



Article

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

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

- Research Institute for Biological Safety Problems, National Holding QazBioPharm, LLP, Guardeyskiy uts 080409, Kazakhstan; k.sultankulova@biosafety.kz (K.T.S.); y.burashev@biosafety.kz (Y.B.); a.melisbek@biosafety.kz (A.M.); m.shirinbekov@biosafety.kz (M.S.); b.myrzakhmetova@biosafety.kz (B.S.M.); a.kerimbayev@biosafety.kz (A.K.); s.nurabayev@biosafety.kz (S.N.); o.chervyakova@biosafety.kz (O.C.); n.kozhabergenov@biosafety.kz (N.K.); l.kutumbetov@biosafety.kz (L.B.K.)
- Institute of Biotechnology, National Academy of Science of Kyrgyzstan, Bishkek 720071, Kyrgyzstan
- Scientific Research Institute of Biology and Biotechnology Problems, al-Farabi Kazakh National University, Almaty 050040, Kazakhstan; smekenovizat@gmail.com
- * Correspondence: b.usserbayev@biosafety.kz

Abstract: Prompt determination of the etiological agent is important in an outbreak of pathogens with pandemic potential, particularly for 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/B1.1/2021, 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/B1.1/2021 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. The variations C241T, F106F, P314L, and D614G found in the 5' UTR, ORF1ab, and S regions were common to all three studied samples, respectively. According to PROVEAN data, the loss-of-function mutations identified in strains KAZ/Britain/2021, KAZ/B1.1/2021, 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/B1.1/2021, 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: genome sequencing; Sanger method; mutation; phylogenetic analysis; COVID-19



Academic Editor: Marta Giovanetti

Received: 30 January 2025 Revised: 6 March 2025 Accepted: 11 March 2025 Published: 14 March 2025

Citation: Usserbayev, B.; Sultankulova, K.T.; Burashev, Y.; Melisbek, A.; Shirinbekov, M.; Myrzakhmetova, B.S.; Zhunushov, A.; Smekenov, I.; Kerimbaev, A.; Nurabaev, S.; et al. Genetic Variations of Three Kazakhstan Strains of the SARS-CoV-2 Virus. *Viruses* 2025, 17, 415. https://doi.org/10.3390/v17030415

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

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 23 February 2025, a total of 777,519,152 confirmed cases have been reported, of which 7,090,776 have resulted in death [2]. The first cases of COVID-19 in the Republic of Kazakhstan were reported on 13 March 2020 [3,4].

The SARS-CoV-2 virion is spherical or ellipsoidal, with an average diameter ranging from 60 to 140 nanometers [5]. The SARS-CoV-2 virus genome consists of ~29.9 kb and is organized in the following order from 5' to 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 [6]

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 [7]. 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), and Omicron (B.1.1.529)) [8,9], having differences in transmission characteristics, ability to cause severe disease, and ability to evade immune response [10].

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 [11]. 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 [12]. 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 [13]. 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 [14] as well as for the creation of diagnostic tests and vaccines against this virus [15].

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,

Viruses 2025, 17, 415 3 of 18

CA, USA) on a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, 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, CA, USA) in a Mastercycler X50 s thermal cycler at the following conditions: 25 $^{\circ}$ C for 10 min; 42 $^{\circ}$ C for 60 min; 85 $^{\circ}$ 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) [16]. The specificity of the primers was checked using NCBI Primer-BLAST [17]. 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 [18]. Oligonucleotides were synthesized on an automatic DNA/RNA Synthesizer H-16 oligonucleotide synthesizer (K&A Labs GmbH, Schaafheim, Germany) using the phosphoramidite method 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 X50 s 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 (TopVision Agarose, Thermo Fisher Scientific Baltics, UAB, Vilnius, Lithuania) stained with ethidium bromide in Tris-acetate buffer at a voltage of 100 volts/cm of gel length for 30 min. 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., Ha-Satat St, Modi'in-Maccabim-Re'ut, 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, CA, 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

Viruses 2025, 17, 415 4 of 18

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, CA, 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 [19]. 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 [20].

2.8. Analysis of Non-Synonymous Mutation Function

PROVEAN v1.1 software 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 [21].

2.9. Phylogenetic Analysis of Nucleotide Sequences

Evolutionary analysis was performed in MEGA 11 [22]. A phylogenetic tree including three samples, a reference genome, and genomes of different SARS-CoV-2 lineages was constructed using the Neighbor-Joining method [23]. The percentage of replicates in which related taxa were grouped together in a bootstrap test (1000 replicates) was shown next to the branches [24]. 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 [25] 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 [18] for the complete SARS-CoV-2 virus genome by PCR according to the manufacturer's protocol described above. Figures 1–3 show the results of the electropherogram using the developed 65 pairs of sequencing primers.

As can be seen in Figures 1–3, fragments of the complete genome of the SARS-CoV-2 virus samples were obtained using the developed sequencing primers [18]. 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.

Viruses **2025**, 17, 415 5 of 18

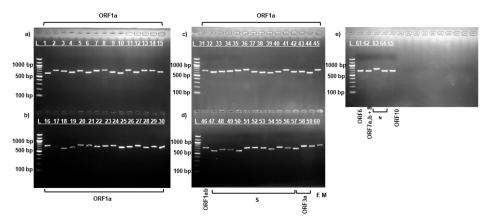


Figure 1. Electrophoresis of amplified fragments of all genes of the SARS-CoV-2/human/KAZ/Britain/2021 strain. Upper and lower gel (**a-d**): lanes 1 to 46—*ORF1ab gene*; lower gel (**d**): Lanes 47 to 56—*S gene*; Lanes 57 to 58—*ORF3a gene*; Lane 59—*E gene*; Lane 60—*M gene*; upper gel (**e**): Lane 61—*ORF6 gene*; Lane 62—*ORF7a, ORF7b,* and *ORF8*; Lanes 63 to 64—*N gene*; Lane 65—*ORF10 gene*; Lane M—100 bp DNA marker.

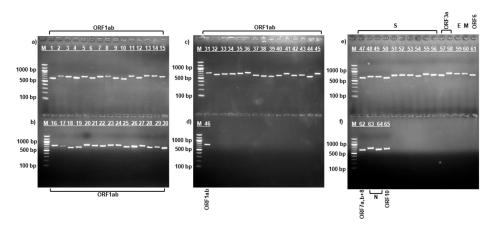


Figure 2. Electrophoresis of amplified fragments of the entire gene of the SARS-CoV-2/human/KAZ/B1.1/2021 strain. Upper and lower gel (**a**-**d**): Lanes 1 to 46—*ORF1ab gene*; upper gel (**e**): Lanes 47 to 56—*S gene*; Lanes 57 to 58—*ORF3a gene*; Lane 59—*E gene*; Lane 60—*M gene*; Lane 61—*ORF6 gene*; lower gel (**f**): Lane 62—*ORF7a, ORF7b,* and *ORF8 genes*; Lanes 63 to 64—*N gene*; Lane 65—*ORF10 gene*; Lane M—100 bp DNA marker.

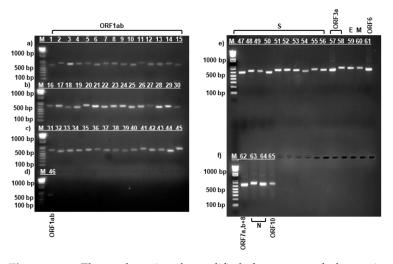


Figure 3. Electrophoresis of amplified fragments of the entire gene of the SARS-CoV-2/human/KAZ/Delta020/2021 strain. Upper and lower gel (**a**–**d**): lanes 1 to 46—*ORF1ab gene*; upper gel (**e**): lanes 47 to 56—*S gene*; lanes 57 to 58—*ORF3a gene*; lane 59—*E gene*; lane 60—*M gene*; lane 61—*ORF6 gene*; lower gel (**f**): lane 62—*ORF7a*, *ORF7b* and *ORF8 genes*; lanes 63 to 64—*N gene*; lane 65—*ORF10 gene*; lane M—100 bp DNA marker.

Viruses 2025, 17, 415 6 of 18

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 contents were 38%, 37.95%, and 38%, respectively [4,26,27].

The nucleotide sequences of Kazakhstan SARS-CoV-2 virus strains were analyzed using the Pangolin COVID-19 database [19]. 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 [20] 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. Mutations 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) [16].

	Data for Strain:										
Protein	Wuha Position	an-Hu-1 ^a Variant	KAZ/Br Variant	itain/2021 Position		1.1/2021 Position		elta020/2021 Position	Type of Mutation	Amino Acid Change	
	106	С	_	_	T	29	-	_	_	106	
5′ UTR b	210	G	_	_	_	_	T	184	_	210	
0 0111	241	Č	T	215	T	164	T	215	-	241	
	344	С	-	-	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	Č	T	2084	_	_	_	_	SNP_silent	N435N	
	2530	A	_	_	G	2453	_	_	SNP_silent	E575E	
	3037	Č	T	3011	Ť	2960	T	3011	SNP_silent	F106F	
	3267	Č	Ť	3241	_	_	_	_	SNP	T183I	
	4181	C G	_	_	_	_	T	4155	SNP	A488S	
	4449	Č	_	_	A	4372	_	_	SNP	T577N	
	4455	C C	_	_	Ť	4378	_	_	SNP	A579V	
	4475	č	_	_	Ť	4398	_	_	SNP	R586C	
	5388	Č	A	5362	_	-	_	_	SNP	A890D	
	5829	Ä	_	-	С	5752	_	_	SNP	K1037T	
	5986	Č	T	5960	_	-	_	_	SNP_silent	F1089F	
	6402	č	_	-	_	_	T	6376	SNP	P1228L	
	6954	Ť	С	6928	_	_	_	-	SNP	I1412T	
	7042	Ġ	Ť	7016	_	_	_	_	SNP	M1441I	
	7124	Č	_	7010	_	_	T	7098	SNP	P1469S	
ORF1ab	8986	Č	_	_	_	_	Ť	8960	SNP_silent	D144D	
	9053	G	_	_	_	_	Ť	9027	SNP	V167L	
	9749	A	_	_	Ğ	9672	_	-	SNP	K399E	
	9867	Ť	_	_	Ğ	9790	_	_	SNP	L438R	
	10,029	Č	_	_	-	9790 —	T	10,003	SNP	T492I	
	10,029	C	_	_	T	10,121	_	10,003	SNP_silent	D48D	
	11,195	C	T	11,169	_	10,121	_	_	SNP SNP	L75F	
	11,193	A	_	11,109	_	_	Ğ	_ 11,175	SNP	T77A	
	11,201	TCTGGTTTT	del	11,261	- del	11,210		11,173	SNP_stop	S106	
				11,201	aei	11,210	- G	11 206	SINT_Stop	V120V	
	11,332	A	_ T					11,306	SNP_silent		
	14,120	C	Ţ	14,085	T	14 222	_ T	14.202	SNP	P218L	
	14,408	Č	T	14,373	T	14,322		14,382	SNP	P314L	
	14,676	Ċ	T	14,641	_ T	14.021	-	_	SNP_silent	P403P	
	15,017	Ċ	_	_ 15.044	T	14,931	_	_	SNP	A517V	
	15,279	C	T	15,244	_	_	_	_ 15 405	SNP_silent	H604H	
	15,451	G	_	-	-	_	Α	15,425	SNP	G662S	
	16,176	Ť	C	16,141	-	_	_	-	SNP	T903T	
	16,466	C	_	_	-	_	Ţ	16,440	SNP	P77L	
	18,271	G	_	_	-	_	A	18,245	SNP	E78K	
	18,337	G	_	_	_	_	T	18,311	SNP	A100S	

Viruses **2025**, 17, 415 7 of 18

Table 1. Cont.

Protein	Data for Strain:										
	Wuhan-Hu-1 ^a Position Variant		KAZ/Britain/2021 Variant Position		KAZ/B1.1/2021 Variant Position		KAZ/Delta020/2021 Variant Position		Type of Mutation	Acid Change	
ORF1ab	19,220	С	_	_	_	_	T	19,194	SNP	A394V	
	20,405	C	T	20,370	_	_	_	_	SNP	P262L	
	20,759	C	_	_	T	20,673	_	_	SNP	A34V	
	21,080	A	_	_	G	20,994	_	_	SNP	K141R	
	21,215	A	G	21,180	_	_	_	_	SNP	H186R	
	21,446	A	-	_	G	21,360	-	_	SNP	K263R	
	27,389	С	_	_	T	27,303	_	_	_	27,389	
3′ UTR	29,733	_	_	_	TA	29,648	_	_	_	29,733	
	29,742	G	_	_	_	-	T	29,716	_	29,742	
	29,755	_	_	_	C	29,672	_	-	_	29,755	
	29,790	_	_	_	T	29,708	_	_	_	29,790	

^a Severe acute respiratory syndrome coronavirus 2 strain Wuhan-Hu-1, complete genome sequence (GenBank accession number NC_045512) [16]. ^b UTR, untranslated region. ^c K81N, the K-to-N change at Position 81.

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) [16].

	Data for Strain:										
Protein		n-Hu-1 ^a	KAZ/Bri	tain/2021	KAZ/B	1.1/2021	KAZ/Delta020/2021				
	Position	Variant	Variant	Position	Variant	Position	Variant	Position			
	21,618	С	_	_	_	_	G	21592			
	21,646	C	_	_	T	21,560	_	_			
	21,648	C	_	_	T	21,562	_	_			
	21,765	TACATG	del	21,729	_	· <u> </u>	_	_			
	21,766	A	_	_	_	_	del	21,739			
	21,784	T	_	_	A	21,698	_	_			
	21,789	T C C	_	_	T	21,703	_	_			
	21,846	C	_	_	T	21,760	_	_			
	21,987	Ğ	_	_	_	_	A	21,961			
	21,993	ATT	del	21,951	_	_	_	_			
	22,185	C	_	_	_	_	T	22,159			
	22,407	A	_	_	_	_	T	22,381			
	22,917	T	_	_	_	_	G	22,891			
S	22,995	C	_	_	_	_	A	22,969			
3	23,014	A	_	_	C	22,928	_	_			
	23,063	A	T	23,019	_	_	_	_			
	23,271	C	A	32,227	_	_	_	_			
	23,403	A	G	23,359	G	23,317	G	23,377			
	23,520	C	-	_	T	23,434	_	_			
	23,604	Č	A	23,560	_	_	G	23,578			
	23,709	C C C	T	23,665	_	_	_	_			
	23,751	C	_	_	T	23,665	_	_			
	23,997	С	_	_	T	23,911	_	_			
	24,000	G	_	_	T	23,914	_	_			
	24,410	G	_		_	_	Α	24,384			
	24,506	T	G	24,462	_	–	_	_			
	24,538	A	_	_	T	24,452	_	_			
	24,914	G	C	24,870	_	_	_	_			

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

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/B1.1/2021, 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 14,408 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

Viruses 2025, 17, 415 8 of 18

at position 314 (P314L), respectively. A deletion of one amino acid residue S106 (serine \rightarrow deletion) was observed in two samples studied (KAZ/Britain/2021 and KAZ/B1.1/2021) in the *NSP6* region.

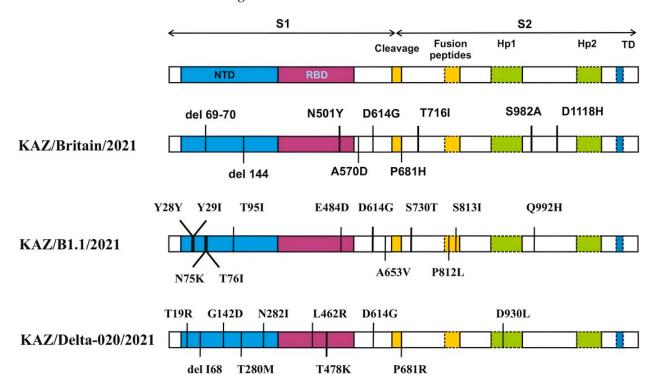


Figure 4. Amino acid changes in the S protein of the studied strains. Description of the SARS-CoV-2 spike mutation in different virus isolates: KAZ/Britain/2021, KAZ/B1.1/2021 and KAZ/Delta020/2021. Colored bars describe the structural domain of the spike protein. NTD—N-terminal domain (blue); RBD—receptor-binding domain (dark purple); Fusion peptides—(orange colour); Hp1—thermal protein 1 (medium spring green); Hp2—thermal protein 2 (medium spring green). TD—transmembrane domain (blue).

Table 2 and Figure 4 show 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) [16].

	Data for Strain:									
Protein	Wuhan- Position		KAZ/Brit Variant	ain/2021 Position	KAZ/B1 Variant		KAZ/Del Variant	ta020/2021 Position	Type of Mutation	Amino Acid Change
ORF3a	25,459 25,688 25,838 26,110	C C G C	– – T –	- 25,794 -	– T – T	25,602 - 26,024	T - - -	25,443 - - -	SNP SNP SNP SNP	S26L ^b A99V W149L P240S
M	26,767 26,895 27,008	T C G	- - -	- - -	– T T	- 26,809 26,922	C - -	26,741 - -	SNP SNP SNP	I82T H125Y K162N
ORF6	27,281 27,285	GG TC	AA AT	27,237 27,241	_ _	-	-	_ _	SNP_stop SNP	W27 NL28KF
ORF7a	27,527 27,638 27,630 27,667 27,739 27,752	C T C G C	- - - - -	- - - - -	- T A T	- 27,544 27,581 27,653	T C - - T	27,501 27,612 - - - 27,726	SNP SNP SNP_silent SNP SNP SNP	P45L V82A A79A E92K L116F T120I
ORF7b	27,874	С	-	_	_	_	T	27,848	SNP	T40I
ORF8	27,919 27,972 28,048 28,095 28,111 28,251 28,253 28,255 28,258	T C G A A T C T A	- T T T G - -	27,928 28,004 28,051 28,067 - - -	- - - - - -	- - - - - - -	C - - - C A A G	27,893 - - - 28,225 28,227 28,229 28,232	SNP SNP_stop SNP_stop SNP_SNP SNP SNP SNP	I9T Q27 R52I K68 Y73C F120L F120L I121N 122 *
N	28,280 28,461 28,881 28,881 28,916 28,977 29,236 29,402 29,436	GAT A GGG G C C C G A	CTA - AAC - T	28,236 28,837 - - 28,933 - - -	- AAC - - - - T	28,837 - - - - - 29,350	- G - T T - T T	28,435 - 28,855 28,890 - 29,210 29,376	SNP	D3L D63G RG203KR R203M G215C S235F G312G D377Y K388I

 $^{^{\}rm a}$ Severe acute respiratory syndrome coronavirus 2 strain Wuhan-Hu-1, complete genome sequence (GenBank accession number NC_045512). $^{\rm b}$ S26L, the S-to-L change at Position 26. 122 *, the studied strain according to the Pangolin database belongs to the AY.122 lineage.

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/B1.1/2021 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/B1.1/2021 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/B1.1/2021 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/B1.1/2021 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

Viruses 2025, 17, 415 10 of 18

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) [28].

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

		KAZ/Britain/2	021		KAZ/B1.1/202	21		KAZ/Delta020/	2021
Protein	Amino Acid Change	PROVEAN Assessment	The Effect of Variation on Protein	Amino Acid Change	PROVEAN Assessment	The Effect of Variation on Protein	Amino Acid Change	PROVEAN Assessment	The Effect of Variation on Protein
ORF1ab	I295L T183I T577N A890D	0.232 0.216 0.240 -1.749	Neutral Neutral Neutral Neutral	L27F A579V R586C K1037T	-0.047 0.011 -0.727 -1.196	Neutral Neutral Neutral Neutral	K81N R365L A488S P1228L	-0.070 -0.939 -0.061 -1.038	Neutral Neutral Neutral Neutral
	I1412T M1441I L75F P218L P314L P262L	$\begin{array}{c} -0.370 \\ 0.263 \\ -2.290 \\ -5.021 \\ -0.446 \\ -0.014 \end{array}$	Neutral Neutral Neutral Deleterious Neutral Neutral	K399E L438R P314L A517V A34V K141R	$\begin{array}{c} -1.877 \\ 0.659 \\ -0.446 \\ -1.291 \\ 1.158 \\ -0.221 \end{array}$	Neutral Neutral Neutral Neutral Neutral Neutral	P1469S V167L T492I T77A P314L G662S	0.338 -0.696 1.435 -0.878 -0.446 -2.475	Neutral Neutral Neutral Neutral Neutral Neutral
	H186R - - -	-0.267 - - -	Neutral - - -	K263R - -	-1.344 - -	Neutral – –	P77L E78K A100S A394V	-6.845 -1.123 1.338 -1.523	Deleterious Neutral Neutral Neutral
	H69del Y145del N501Y A570D D614G	0.260 0.853 -0.090 -0.682 0.598	Neutral Neutral Neutral Neutral Neutral	Y28Y T29I N74K T76I T95I	0.000 -1.538 -1.309 -0.115 -1.214	Neutral Neutral Neutral Neutral Neutral	T19R I68del G142D T208M N282I	-0.839 -0.821 -0.277 -0.314 -3.717	Neutral Neutral Neutral Neutral Deleterious
S	P681H T716I S982A D1118H	0.060 -3.293 -1.505 -1.142	Neutral Neutral Deleterious Neutral Neutral	E484D D614G A653V S730T P812L S813I	-0.210 0.598 -0.715 -0.040 -0.868 -2.867	Neutral Neutral Neutral Neutral Neutral Neutral Deleterious	L452R T478K D614G P681R D950N	0.559 -0.524 0.598 0.741 -1.631	Neutral Neutral Neutral Neutral Neutral
				Q992H	-4.059	Deleterious	<u> </u>	_	
ORF3a	– W149L	- -9.419	– Deleterious	A99V	-1.962 -	Neutral	S26L - -	-2.314 - -	Neutral – –
	_	_		P240S	-1.495	Neutra	- I82T	-3.853	 Deleterious
Μ	- - -	_ _ _	- - -	H125Y K162N	0.799 0.501	Neutral Neutral	- -	-3.633 - -	– – –
	-	<u>-</u> -	<u>-</u> -				P45L V82A	-10.000 -2.667	Deleterious Deleterious
ORF7a	- - -	- - -	- - -	A79A E92K L116F –	$0.000 \\ -1.842 \\ -1.263$	Neutral Neutral Neutral	– – – T120I	- - - -1.789	– – – Neutral
ORF7b							T40I	-2.000	Neutral
				_	_	_	I9T	-1.333	Neutral
ORF8	R52I Y73C	$ \begin{array}{r} -6.417 \\ -4.500 \end{array} $	Deleterious Deleterious	- - -	_ _ _	- - -	– – F120L	- - -2.667	– – Deleterious
	-			-			F120L I121N	-2.667 -0.667	Deleterious Neutral
	D3L	-0.230	Neutral	_		_	D63G	-0.929	Neutral
N	- - -	- - -	_ _ _ _	- - -	- - -	_ _ _ _	R203M G215C S235F	-0.929 -3.304 -0.953 -1.738	Deleterious Neutral Neutral
	- -	<u>-</u> -	- -	- K388I	_ -1.204	– Neutral	D377Y -	-1.779 -	Neutral –

The 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/B1.1/2021, 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.

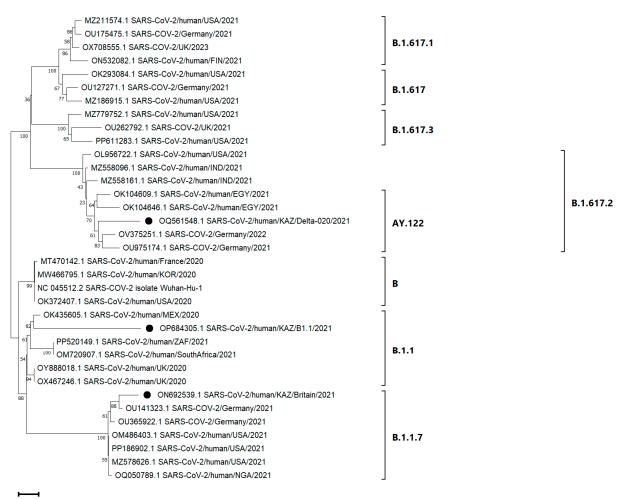


Figure 5. Phylogenetic analysis of the studied strains (black circle) of SARS-CoV-2 and 35 global strains belonging to different virus lineages, such as B.1.617.1, B.1.617.2, B.1.617, B.1.617.3, AY.122, B.1.1.7, B.1.1, and B, which were obtained from the NCBI GenBank database. Here, the x-axis represents 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 [29]. Their transmission from animal to human and from human to human depends on many factors, including potential adaptive evolution to virulent strains [30].

RNA viruses are characterized by higher replication fidelity ($\sim 10^{-4}$ error/site/cycle) and genetically diverse RNA polymerases [31]. 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 [32]. 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 [33,34], 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 [35].

The rapid evolution of the SARS-CoV-2 genome highlights the need to develop antiviral drugs against the virus [36]. 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 [37].

Sanger sequencing is considered the most optimal method for sequencing short fragments (<1000 bp) and is useful for filling gaps in partial whole genomes [38,39]. 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 [32].

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 [18]. 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 $^{\circ}$ C. The PCR products of the studied samples were obtained using the designed sequencing primers.

The length of the amplicons obtained (Figures 1–3) 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 [19]. 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 [40,41]. The mutations identified in the

Viruses 2025, 17, 415 13 of 18

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 [42-44]. 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 [44,45]. Some studies describe H69-V70, N501Y, and P681H, which may affect viral infectivity [42,46]. N501Y increases viral infectivity by 70-80% and enhances the binding affinity of the viral S protein to human ACE2 [42,46–48]. 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 [48,49]. D614G was found in all three isolates and has become the most common mutation among SARS-CoV-2 variants during the global pandemic [50]. Lubinski B. et al. showed that P681H can increase its cleavage by furin-like proteases, although this process does not lead to viral entry [51]. According to Pangolin COVID-19 data, strain KAZ/B1.1/2021 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 [52–54].

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 SARS-CoV-2 Lineage Tree, 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 [55–58]. 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 [59,60].

The resulting nucleotide sequences were tested using the COVID-19 Genome Annotator [20] 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 (Tables 1–3 and Figure 4). 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 [61]. 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 [12,62,63]. 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 [12,63]. The missense mutation P314L, found in the *NSP12* region of the *ORF1b* gene, is considered to be part of the core

Viruses 2025, 17, 415 14 of 18

replication/transcription complex and is a conserved protein in coronaviruses [6,64]. Thus, the P314L mutation affects SARS-CoV-2 RNA replication by participating in the activity of *RdRp* (*RNA-dependent RNA polymerase*) [65,66]. In addition, *RdRp* plays an important role in the process of SARS-CoV-2 viral replication and transcription [67]. D614G was detected in all three isolates and has become the most common mutation among SARS-CoV-2 variants during the global pandemic [50].

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 [68]. In mid-June 2021, a mutated Delta variant (B.1.617.2), known as the Delta plus, was identified in India [69].

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

Funding: The work was carried out within the framework of the scientific and technical program on the project "Development of a vaccine against coronavirus infection COVID-19" (IRN No.64356/PTsF-MON-RK-OT-20) with the support of the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan.

Institutional Review Board Statement: Ethical review and approval were not required for this study, as it did not involve human participants. The research was conducted using biological material obtained from the Scientific and Practical Centre for Sanitary-Epidemiological Expertise and Monitoring, a branch of the National Centre for Public Health, a republican state enterprise under the right of economic management of the Ministry of Health of the Republic of Kazakhstan, which did not involve direct human interaction.

Viruses 2025, 17, 415 15 of 18

Informed Consent Statement: Informed consent was not required for this study because the samples were obtained from the Scientific and Practical Centre for Sanitary-Epidemiological Expertise and Monitoring, a branch of the National Centre for Public Health, a republican state enterprise under the right of economic management of the Ministry of Health of the Republic of Kazakhstan. The researchers had no direct contact with people and did not collect samples from them.

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 conflicts of interest.

References

- 1. Shereen, M.A.; 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. [CrossRef] [PubMed]
- 2. World Health Organization. COVID-19 Dashboard. Available online: https://covid19.who.int/ (accessed on 6 March 2025).
- 3. Zhugunissov, K.; Zakarya, K.; Khairullin, B.; Orynbayev, M.; Abduraimov, Y.; Kassenov, M.; Sultankulova, K.; Kerimbayev, A.; Nurabayev, S.; Myrzakhmetova, B.; et al. Development of the Inactivated QazCovid-in Vaccine: Protective Efficacy of the Vaccine in Syrian Hamsters. *Front. Microbiol.* **2021**, *12*, 720437. [CrossRef] [PubMed]
- 4. Usserbayev, B.; Zakarya, K.; Kutumbetov, L.; Orynbayev, M.; Sultankulova, K.; Abduraimov, Y.; Myrzakhmetova, B.; Zhugunissov, K.; Kerimbayev, A.; Melisbek, A.; et al. Near-complete genome sequence of a SARS-CoV-2 variant B. 1.1. 7 virus strain isolated in Kazakhstan. *Microbiol. Resour. Announc.* 2022, 11, e0061922. [CrossRef]
- 5. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733. [CrossRef]
- 6. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef]
- 7. Cosar, B.; Karagulleoglu, Z.Y.; Unal, S.; Ince, A.T.; Uncuoglu, D.B.; Tuncer, G.; Kilinc, B.R.; Ozkan, Y.E.; Ozkoc, H.C.; Demir, I.N.; et al. SARS-CoV-2 Mutations and their Viral Variants. *Cytokine Growth Factor. Rev.* **2022**, *63*, 10–22. [CrossRef] [PubMed]
- 8. Safari, I.; Elahi, E. Evolution of the SARS-CoV-2 genome and emergence of variants of concern. *Arch. Virol.* **2022**, *167*, 293–305. [CrossRef]
- 9. Andre, M.; Lau, L.S.; Pokharel, M.D.; Ramelow, J.; Owens, F.; Souchak, J.; Akkaoui, J.; Ales, E.; Brown, H.; Shil, R.; et al. From alpha to omicron: How different variants of concern of the SARS-Coronavirus-2 impacted the world. *Biology* **2023**, *12*, 1267. [CrossRef]
- 10. Bhardwaj, P.; Mishra, S.K.; Behera, S.P.; 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.
- 11. Vidanović, D.; Tešović, B.; Volkening, J.D.; Afonso, C.L.; Quick, J.; Šekler, M.; Knežević, A.; Janković, M.; Jovanović, T.; Petrović, T.; et al. First whole-genome analysis of the novel coronavirus (SARS-CoV-2) obtained from COVID-19 patients from five districts in Western Serbia. *Epidemiol Infect* **2021**, *149*, e246. [CrossRef]
- 12. Mercatelli, D.; Giorgi, F.M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* **2020**, *11*, 1800. [CrossRef]
- 13. LaTourrette, K.; Garcia-Ruiz, H. Determinants of Virus Variation, Evolution, and Host Adaptation. *Pathogens* **2022**, *11*, 1039. [CrossRef]
- 14. Márquez, S.; Prado-Vivar, B.; Guadalupe, J.J.; Gutierrez, B.; Jibaja, M.; Tobar, M.; Mora, F.; Gaviria, J.; García, M.; Espinosa, F.; et al. Genome sequencing of the first SARS-CoV-2 reported from patients with COVID-19 in Ecuador. *medRxiv* **2020**. [CrossRef]
- 15. Mohammadi, E.; Shafiee, F.; Shahzamani, K.; Ranjbar, M.M.; Alibakhshi, A.; Ahangarzadeh, S.; Beikmohammadi, L.; Shariati, L.; Hooshmandi, S.; Ataei, B.; et al. Novel and emerging mutations of SARS-CoV-2: Biomedical implications. *Biomed. Pharmacother.* **2021**, *139*, 111599. [CrossRef]
- 16. National Center for Biotechnology Information (NCBI). SARS-CoV-2 Reference Genome. Available online: https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2 (accessed on 6 March 2025).
- 17. National Center for Biotechnology Information (NCBI). BLAST: Basic Local Alignment Search Tool. Available online: https://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed on 6 March 2025).

Viruses 2025, 17, 415 16 of 18

18. Burashev, Y. Primers for Whole Genome Sequencing of the Sars-Cov-2 Virus. Zenodo. 2022. Available online: https://zenodo.org/records/7264509 (accessed on 6 March 2025).

- 19. O'Toole, Á.; Scher, E.; Underwood, A.; Jackson, B.; Hill, V.; McCrone, J.T.; Colquhoun, R.; Ruis, C.; Abu-Dahab, K.; Taylor, B.; et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* 2021, 7, veab064. [CrossRef] [PubMed]
- 20. COVID-19 genome annotator. Available online: http://giorgilab.unibo.it/coronannotator/ (accessed on 6 March 2025).
- 21. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [CrossRef] [PubMed]
- 22. Tamura, K.; Stecher, G.; Kumar, S. MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]
- 23. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 1987, 4, 406–425.
- 24. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1985**, *39*, 783–791. [CrossRef] [PubMed]
- 25. Tamura, K.; Nei, M.; Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11030–11035. [CrossRef]
- 26. Burashev, Y.; Usserbayev, B.; Kutumbetov, L.; Abduraimov, Y.; Kassenov, M.; Kerimbayev, A.; Myrzakhmetova, B.; Melisbek, A.; Shirinbekov, M.; Khaidarov, S.; et al. Coding Complete Genome Sequence of the SARS-CoV-2 Virus Strain, Variant B.1.1, Sampled from Kazakhstan. *Microbiol. Resour. Announc.* 2022, 11, e0111422. [CrossRef]
- 27. Usserbayev, 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, e0030123. [CrossRef] [PubMed]
- 28. Cruz, C.A.K.; Medina, P.M.B. Temporal changes in the accessory protein mutations of SARS-CoV-2 variants and their predicted structural and functional effects. *J. Med. Virol.* **2022**, *94*, 5189–5200. [CrossRef] [PubMed]
- 29. Cleaveland, S.; Laurenson, M.K.; Taylor, L.H. 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*, 991–999. [CrossRef] [PubMed]
- 30. Padhan, K.; Parvez, M.K.; Al-Dosari, M.S. Comparative sequence analysis of SARS-CoV-2 suggests its high transmissibility and pathogenicity. *Future Virol.* **2021**, *16*, 245–254. [CrossRef]
- 31. Parvez, M.K.; Parveen, S. Evolution and Emergence of Pathogenic Viruses: Past, Present, and Future. *Intervirology* **2017**, *60*, 1–7. [CrossRef]
- 32. Lee, S.H. A Routine Sanger Sequencing Target Specific Mutation Assay for SARS-CoV-2 Variants of Concern and Interest. *Viruses* **2021**, *13*, 2386. [CrossRef]
- 33. 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*, 501–511. [CrossRef]
- 34. Li, Y.; Lai, D.Y.; Zhang, H.N.; Jiang, H.W.; Tian, X.; Ma, M.L.; Qi, H.; Meng, Q.F.; Guo, S.J.; Wu, Y.; et al. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol. Immunol.* **2020**, *17*, 1095–1097. [CrossRef]
- 35. Abbasian, M.H.; Mahmanzar, M.; Rahimian, K.; Mahdavi, B.; Tokhanbigli, S.; Moradi, B.; Sisakht, M.M.; Deng, Y. Global landscape of SARS-CoV-2 mutations and conserved regions. *J. Transl. Med.* **2023**, 21, 152. [CrossRef]
- 36. Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; COVID-19 Genomics UK (COG-UK) Consortium; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [CrossRef] [PubMed]
- 37. Esman, A.; Dubodelov, D.; Khafizov, K.; Kotov, I.; Roev, G.; Golubeva, A.; Gasanov, G.; Korabelnikova, M.; Turashev, A.; Cherkashin, E.; et al. Development and Application of Real-Time PCR-Based Screening for Identification of Omicron SARS-CoV-2 Variant Sublineages. *Genes* 2023, 14, 1218. [CrossRef]
- 38. Singh, L.; San, J.E.; Tegally, H.; Brzoska, P.M.; Anyaneji, U.J.; Wilkinson, E.; Clark, L.; Giandhari, J.; Pillay, S.; Lessells, R.J.; et al. Targeted Sanger sequencing to recover key mutations in SARS-CoV-2 variant genome assemblies produced by next-generation sequencing. *Microb. Genom.* 2022, 8, 000774. [CrossRef] [PubMed]
- 39. World Health Organization. *Genomic Sequencing of SARS-CoV-2: A Guide to Implementation for Maximum Impact on Public Health;* World Health Organization: Geneva, Switzerland, 2021.
- 40. Supasa, P.; Zhou, D.; Dejnirattisai, W.; Liu, C.; Mentzer, A.J.; Ginn, H.M.; Zhao, Y.; Duyvesteyn, H.M.E.; Nutalai, R.; Tuekprakhon, A.; et al. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* 2021, 184, 2201–2211.e7. [CrossRef] [PubMed]
- 41. 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. [CrossRef]
- 42. Majumdar, P.; Niyogi, S. SARS-CoV-2 mutations: The biological trackway towards viral fitness. *Epidemiol. Infect.* **2021**, 149, e110. [CrossRef]

Viruses 2025, 17, 415 17 of 18

43. Meng, B.; Kemp, S.A.; Papa, G.; Datir, R.; Ferreira, I.A.T.M.; Marelli, S.; Harvey, W.T.; Lytras, S.; Mohamed, A.; Gallo, G.; et al. 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*, 109292. [CrossRef]

- 44. 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*, e0219121. [CrossRef]
- 45. McCarthy, K.R.; Rennick, L.J.; Nambulli, S.; Robinson-McCarthy, L.R.; Bain, W.G.; Haidar, G.; Duprex, W.P. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* **2021**, *371*, 1139–1142. [CrossRef]
- Davies, N.G.; Abbott, S.; Barnard, R.C.; Jarvis, C.I.; Kucharski, A.J.; Munday, J.; Pearson, C.A.; Russell, T.W.; Tully, D.C.; Washburne, A.D.; et al. Estimated transmissibility and severity of novel SARS-CoV-2 variant of concern 202012/01 in England. medRxiv. 2020. [CrossRef]
- 47. Liu, H.; Yuan, M.; Huang, D.; Bangaru, S.; Zhao, F.; Lee, C.D.; Peng, L.; Barman, S.; Zhu, X.; Nemazee, D.; 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. [CrossRef] [PubMed]
- 48. Khetran, S.R.; 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. [CrossRef] [PubMed]
- 49. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef]
- 50. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182*, 812–827.e19. [CrossRef]
- 51. Lubinski, B.; Fernandes, M.H.V.; Frazier, L.; Tang, T.; Daniel, S.; Diel, D.G.; Jaimes, J.A.; Whittaker, G.R. Functional evaluation of the P681H mutation on the proteolytic activation the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *bioRxiv* **2021**. [CrossRef]
- 52. SARS-CoV-2 Lineage Tree. Available online: https://observablehq.com/embed/6475ff63fc3ebfb3 (accessed on 6 March 2025).
- 53. Rambaut, A.; Holmes, E.C.; O'Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; Du Plessis, L.; Pybus, O.G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403–1407. [CrossRef]
- 54. O'Toole, Á.; Pybus, O.G.; Abram, M.E.; Kelly, E.J.; Rambaut, A. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genom.* **2022**, 23, 121. [CrossRef] [PubMed]
- 55. Dhawan, M.; Sharma, A.; Priyanka Thakur, N.; Rajkhowa, T.K.; Choudhary, O.P. Delta variant (B.1.617.2) of SARS-CoV-2: Mutations, impact, challenges and possible solutions. *Hum. Vaccin. Immunother.* **2022**, *18*, 2068883. [CrossRef]
- 56. Rahman, F.I.; Ether, S.A.; Islam, M.R. 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*, 109–112. [CrossRef]
- 57. Kumar, S.; Thambiraja, T.S.; Karuppanan, K.; Subramaniam, G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. *J. Med. Virol.* **2022**, *94*, 1641–1649. [CrossRef]
- 58. Planas, D.; Veyer, D.; Baidaliuk, A.; Staropoli, I.; Guivel-Benhassine, F.; Rajah, M.M.; Planchais, C.; Porrot, F.; Robillard, N.; Puech, J.; et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **2021**, *596*, 276–280. [CrossRef] [PubMed]
- 59. Mlcochova, P.; Kemp, S.A.; Dhar, M.S.; Papa, G.; Meng, B.; Ferreira, I.A.T.M.; Datir, R.; Collier, D.A.; Albecka, A.; Singh, S.; et al. SARS-CoV-2 B.1.617.2 delta variant replication and immune evasion. *Nature* **2021**, *599*, 114–119. [CrossRef] [PubMed]
- 60. 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*, 220. [CrossRef]
- 61. Periwal, N.; Rathod, S.B.; Sarma, S.; Johar, G.S.; Jain, A.; Barnwal, R.P.; Srivastava, K.R.; Kaur, B.; Arora, P.; Sood, V. Time Series Analysis of SARS-CoV-2 Genomes and Correlations among Highly Prevalent Mutations. *Microbiol. Spectr.* **2022**, *10*, e0121922. [CrossRef]
- 62. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10. [CrossRef]
- 63. Hossain, M.S.; Pathan, A.Q.M.S.U.; Islam, M.N.; Tonmoy, M.I.Q.; Rakib, M.I.; Munim, M.A.; Saha, O.; Fariha, A.; Reza, H.A.; Roy, M.; et al. 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. [CrossRef]
- 64. 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*, 122–130. [CrossRef] [PubMed]
- 65. Haddad, D.; John, S.E.; Mohammad, A.; Hammad, M.M.; Hebbar, P.; Channanath, A.; Nizam, R.; Al-Qabandi, S.; Al Madhoun, A.; Alshukry, A.; et al. SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLoS ONE* **2021**, *16*, e0251368. [CrossRef]

Viruses 2025, 17, 415 18 of 18

66. Archana, A.; Long, C.; Chandran, K. 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. [CrossRef]

- 67. Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **2020**, *368*, 779–782. [CrossRef]
- 68. Ferreira, I.A.T.M.; Kemp, S.A.; Datir, R.; Saito, A.; Meng, B.; Rakshit, P.; Takaori-Kondo, A.; Kosugi, Y.; Uriu, K.; Kimura, I.; et al. SARS-CoV-2 B.1.617 Mutations L452R and E484Q Are Not Synergistic for Antibody Evasion. *J. Infect. Dis.* **2021**, 224, 989–994. [CrossRef] [PubMed]
- 69. Sharma, M. New 'Delta Plus' Variant of SARS-CoV-2 Identified; Here's What We Know So Far. *India Toda*. 2021. Available online: https://www.indiatoday.in/coronavirus-outbreak/story/delta-plus-variant-covid-corona-coronavirus-sarscov2-181 4768-2021-06-14 (accessed on 6 March 2025).

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.