, A.A.K., O.V.S. and S.S.N., project administration, Y.B.D. and L.B.K. funding acquisition, A.A.K. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement. The complete genome sequence of SARS-CoV-2 in this study is available in GenBank under accession numbers [ON692539.1](#), [OP684305.1](#) and [OQ561548.1](#).

Acknowledgments. We thank the management of the Research Institute of Biomedical Biotechnology for financial support, as well as the branch of the Scientific and Practical Center for Sanitary and Epidemiological Expertise and Monitoring of the Republican State Enterprise on the Right of Economic Management of the Ministry of Health of the Republic of Kazakhstan "National Center for Public Health" for the transfer of clinical samples from patients for molecular genetic studies on COVID-19.

Conflicts of Interest. The authors declare no conflict of interest.

References

1. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J Adv Res.* 2020, 24, 91-98
2. Zhugunissov K, Zakarya K, Khairullin B, Orynbayev M, Abduraimov Y, Kassenov M, Sultankulova K, Kerimbayev A, Nurabayev S, Myrzakhmetova B, Nakhanov A, Nurpeisova A, Chervyakova O, Assanzhanova N, Burashev Y, Mambetaliyev M, Azanbekova M, Kopeyev S, Kozhabergenov N, Issabek A, Tuyskanova M, Kutumbetov L. Development of the Inactivated QazCovid-in Vaccine: Protective Efficacy of the Vaccine in Syrian Hamsters. *Front Microbiol.* 2021, 12, 720437.
3. Usserbayev B, Zakarya K, Kutumbetov L, Orynbayev M, Sultankulova K, Abduraimov Y, Myrzakhmetova B, Zhugunissov K, Kerimbayev A, Melisbek A, Shirinbekov M, Khaidarov S, Zhunushov A, Burashev Y. // *Microbiol Resour Announc.* 2022, 11, 9:e0061922.
4. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. // *N Engl J Med.* 2020, 382.8,727-733.

5. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. *Nature*. 2020, 579, 265-269.
6. Cosar B, Karagulleoglu ZY, Unal S, Ince AT, Uncuoglu DB, Tuncer G, Kilinc BR, Ozkan YE, Ozkoc HC, Demir IN, Eker A, Karagoz F, Simsek SY, Yasar B, Pala M, Demir A, Atak IN, Mendi AH, Bengi VU, Cengiz Seval G, Gunes Altuntas E, Kilic P, Demir-Dora D. *S. Cytokine Growth Factor Rev*. 2022, 63, 10-22.
7. Safari I, Elahi E. Evolution of the SARS-CoV-2 genome and emergence of variants of concern. *Arch Virol*. 2022, 167(2), 293-305
8. Andre M, Lau LS, Pokharel MD, Ramelow J, Owens F, Souchak J, Akkaoui J, Ales E, Brown H, Shil R, Nazaire V, Manevski M, Paul NP, Esteban-Lopez M, Ceyhan Y, El-Hage N. *Biology*. 2023, 12, 1267
9. Bhardwaj P, Mishra SK, Behera SP, Zaman K, Kant R, Singh R. Genomic evolution of the SARS-CoV-2 Variants of Concern: COVID-19 pandemic waves in India. *EXCLI J*. 2023, 22, 451-465
10. Vidanović D, Tešović B, Volkening JD, Afonso CL, Quick J, Šekler M, Knežević A, Janković M, Jovanović T, Petrović T, Đeri BB. *Epidemiol Infect*. 2021, 149, e246
11. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front Microbiol*. 2020, 11, 1800
12. LaTourrette K, Garcia-Ruiz H. Determinants of Virus Variation, Evolution, and Host Adaptation. *Pathogens*. 2022, 11(9), 1039
13. Márquez S, Prado-Vivar B, Guadalupe JJ, Gutierrez B, Jibaja M, Tobar M, Mora F, Gaviria J, García M, Espinosa F, Ligna E, Reyes J, Barragán V, Rojas-Silva P, Trueba G, Grunauer M, Cárdenas P. Genome sequencing of the first SARS-CoV-2 reported from patients with COVID-19 in Ecuador. *medRxiv [Preprint]*. 2020, 2020.06.11.20128330.
14. Mohammadi E, Shafiee F, Shahzamani K, Ranjbar MM, Alibakhshi A, Ahangarzadeh S, Beikmohammadi L, Shariati L, Hooshmandi S, Ataei B, Javanmard SH. Novel and emerging mutations of SARS-CoV-2: Biomedical implications. *Biomed Pharmacother*. 2021, 139, 111599
15. Choi Y, Chan A.P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015, 31(16), 2745-2747
16. Tamura K., Stecher G., and Kumar S.. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*. 2021
17. Saitou N. and Nei M.. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 1987, 4, 406-425
18. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*. 1985, 39, 783-791
19. Tamura K., Nei M., and Kumar S.. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences (USA)*. 2004, 101, 11030-11035.
20. Burashev Y, Ussebayev B, Kutumbetov L, Abduraimov Y, Kassenov M, Kerimbayev A, Myrzakhmetova B, Melisbek A, Shirinbekov M, Khaidarov S, Tulman ER. Coding Complete Genome Sequence of the SARS-CoV-2 Virus Strain, Variant B.1.1, Sampled from Kazakhstan. *Microbiol Resour Announc*. 2022, 11(12):e0111422.
21. Ussebayev B, Abduraimov Y, Kozhabergenov N, Melisbek A, Shirinbekov M, Smagul M, Nusupbayeva G, Nakhanov A, Burashev Y. Complete Coding Sequence of a Lineage AY.122 SARS-CoV-2 Virus Strain Detected in Kazakhstan. *Microbiol Resour Announc*. 2023, 12(7):e0030123
22. Christian Alfredo K. Cruz, Paul Mark B. Medina. Temporal changes in the accessory protein mutations of SARS-CoV-2 variants and their predicted structural and functional effects. 2022, 94
23. Cleaveland S, Laurenson MK, Taylor LH. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos Trans R Soc Lond B Biol Sci*. 2001, 356(1411), 991-9
24. Padhan K, Parvez MK, Al-Dosari MS. Comparative sequence analysis of SARS-CoV-2 suggests its high transmissibility and pathogenicity. *Future Virol*. 2021, 10.2217/fvl-2020-0204.
25. Parvez MK, Parveen S. Evolution and Emergence of Pathogenic Viruses: Past, Present, and Future. *Intervirology*. 2017, 60(1-2):1-7

26. Lee SH. A Routine Sanger Sequencing Target Specific Mutation Assay for SARS-CoV-2 Variants of Concern and Interest. *Viruses*. 2021, 13(12): 2386
27. Li X, Wang W, Zhao X, Zai J, Zhao Q, Li Y, Chaillon A. Transmission dynamics and evolutionary history of 2019-nCoV. *J Med Virol*. 2020, 92(5):501-511
28. Li Y, Lai DY, Zhang HN, Jiang HW, Tian X, Ma ML, Qi H, Meng QF, Guo SJ, Wu Y, Wang W, Yang X, Shi DW, Dai JB, Ying T, Zhou J, Tao SC. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol*. 2020, 17(10), 1095-1097
29. Abbasian MH, Mahmanzar M, Rahimian K, Mahdavi B, Tokhanbigli S, Moradi B, Sisakht MM, Deng Y. Global landscape of SARS-CoV-2 mutations and conserved regions. *J Transl Med*. 2023, 21, 152
30. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A; COVID-19 Genomics UK (COG-UK) Consortium; Peacock SJ, Robertson DL. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol*. 2021, 19, 409-424
31. Esman A, Dubodelov D, Khafizov K, Kotov I, Roev G, Golubeva A, Gasanov G, Korabelnikova M, Turashev A, Cherkashin E, Mironov K, Cherkashina A, Akimkin V. Development and Application of Real-Time PCR-Based Screening for Identification of Omicron SARS-CoV-2 Variant Sublineages. *Genes (Basel)*. 2023, 14(6), 1218
32. Singh L, San JE, Tegally H, Brzoska PM, Anyaneji UJ, Wilkinson E, Clark L, Giandhari J, Pillay S, Lessells RJ, Martin DP, Furtado M, Kiran AM, de Oliveira T. Targeted Sanger sequencing to recover key mutations in SARS-CoV-2 variant genome assemblies produced by next-generation sequencing. *Microb Genom*. 2022;;8(3), 000774
33. World Health Organization Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. Geneva: World Health Organization; 2021
34. Supasa P, Zhou D, Dejnirattisai W, Liu C, Mentzer AJ, Ginn HM, Zhao Y, Duyvesteyn HME, Nutalai R, Tuekprakhon A, Wang B, Paesen GC, Slon-Campos J, López-Camacho C, Hallis B, Coombes N, Bewley KR, Charlton S, Walter TS, Barnes E, Dunachie SJ, Skelly D, Lumley SF, Baker N, Shaik I, Humphries HE, Godwin K, Gent N, Sienkiewicz A, Dold C, Levin R, Dong T, Pollard AJ, Knight JC, Klennerman P, Crook D, Lambe T, Clutterbuck E, Bibi S, Flaxman A, Bittaye M, Belij-Rammerstorfer S, Gilbert S, Hall DR, Williams MA, Paterson NG, James W, Carroll MW, Fry EE, Mongkolsapaya J, Ren J, Stuart DI, Screaton GR. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell*. 2021, 184(8), 2201-2211.e7
35. Li X, Zhang L, Chen S, Ji W, Li C, Ren L. Recent progress on the mutations of SARS-CoV-2 spike protein and suggestions for prevention and controlling of the pandemic. *Infect Genet Evol*. 2021, 93, 104971
36. Majumdar P, Niyogi S. SARS-CoV-2 mutations: the biological trackway towards viral fitness. *Epidemiol Infect*. 2021, 149:e110
37. Meng B, Kemp SA, Papa G, Datir R, Ferreira IATM, Marelli S, Harvey WT, Lytras S, Mohamed A, Gallo G, Thakur N, Collier DA, Mlcochova P; COVID-19 Genomics UK (COG-UK) Consortium; Duncan LM, Carabelli AM, Kenyon JC, Lever AM, De Marco A, Saliba C, Culap K, Cameroni E, Matheson NJ, Piccoli L, Corti D, James LC, Robertson DL, Bailey D, Gupta RK. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep*. 2021, 35(13), 109292
38. Weng S, Zhou H, Ji C, Li L, Han N, Yang R, Shang J, Wu A. Conserved Pattern and Potential Role of Recurrent Deletions in SARS-CoV-2 Evolution. *Microbiol Spectr*. 2022, 10(2):e0219121
39. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, Duprex WP. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science*. 2021, 371(6534), 1139-1142
40. Davies NG et al. Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. *medRxiv*. 2020
41. Liu H, Yuan M, Huang D, Bangaru S, Zhao F, Lee CD, et al. A combination of cross-neutralizing antibodies synergizes to prevent SARS-CoV-2 and SARS-CoV pseudovirus infection. *Cell Host Microbe* 2021, 29, 806-818
42. Khetran SR, Mustafa R. Mutations of SARS-CoV-2 Structural Proteins in the Alpha, Beta, Gamma, and Delta Variants: Bioinformatics Analysis. *JMIR Bioinform Biotech* 2023, 4:e43906

43. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020, 579(7798), 270-273
44. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI; Sheffield COVID-19 Genomics Group; McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020, 182(4), 812-827.e19
45. Lubinski B, Fernandes MHV, Frazier L, Tang T, Daniel S, Diel DG, Jaimes JA, Whittaker GR. Functional evaluation of the P681H mutation on the proteolytic activation the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *bioRxiv [Preprint]*. 2021, 04.06.438731
46. Dhawan M, Sharma A, Priyanka, Thakur N, Rajkhowa TK, Choudhary OP. Delta variant (B.1.617.2) of SARS-CoV-2: Mutations, impact, challenges and possible solutions. *Hum Vaccin Immunother*. 2022, 18(5), 2068883
47. Rahman FI, Ether SA, Islam MR. The "Delta Plus" COVID-19 variant has evolved to become the next potential variant of concern: mutation history and measures of prevention. *J Basic Clin Physiol Pharmacol*. 2021, 33(1), 109-112
48. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. *J Med Virol*. 2022, 94(4), 1641-1649
49. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, Prot M, Gallais F, Gantner P, Velay A, Le Guen J, Kassis-Chikhani N, Edriss D, Belec L, Seve A, Courtellemont L, Péré H, Hocqueloux L, Fafi-Kremer S, Prazuck T, Mouquet H, Bruel T, Simon-Lorière E, Rey FA, Schwartz O. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*. 2021, 596(7871), 276-280
50. Mlcochova P, et al. SARS-CoV-2 B.1.617.2 delta variant replication and immune evasion. *Nature*. 2021, 599(7883), 114-9
51. Abavisani M, Rahimian K, Mahdavi B, Tokhanbigli S, Mollapour Siasakht M, Farhadi A, Kodori M, Mahmanzar M, Meshkat Z. Mutations in SARS-CoV-2 structural proteins: a global analysis. *Virol J*. 2022, 19(1), 220
52. Periwal N, Rathod SB, Sarma S, Johar GS, Jain A, Barnwal RP, Srivastava KR, Kaur B, Arora P, Sood V. Time Series Analysis of SARS-CoV-2 Genomes and Correlations among Highly Prevalent Mutations. *Microbiol Spectr*. 2022, 10(5), e0121922
53. 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.e10
54. Hossain MS, Pathan AQMSU, Islam MN, Tonmoy MIQ, Rakib MI, Munim MA, Saha O, Fariha A, Reza HA, Roy M, Bahadur NM, Rahaman MM. Genome-wide identification and prediction of SARS-CoV-2 mutations show an abundance of variants: Integrated study of bioinformatics and deep neural learning. *Inform Med Unlocked*. 2021, 27, 100798
55. Subissi L, Imbert I, Ferron F, Collet A, Coutard B, Decroly E, Canard B. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. *Antiviral Res*. 2014, 101, 122-30
56. Haddad D, John SE, Mohammad A, Hammad MM, Hebbar P, Channanath A, Nizam R, Al-Qabandi S, Al Madhoun A, Alshukry A, Ali H, Thanaraj TA, Al-Mulla F. SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLoS One*. 2021, 16(5):e0251368
57. Anand Archana, Chenghua Long, Kartik Chandran. Analysis of SARS-CoV-2 amino acid mutations in New York City Metropolitan wastewater (2020-2022) reveals multiple traits with human health implications across the genome and environment-specific distinctions. *medRxiv* 2022, 07.15.22277689
58. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. 2020, 368(6492), 779-782

59. Ferreira IATM, Kemp SA, Datir R, Saito A, Meng B, Rakshit P, Takaori-Kondo A, Kosugi Y, Uriu K, Kimura I, Shirakawa K, Abdullahi A, Agarwal A, Ozono S, Tokunaga K, Sato K, Gupta RK; CITIID-NIHR BioResource COVID-19 Collaboration, Indian SARS-CoV-2 Genomics Consortium; Genotype to Phenotype Japan (G2P-Japan) Consortium. SARS-CoV-2 B.1.617 Mutations L452R and E484Q Are Not Synergistic for Antibody Evasion. *J Infect Dis.* 2021, 224(6), 989-994
60. Sharma, Milan New 'Delta Plus' variant of SARS-CoV-2 identified; here's what we know so far. *India Today.* 2023

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.