



Article

# Detection and Characterisation of SARS-CoV-2 in Eastern Province of Zambia: A Retrospective Genomic Surveillance Study

Doreen Mainza Shempela <sup>1,\*</sup>, Herman M. Chambaro <sup>2</sup>, Jay Sikalima <sup>1</sup>, Fatim Cham <sup>3</sup>, Michael Njuguna <sup>3</sup>, Linden Morrison <sup>3</sup>, Steward Mudenda <sup>4</sup>, Duncan Chanda <sup>5</sup>, Maisa Kasanga <sup>6</sup>, Victor Daka <sup>7</sup>, Geoffrey Kwenda <sup>8</sup>, Kunda Musonda <sup>9</sup>, Sody Munsaka <sup>8</sup>, Roma Chilengi <sup>9</sup>, Karen Sichinga <sup>1</sup> and Edgar Simulundu <sup>10,11,\*</sup>

- <sup>1</sup> Churches Health Association of Zambia, Lusaka 10101, Zambia; jay.sikalima@chaz.org.zm (J.S.); karen.sichinga@chaz.org.zm (K.S.)
- <sup>2</sup> Virology Unit, Central Veterinary Research Institute, Ministry of Fisheries and Livestock, Lusaka 10101, Zambia; hermcham@gmail.com
- <sup>3</sup> Global Fund to Fight AIDS, Tuberculosis and Malaria (GFATM), 1201 Geneva, Switzerland; fatim.jallow@theglobalfund.org (F.C.); michael.njuguna@theglobalfund.org (M.N.); linden.morrison@theglobalfund.org (L.M.)
- <sup>4</sup> Department of Pharmacy, School of Health Sciences, University of Zambia, Lusaka 10101, Zambia; steward.mudenda@unza.zm
- <sup>5</sup> University Teaching Hospital, Ministry of Health, Lusaka 10101, Zambia; duncanchanda@gmail.com
- <sup>6</sup> Department of Epidemiology and Biostatistics, School of Public Health, Zhengzhou University, Zhengzhou 450001, China; kasangaanita@gmail.com
- <sup>7</sup> Public Health Department, Michael Chilufya Sata School of Medicine, Copperbelt University, Ndola 21692, Zambia; dakavictorm@gmail.com
- <sup>8</sup> Department of Biomedical Sciences, School of Health Sciences, University of Zambia, Lusaka 10101, Zambia; kwenda.geoffrey@unza.zm (G.K.); s.munsaka@unza.zm (S.M.)
- <sup>9</sup> Zambia National Public Health Institute, Ministry of Health, Lusaka 10101, Zambia; kunda.musonda@znphi.com (K.M.); roma.chilengi@znphi.co.zm (R.C.)
- <sup>10</sup> Department of Disease Control, School of Veterinary Medicine, University of Zambia, Lusaka 10101, Zambia
- <sup>11</sup> Macha Research Trust, Choma 20100, Zambia
- \* Correspondence: doreen.shempela@chaz.org.zm (D.M.S.); edgar.simulundu@macharesearch.org (E.S.)



**Citation:** Shempela, D.M.; Chambaro, H.M.; Sikalima, J.; Cham, F.; Njuguna, M.; Morrison, L.; Mudenda, S.; Chanda, D.; Kasanga, M.; Daka, V.; et al. Detection and Characterisation of SARS-CoV-2 in Eastern Province of Zambia: A Retrospective Genomic Surveillance Study. *Int. J. Mol. Sci.* **2024**, *25*, 6338. <https://doi.org/10.3390/ijms25126338>

Academic Editor: Masahiro Fujimuro

Received: 9 April 2024

Revised: 27 May 2024

Accepted: 30 May 2024

Published: 7 June 2024



**Copyright:** © 2024 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/>).

**Abstract:** Mutations have driven the evolution and development of new variants of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with potential implications for increased transmissibility, disease severity and vaccine escape among others. Genome sequencing is a technique that allows scientists to read the genetic code of an organism and has become a powerful tool for studying emerging infectious diseases. Here, we conducted a cross-sectional study in selected districts of the Eastern Province of Zambia, from November 2021 to February 2022. We analyzed SARS-CoV-2 samples ( $n = 76$ ) using high-throughput sequencing. A total of 4097 mutations were identified in 69 SARS-CoV-2 genomes with 47% (1925/4097) of the mutations occurring in the spike protein. We identified 83 unique amino acid mutations in the spike protein of the seven Omicron sublineages (BA.1, BA.1.1, BA.1.14, BA.1.18, BA.1.21, BA.2, BA.2.23 and XT). Of these, 43.4% (36/83) were present in the receptor binding domain, while 14.5% (12/83) were in the receptor binding motif. While we identified a potential recombinant XT strain, the highly transmissible BA.2 sublineage was more predominant (40.8%). We observed the substitution of other variants with the Omicron strain in the Eastern Province. This work shows the importance of pandemic preparedness and the need to monitor disease in the general population.

**Keywords:** SARS-CoV-2; SAR-CoV-2 mutations; Omicron variant; Eastern Province; Zambia; genomic surveillance

## 1. Introduction

Coronaviruses (family *Coronaviridae*) are the largest group of the order Nidovirales and can be present in humans and different animal species [1,2]. Emerging and re-emerging coronaviruses threaten global public health and the socio-economic well-being of populations [3,4]. The emergence of the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003 and the Middle Eastern respiratory syndrome CoV (MERS-CoV) in 2012 were of serious public health concern [5]. Recently, the SARS-CoV-2, a novel beta coronavirus that is genetically distinct from the SARS-CoV and MERS-CoV, has been responsible for the largest global health crisis to date [6,7].

Since the first case of coronavirus disease 2019 (COVID-19) in Wuhan, China in 2019, the SARS-CoV-2, the causative agent for COVID-19, has continued to evolve, acquiring mutations with the potential for evading host immune response [8–11]. In early 2020, the SARS-CoV-2 appeared to evolve relatively slowly for an RNA virus, and it was believed that the proposed vaccines would effectively control infections. However, by the end of December 2020, a highly mutated variant of concern (VOC), the Alpha variant, was first reported in the United Kingdom and spread globally [12]. The Alpha variant was responsible for driving the second wave of the pandemic. Among the notable mutations of the Alpha variant was the N501Y mutation in the S gene, which was associated with the increased binding affinity of the SARS-CoV-2 to the angiotensin-converting enzyme 2 (ACE2) receptor [13]. Similarly, a H69del/V70del mutation of the Alpha variant was associated with immune evasion and S gene target failure (SGTF) [14,15]. Consequently, the S gene target failure was used as a marker of the Alpha variant B.1.1.7 [15]. Similar to what was observed with the Alpha variant, the Beta and Gamma VOC were characterized by mutations in the S gene that resulted in increased transmission and potential for immune escape [14,16–18]. The Delta VOC, responsible for the third wave of the pandemic, became the dominant lineage globally and was characterized by the T478K S gene mutation, which resulted in increased virus binding affinity to ACE2 receptors [19].

In early November 2021, the Network for Genomic Surveillance in South Africa (NGS-SA) reported the emergence of a new and rapidly spreading variant, designated as the Omicron (B.1.1.529) VOC by the World Health Organization (WHO) [20]. The Omicron variant was remarkably genetically distinct, exhibiting over 40 amino acid residue changes in the spike protein [21]. The mutations were reported to increase transmissibility, high viral binding affinity and high potential for immune escape [22–24]. The increased viral fitness of the Omicron VOC was responsible for the fourth global wave of the SARS-CoV-2 pandemic [25,26]. The emergence of VOCs (Alpha, Beta, Gamma, Delta and Omicron) and variants of interest (VOIs; Epsilon, Eta, Iota, Kappa, Zeta and Mu) in the recent past highlights the continuous evolution of the SARS-CoV-2 [27,28]. Similar to what has been observed with the Omicron VOC, it is anticipated that the further evolution of the SARS-CoV-2 will result in variants with potential for antigenic shift, antigenic escape and increased transmissibility [29]. For instance, the B.1.351 (Beta) variant in Zambia was shown to coincide with a sharp increase in confirmed cases between December 2020 and January 2021 [30].

Genomic analysis has proved vital in understanding the continuous evolution of the SARS-CoV-2 [31,32]. However, the slow generation of genomic surveillance data in resource-limited countries like Zambia underlines the knowledge gaps that exist in SARS-CoV-2 surveillance. While recent genomic surveillance efforts have shed some light on the molecular epidemiology of SARS-CoV-2 in Zambia [30,33,34], there is still a paucity of information. This presents a missed opportunity for the early detection of emerging VOCs and VOIs. Moreover, the WHO emphasized the need to conduct the genomic surveillance of pathogens with epidemic and pandemic potential [35]. The Government of the Republic of Zambia has increased its capacity to conduct genomic surveillance for emerging and re-emerging pathogens with pandemic potential [12].

In this study, we performed genomic surveillance for SARS-CoV-2 in humans in selected districts in the Eastern Province of Zambia from November 2021 to February 2022.

## 2. Results

### 2.1. Demographic Summary

In total, 115 deidentified samples positive for SARS-CoV-2 on the rapid antigen test were submitted to the Churches Health Association of Zambia (CHAZ) laboratory for whole genome sequencing (WGS). The samples were collected from selected districts in the Eastern Province of Zambia. Of these, 58.3% (67/115) were obtained from female participants, while 40.2% were collected from male participants. Three samples (2.6%) had missing data. Most of the samples (33.0%) were collected from the 10–19 years age group, signifying a substantial representation of young individuals among SARS-CoV-2 cases, while only eight samples were collected from those who were above 50 years old. Twenty-eight samples (24.3%) had missing demographic data. The majority of the samples 37.4% (43/115) were obtained from Nyimba District while the remaining samples were obtained from Mambwe 18.3% (21/115), Chipangali 13.9% (16/115), Katete 12.2% (14/115), Lundazi 10.4% (12/115) and Chadiza 7.8% (9/115) Districts. The mean number of samples collected per district was 19.2% (Table 1).

**Table 1.** Demographics of participants who tested positive for SARS-CoV-2 in the Eastern Province of Zambia.

Description		N	%
Gender ( <i>n</i> = 112)	Male	45	40.2
	Female	67	59.8
	Unknown	3	2.6
Age ( <i>n</i> = 115)	0–9 years	3	3.4
	10–19 years	38	43.7
	20–29 years	18	20.7
	30–39 years	12	13.8
	40–49 years	8	9.2
	50+	8	9.2
	Unknown	28	24.3
District ( <i>n</i> = 115)	Chadiza	9	7.8
	Chipangali	16	13.9
	Katete	14	12.2
	Lundazi	12	10.4
	Mambwe	21	18.3
	Nyimba	43	37.4

### 2.2. SARS-CoV-2 Genome Detection and Assembly

Of the 115 samples analyzed by RT-qPCR, 35 (30.4%) were positive on three genes (ORF1ab, N, S), while the rest were positive on two genes (ORF1ab, N) (Table 2). For Whole Genome Sequencing (WGS) sample inclusion criteria, samples with Ct values  $\leq 30$  on either the ORF1ab, N or S gene targets were acceptable to undergo sequencing. Thus, all of the samples (*n* = 115; Ct < 30) were subjected to WGS on the Illumina NextSeq2000 platform. Raw reads generated by the Illumina NextSeq2000 platform were assembled into consensus sequences using the DRAGEN COVID Pipeline v1.1.0 available at [https://emea.support.illumina.com/sequencing/sequencing\\_software/dragen-covid-pipeline/downloads.html](https://emea.support.illumina.com/sequencing/sequencing_software/dragen-covid-pipeline/downloads.html) (accessed on 1 February 2022). The non-gap ambiguity fraction rate ranged from 0 to 38.0% (average 10.3%). The mean sequence length was 29,664 (Std Dev: 233.7) while the average GC content was 37.9%. However, out of the 115 consensus sequences generated by the DRAGEN COVID Pipeline, 39 had stretches of more than 10% ‘NNNNNNNN’ and were

thus removed from subsequent analysis. Thus, our final dataset of 76 sequences had an average sequence length of 29,752 (Std Dev: 27.5) and a non-gap ambiguity rate of 3%.

**Table 2.** Summary of TaqPath COVID-19 CE-IVD RT-qPCR assay.

Target			
S Gene	ORF1ab	N Gene	Number of Samples
—	+	+	80
+	+	+	35
		Total	115

### 2.3. Genetic Diversity in SARS-CoV-2 Sequences

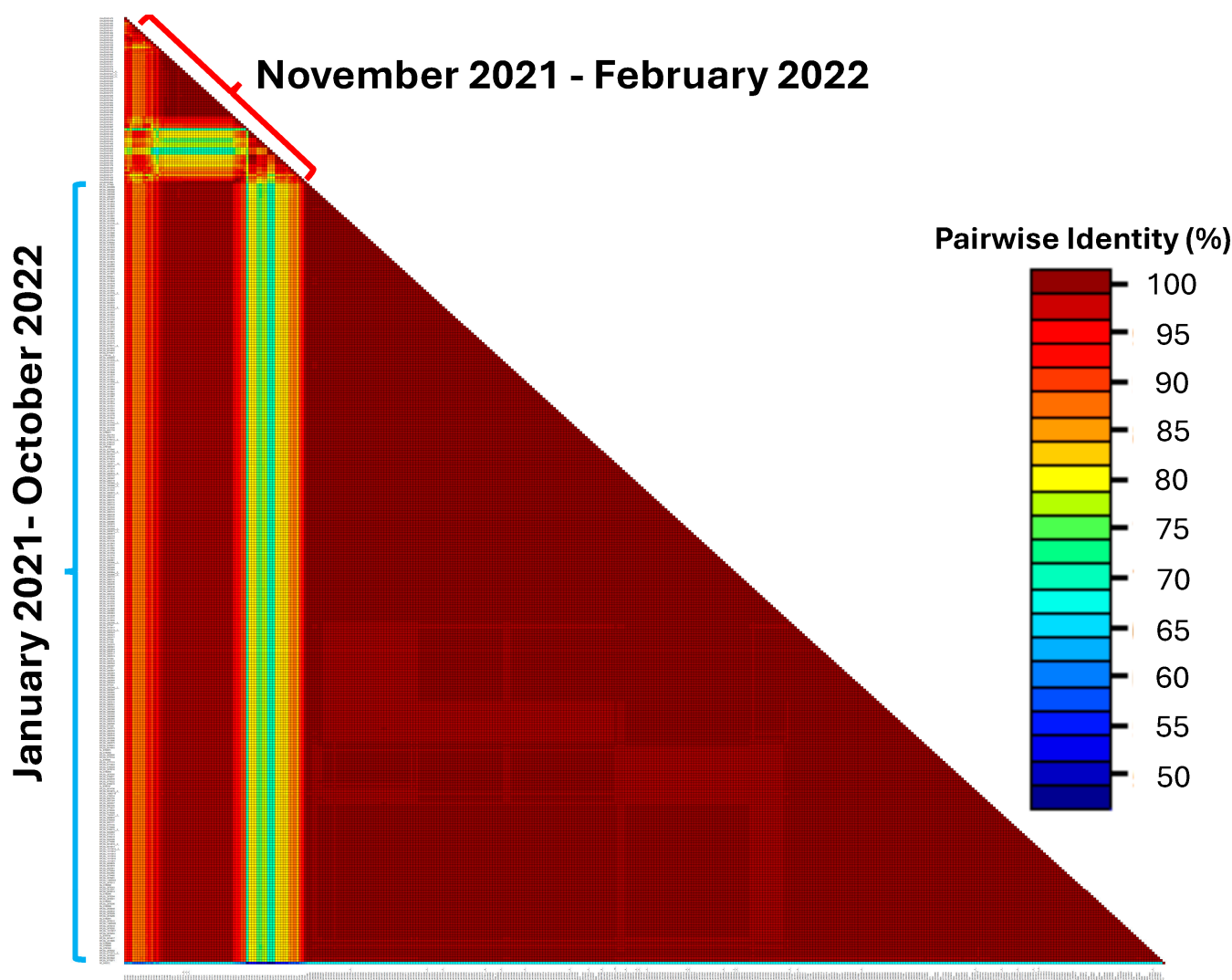
Nucleotide diversity ( $\pi$ ) was calculated across the genome in the 1 kb sliding window with a step size of 200 bp using the recombinant detection program (RDP4) and the Kimura two-parameter model on the nucleotide alignment generated by the MAFFT software [36,37]. There was a noticeable variation (0.00–0.337) in nucleotide diversity across the SARS-CoV-2 genome (Figure 1). Expectedly, high genetic diversity was observed in the S gene (area margined in with red dots), while the rest of the genes had relatively low genetic diversity.



**Figure 1.** Genetic distance plot. Nucleotide diversity of SARS-CoV sequences from this study was calculated using the Wuhan HU-1 reference sequence (accession no. NC\_045512).

To further assess the diversity of SARS-CoV-2 strains in the present study, we used the Sequence Demarcation Tool (SDT) version 1.2 [38] to compare the complete nucleotide sequences of the S gene of SARS-CoV-2 strains from the present study to those detected in Zambia before 1 November 2021. In comparison to SARS-CoV-2 strains reported before 1 November 2021, viruses from this study were genetically diverse (Figure 2), suggesting the emergence of a highly divergent strain.

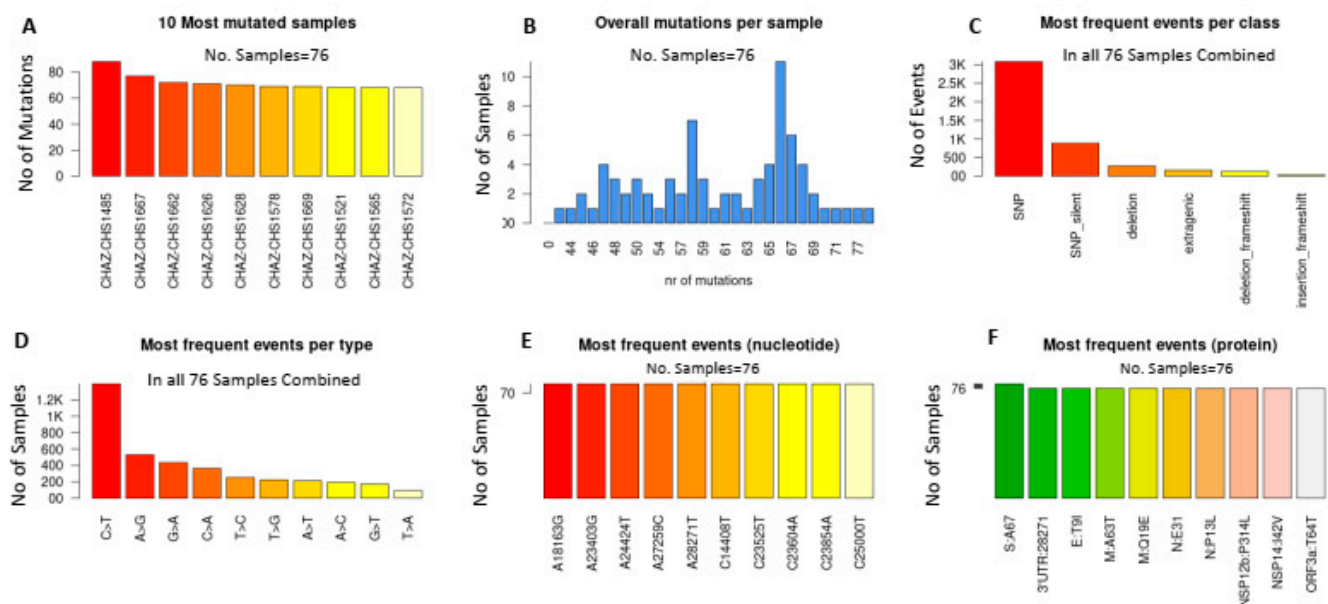




**Figure 2.** Pairwise genetic identity matrices of nucleotide sequences of the complete open reading frame of the SARS-CoV-2 spike protein. Viruses from this study (1 November 2021 to 28 February 2022). Reference sequences detected in Zambia between 1 January 2021 to 31 October 2021 are in black text denoted by blue right brackets. The colour indicates the homology level between sequences. Pairwise matrices were generated using the Sequence Demarcation Tool v.1.2 [38].

#### 2.4. Synonymous and Non-Synonymous Mutations in SARS-CoV-2 Genomes

With respect to the Wuhan HU-1 reference sequence (accession NC\_045512.2), we identified a total of 4097 mutations in the 69 unique SARS-CoV-2 genomes from the present study. Ten of the most mutated samples had at least 60 mutations (Figure 3A,B), while the most encountered events (>2500) were single nucleotide polymorphisms (SNPs) (Figure 3C). Similarly, the C > T transition was the most commonly encountered mutational event (>1200 events; Figure 3D). The most frequently encountered nucleotide substitutions were A18163G, A24424T, A28271T, C23525T and C23854A (Figure 3E), whereas the S: A67, E: T9I, M: Q19E, N: P13L and NSP14:I42V amino acid substitutions were observed in all the analyzed samples (Figure 3F).



**Figure 3.** Mutation analysis of SARS-CoV-2 genomes in the present study. (A) Ten of the most mutated samples. (B) Number of overall mutations per sample. (C) Most frequently observed variant classifications. (D) Most frequently encountered substitution type. (E) Frequently observed nucleotide substitutions. (F) Most frequently observed amino acid mutations. Analysis was performed on 69 complete unique sequences using the Coronapp [39].

A comparative analysis of amino acid sequences of the complete spike protein revealed remarkably high mutation frequencies. A total of 1925 mutations and 83 variant classes were identified in the spike protein. The spike protein accounted for the majority (47.0%; 1925/4097) of the observed mutational events (Table 2). High mutation counts (>100 count) were observed in the NSP3, NSP4, NSP5, NSP6, NSP12b, ORF3ab and M proteins. However, NSP2, NSP16, ORF7a, ORF8 and ORF10 had a low mutational count (<20 counts), suggesting that these genes are relatively stable (Table 3).

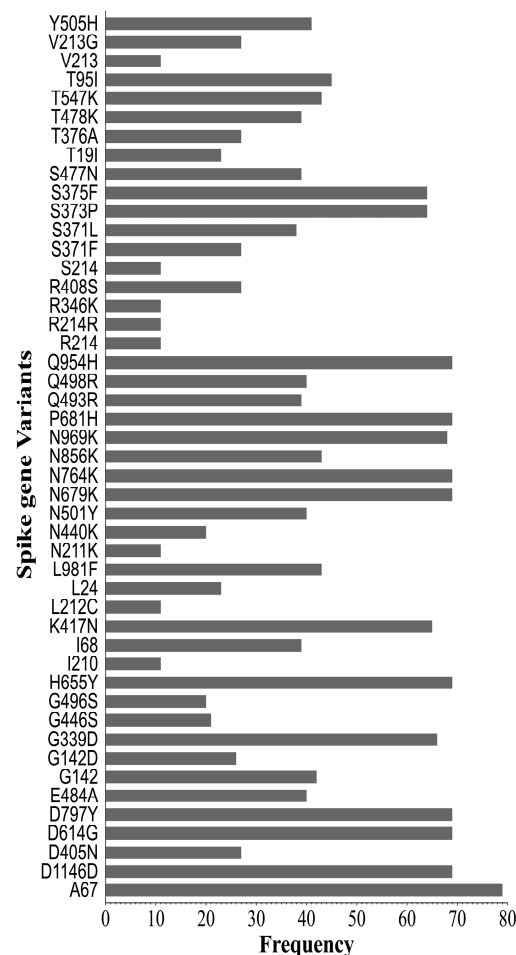
**Table 3.** Mutation count of SARS-CoV-2 Strains with respect to the Wuhan HU-1 reference sequence (accession no. NC\_045512).

Genomic Region	Mutation Count	Annotation
5'UTR	41	5' Untranslated region
NSP1	27	RNA-dependent RNA polymerase
NSP2	8	
NSP3	378	
NSP4	156	
NSP5	134	
NSP6	122	
NSP9	39	
NSP10	43	
NSP12b	146	
NSP13	33	
NSP14	90	
NSP15	68	
NSP16	10	
S	1925	Spike
ORF3a	105	Open reading frame 3a protein
E	70	Envelope protein
M	198	Membrane protein
ORF6	96	Open reading frame 6 protein
ORF7a	5	Open reading frame 7a protein

Table 3. Cont.

Genomic Region	Mutation Count	Annotation
ORF7b	59	Open reading frame 7b protein
ORF8	13	Open reading frame 8 protein
N	235	Nucleocapsid protein
ORF10	2	Open reading frame 10 protein
3'UTR	94	3' Untranslated region
Total	4097	

Of the 83 variant classes, nine (i.e., Q954H, P681H, N764K, N679K, H655Y, D797Y, D614G, D1146D and A67) were present in all of the analyzed samples, while 41% (34/83) (A163A, A27S, A372, A397A, A475A, A845S, H69, I410V, I434M, K182K, K986T, L212L, L223L, L368I, N1098N, N211D, N234S, N370S, R190G, R237G, R328G, R403G, R454R, S399S, T333A, T376, T385A, T430A, T881T, V1264V, V367I, Y160C, Y170C and Y200C) were unique to only one sample (Figure 4).



**Figure 4.** Frequency of observed variants in the spike protein of SARS-CoV-2 strains from the present study. The dotted blue line denotes the total number of analyzed samples ( $n = 69$ ). Analysis was performed on the complete S protein of 69 unique sequences using the Coronapp [39].

In the receptor binding domain (RBD), the most predominant variants (>19 counts) were G339D, K417N, S375F, S373P, Y505H, E484A, Q498R, N501Y, S477N, T478K, Q493R, S371L, D405N, T376A, S371F, R408S, G446S, G496S and N440K (Table 4). The least encountered variants (<12 count) were R346K, T547K, A372, L368I, N370S, T430A, A397A, T376, I410V, T385A, I434M, R403G, S399S, V367I, A475A, R454R and T333A. Within the receptor

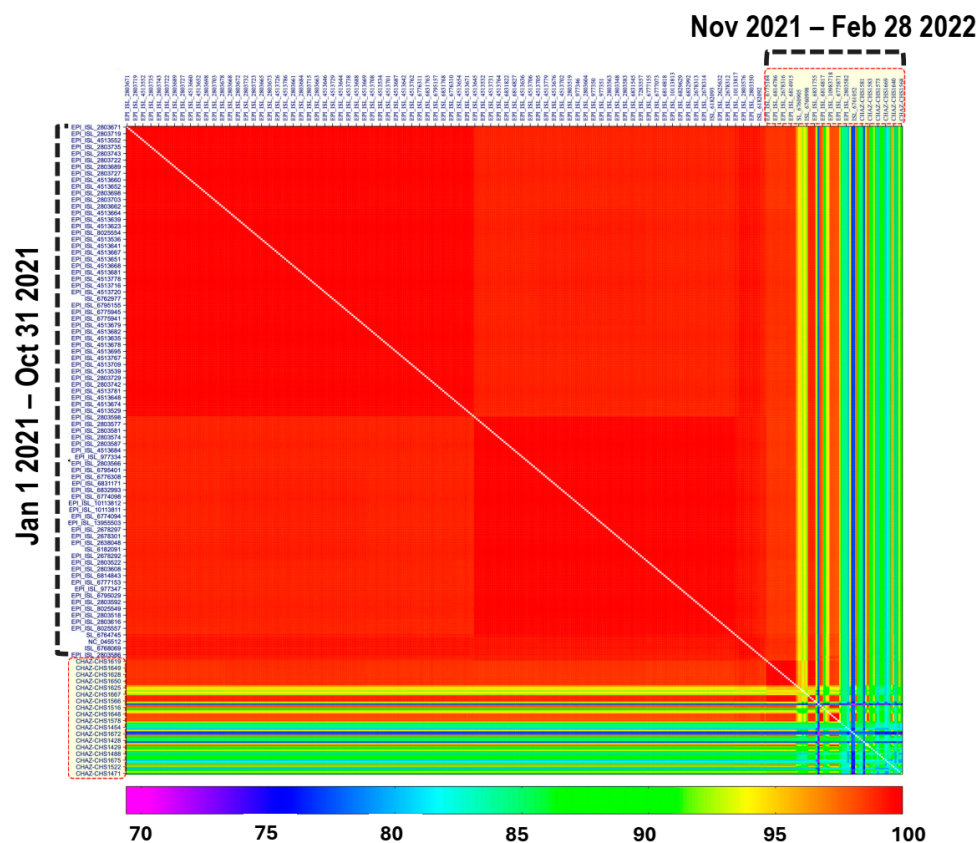
binding motif, 12 variant classes, that is, Y505H, Q498R, E484A, N501Y, T478K, S477N, Q493R, G446S, G496S, N440K, R454R and A475A were identified. Y505H, Q498R, E484A, N501Y, T478K, S477N, Q493R, G446S and G496S were the most encountered mutations (>19 count), while N440K, R454R, and A475A were rare variants (<3 count).

**Table 4.** Mutational count in the receptor binding domain and receptor binding motif.

Receptor Binding Domain		
Variant Class	Count	%
G339D	66	1.6
K417N	65	1.6
S375F	64	1.6
S373P	64	1.6
Y505H	41	1.0
E484A	40	1.0
Q498R	40	1.0
N501Y	40	1.0
S477N	39	1.0
T478K	39	1.0
Q493R	39	1.0
S371L	38	0.9
D405N	27	0.7
T376A	27	0.7
S371F	27	0.7
R408S	27	0.7
G446S	21	0.5
G496S	20	0.5
N440K	20	0.5
R346K	11	0.3
T547K	6	0.1
A372	1	0.0
L368I	1	0.0
N370S	1	0.0
T430A	1	0.0
A397A	1	0.0
T376	1	0.0
I410V	1	0.0
T385A	1	0.0
I434M	1	0.0
R403G	1	0.0
S399S	1	0.0
V367I	1	0.0
A475A	1	0.0
R454R	1	0.0
T333A	1	0.0
Total	776	18.9
Receptor Binding Motif		
Y505H	41	1.0
Q498R	40	1.0
E484A	40	1.0
N501Y	40	1.0
T478K	39	1.0
S477N	39	1.0
Q493R	39	1.0
G446S	21	0.5
G496S	20	0.5
N440K	2	0.0
R454R	1	0.0
A475A	1	0.0
Total	323	7.9

### 2.5. Highly Divergent Strain Rapidly Replaced Other Variants after 1 November 2021

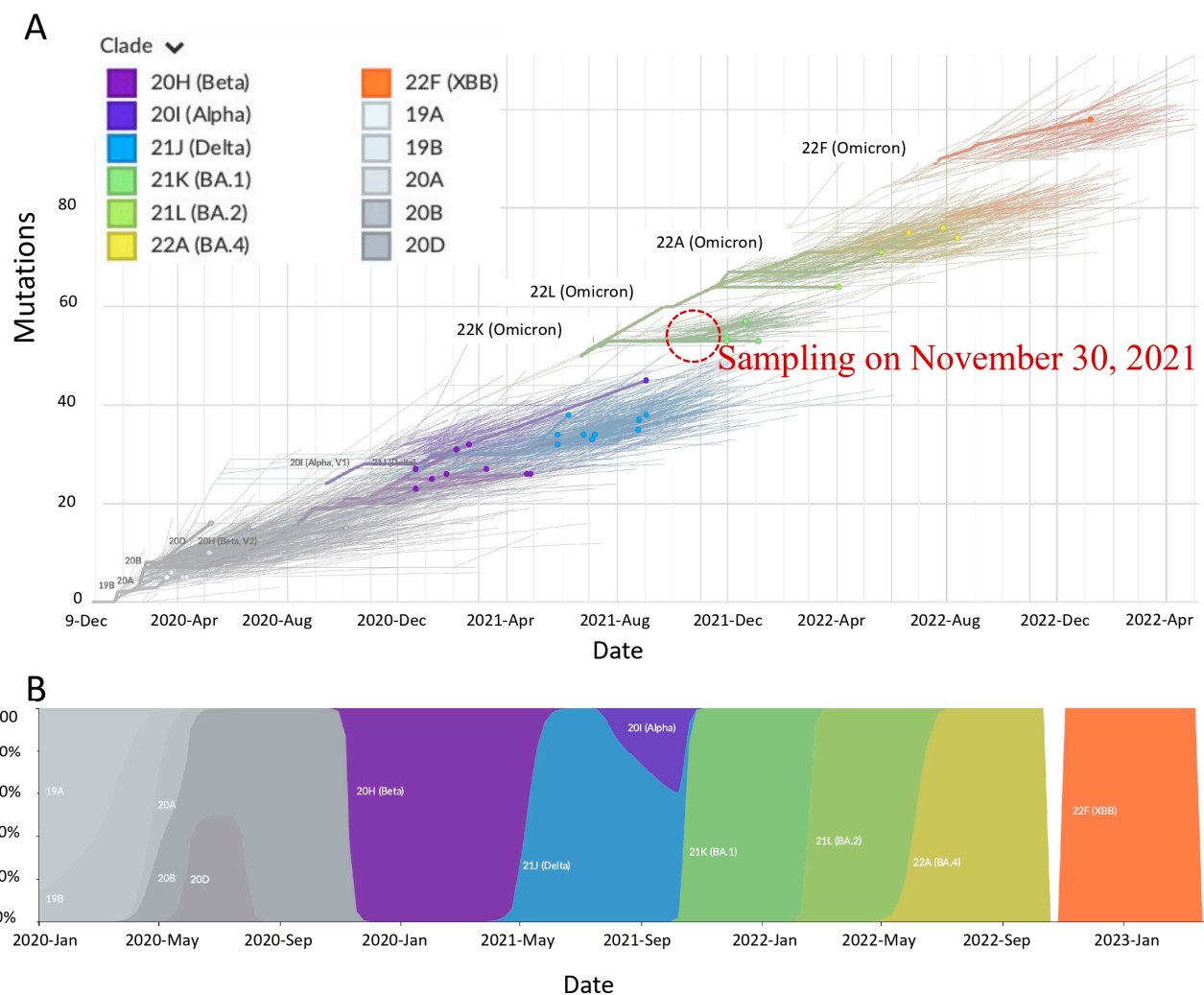
A pairwise matrix was generated in the Geneious R11 software using MAFFT (Figure 5) [37] available at <https://mafft.cbrc.jp/alignment/server/> (accessed on 5 May 2024).



**Figure 5.** Pairwise nucleotide alignment of complete SARS-CoV-2 genomes from the present study and those downloaded from the GISAID database at <https://www.epicov.org/epi3/> (accessed on 5 May 2024). Viruses from this study are in blue text and yellow highlight. Reference sequences are denoted in blue text. The horizontal bar denotes the percent similarity between sequences.

The pairwise nucleotide alignment of the complete SARS-CoV-2 genomes from this study and those downloaded from the GISAID database at <https://www.epicov.org/epi3/> (accessed on 5 May 2023) showed that viruses detected after 1 November 2021 (this study) were genetically diverse as compared to those reported before 1 November 2021 (Figure 6). A time series analysis of the variants detected in Zambia so far showed the rapid replacement of other variants with the Omicron strain (Figure 6A,B). An inspection of nucleotide sequences deposited in the GISAID database shows that the first Omicron (BA.1, BA.1.17.2, and BA.1.19) samples in Zambia were collected 3 weeks (30 November 2021) after the first report of the Omicron variant in Botswana and South Africa [20]. Moreover, as of 4 February 2024, only Omicron and its subvariants have been reported from Zambia (GISAID; accessed 4 February 2024) so far. Notably, the early presence of the S gene target present (SGTP) BA.2 lineage in Eastern Province and other parts of the country resulted in some samples being misdiagnosed as Delta variant resurgence at the PCR stage. Taken together, our results suggest that the Omicron strain might have been present in Zambia earlier than previously thought. Likewise, the lack of detection of the other variants in samples collected after 1 November 2021 intimates that the Omicron variant was the dominant circulating strain.

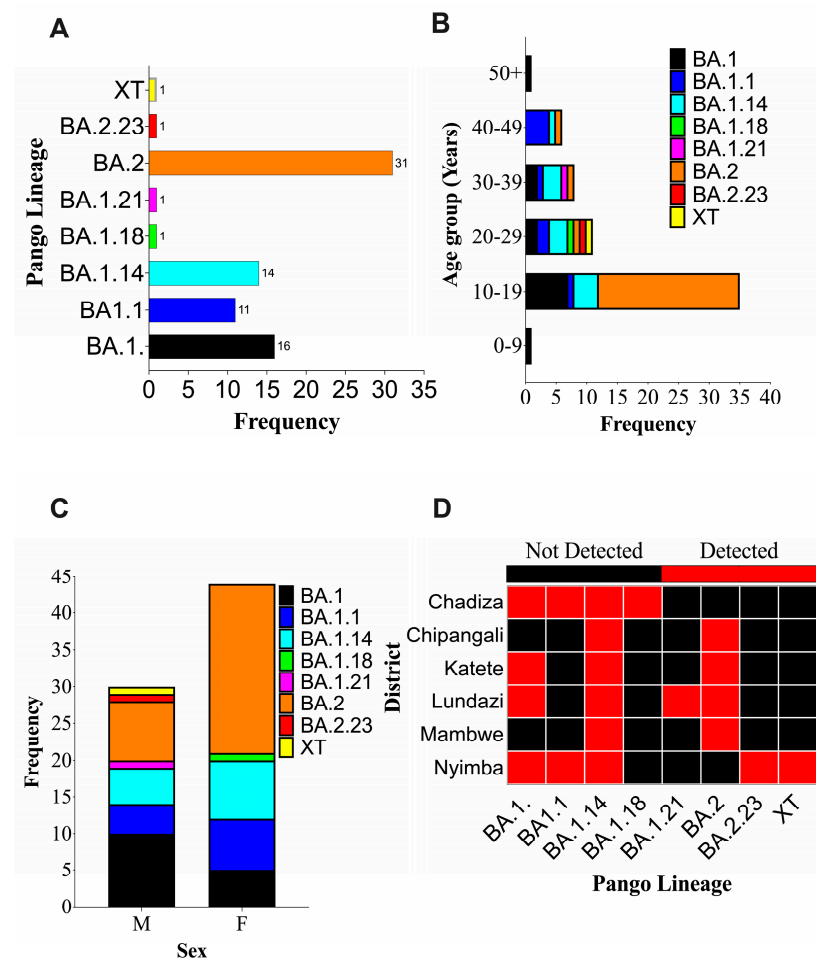




**Figure 6.** SARS-CoV-2 lineages identified in this study. (A) Frequency of SARS-CoV-2 Pango lineages identified in Eastern Province. (B) Distribution of SARS-CoV-2 Pango lineages in Eastern Province.

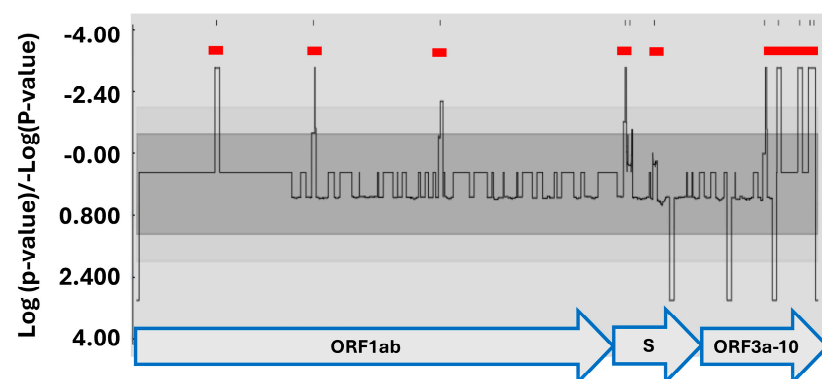
## 2.6. Lineage, Clade Assignment and Recombination Analysis

The PANGOLIN [40] (<https://pangolin.cog-uk.io/>; accessed on 7 May 2023) and Nextclade [41] (<https://clades.nextstrain.org/>; accessed on 7 May 2021) application softwares were used for lineage and clade classification. Samples ( $n = 76$ ) from the present study were classified into eight Pango lineages, that is, BA.1, BA.1.1, BA.1.14, BA.1.18, BA.1.21, BA.2, BA.2.23 and XT. Interestingly, the XT strain (GSAID accession number EPI-1S1-10103848) was classified as a recombinant variant on the Nextclade analysis suit. Among the eight identified lineages, BA.2 was predominant, accounting for 40.8% (31/76) of all identified lineages (Figure 7A) while BA.1.21, BA.1.18, BA.2.23 and XT were the least (1.3%, 1/76) encountered lineages (Figure 7A). Intriguingly, the BA.2 lineage was more predominant in the 10–19-year-olds (Figure 7B) and also accounted for the majority of the identified lineages in female participants (Figure 7C). High lineage diversity was observed in 20–29-year-olds in whom all eight lineages (BA.1, BA.1.1, BA.1.14, BA.1.18, BA.1.21, BA.2, and BA.2.23) were identified (Figure 7B). With respect to lineage distribution, BA.1.14 was identified in all of the study sites, while the rare variants (BA.1.21, BA.1.18, BA.2.23 and XT) were detected only in Chadiza (BA.1.18), Lundazi (BA.1.21) and Nyimba (BA.2.23, XT) Districts. Furthermore, the eight lineages were classified into clades 21K (43/76) and 21L (32/76).



**Figure 7.** SARS-CoV-2 Lineages identified in this study. (A) Frequency of SARS-CoV-2 Pango lineages identified in Eastern Province. (B) Distribution of SARS-CoV-2 Pango lineages by age groups. (C) Distribution of Pango lineages by sex. (D) Detected Pango lineages by District.

To account for the observed potential recombination event, (GSAID accession no. EPI-1S1-10103848), nucleotide sequences from this study together with those downloaded from the GISAID database were analyzed using the Recombination Detection Program version 4.101 (RDP4) [40]. Recombination breakpoints were detected in the ORF1ab, S, M and N (Figure 8).

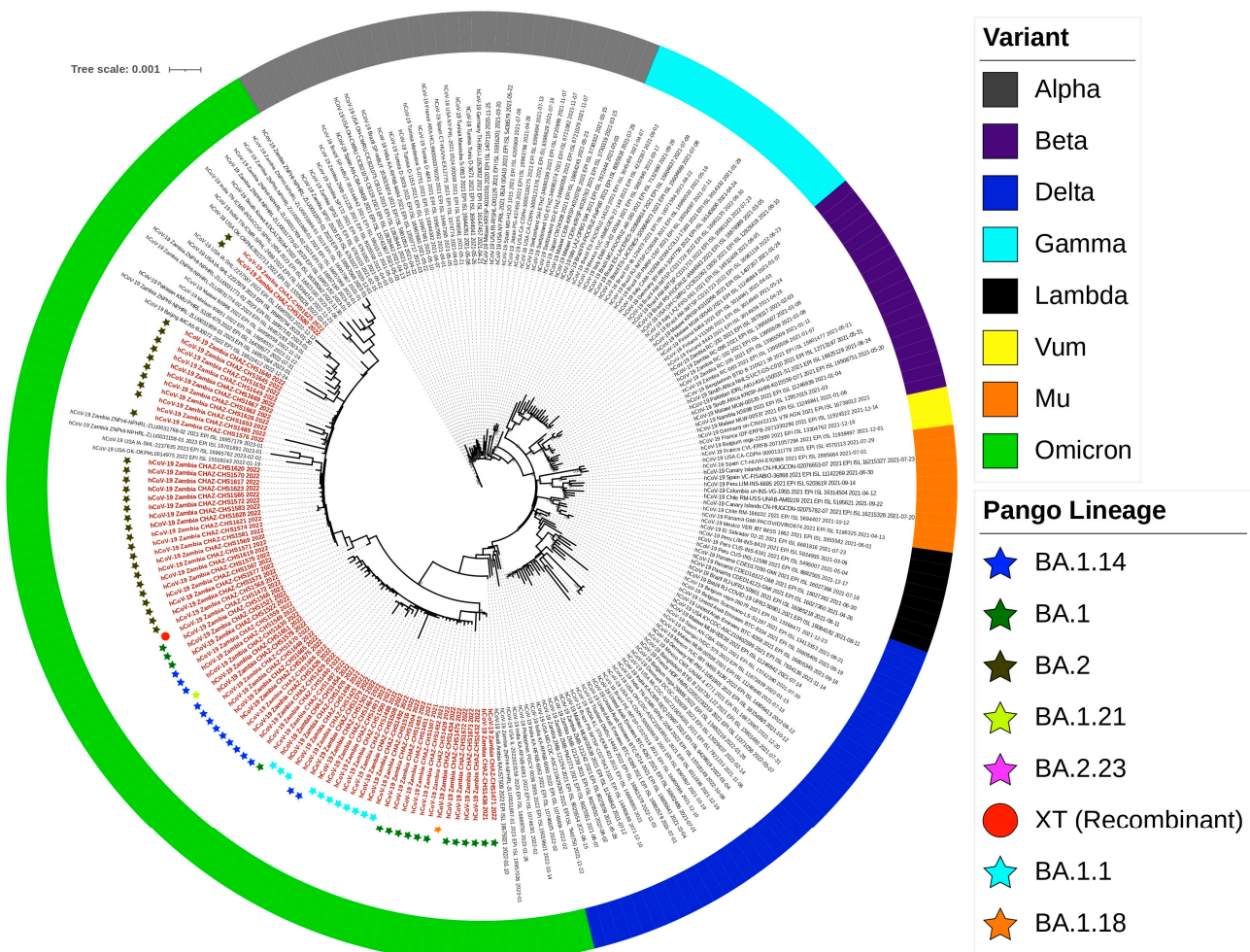


**Figure 8.** Detection of recombination hot/cold spots. Analysis was performed in RPD4 using a 200-base pair (bp) window at a 20-bp step and the Kimura two-parameter model on a nucleotide alignment generated by MAFFT. Recombination hotspots are denoted by red horizontal bars.

## 2.7. Phylogenetic Analysis

On phylogeny, sequences from the present study formed a monophyletic group with other previously reported Omicron strains from Asia, Europe, America and Africa. This was suggestive of probable local and/or international transmission. Furthermore, Omicron strains shared a common ancestor with the Delta variant, exhibiting a descendant-like pattern. This finding supports earlier evidence of the emergence of the Omicron strain from the Delta variant. Topologically, Omicron strains from this study were classified into eight distinct lineages, that is, BA.1, BA.1.1, BA.1.14, BA.1.18, BA.1.21, BA.2, BA.2.23 and XT (Figure 8).

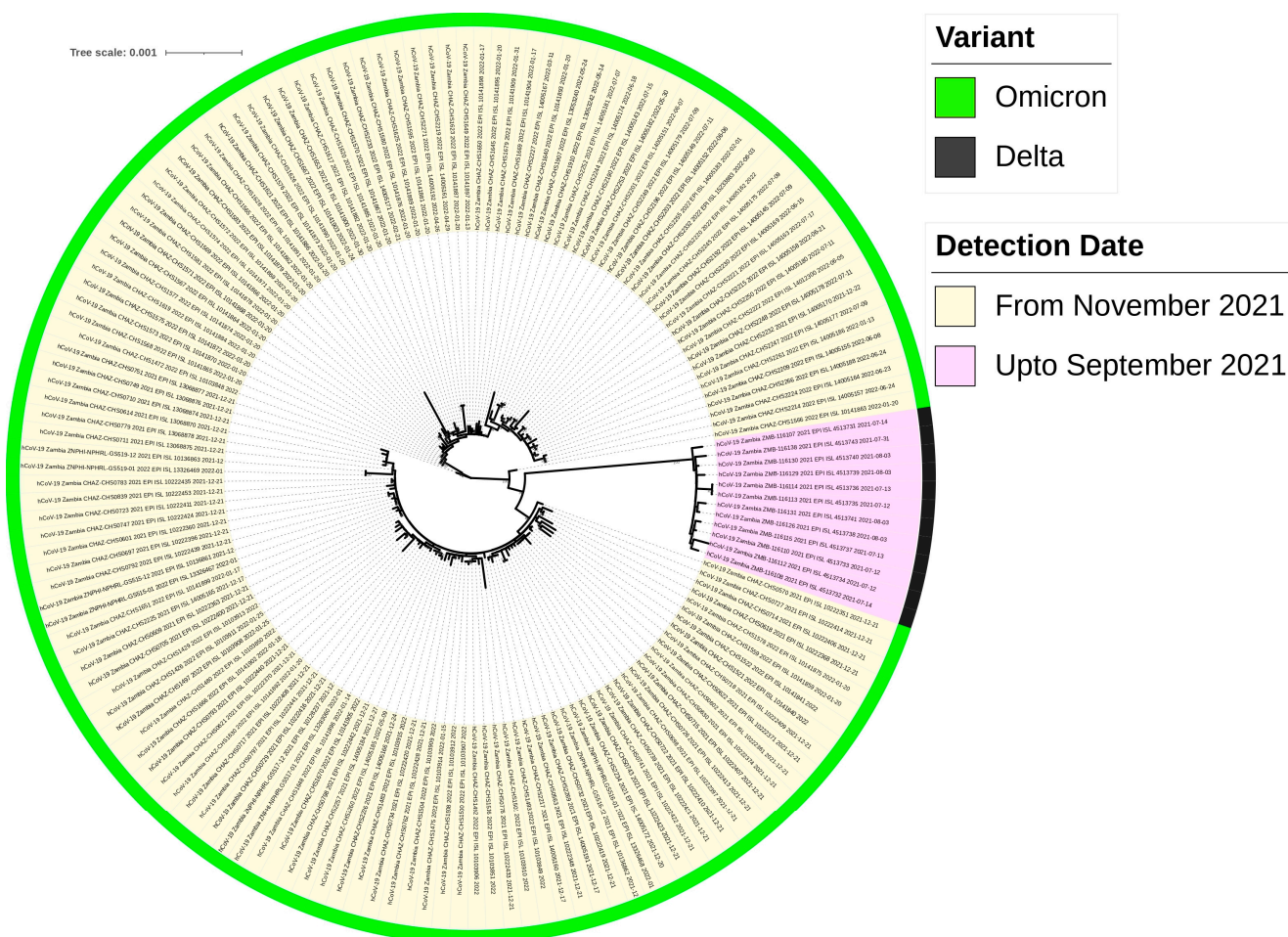
The phylogenetic tree was implemented in the IQ TREE [41] according to the best nucleotide substitution model (GTR + F+I + G4) in ModelFinder [42]. The reliability of the Phylogenetic tree was evaluated by 10,000 ultrafast bootstrap replicates [43]. Sequences generated in this study are denoted in red text, while reference sequences are in purple text. Coloured strips represent the SARS-CoV-2 variant. The coloured star denotes the assigned Pango Lineage. Bar, number of substitutions per site (Figure 9).



**Figure 9.** Maximum likelihood phylogenetic tree of SARS-CoV-2 genomes from Zambia and reference sequences retrieved from the GISAID database. The tree was implemented in IQ TREE [41] based on the best nucleotide substitution model (GTR + F+I + G4) in ModelFinder [42].

Furthermore, the phylogenetic analysis of all samples collected in Eastern Province so far showed the rapid replacement of the Delta variant with the Omicron strain (Figure 10). This supports our earlier observation (Figures 5 and 6) of the rapid replacement of the Delta variant after 1 November 2021.





**Figure 10.** Maximum likelihood phylogenetic tree of SARS-CoV-2 genomes from the Eastern Province of Zambia collected between September 2021 and October 2022. The tree was implemented in IQ TREE [41] based on the best nucleotide substitution model (GTR + F+I + G4) in ModelFinder [42]. Phylogenetic tree reliability was evaluated by 10,000 ultrafast bootstrap replicates [43]. Coloured strips represent SARS-CoV-2 variants. Bar, number of substitutions per site.

### 3. Discussion

This study investigated mutations and their evolutionary relationships with other variants in the SARS-CoV-2 Omicron strain detected among residents of six selected districts in the Eastern Province of Zambia. This area was disproportionately affected by the surging numbers of COVID-19 infections at a time when numbers were receding in other provinces. We observed high genetic diversity in the spike protein. Similarly, high mutational counts were observed in the NSP3 to NSP6, NSP12b, ORF3ab and the M protein. Our findings are consistent with previous reports of high variations in the SARS-CoV-2 genome [44,45].

The present study revealed that there were more cases of the Omicron SARS-CoV-2 observed in females compared to males. Our findings are consistent with those of other studies that have reported a higher prevalence of the Omicron SARS-CoV-2 among females compared to males in Malawi, the United States of America and Zambia [34,46,47]. However, other studies have reported contrasting results, with more males being infected with the Omicron SARS-CoV-2 compared to females [48] with consistently more severe disease in males than in females [49,50]. These differences could be explained by the differences in transmission dynamics between the different settings, although our relatively limited sample size could also have influenced our findings.

The predominant age of patients was 10–19 years old, signifying a substantial level of SARS-CoV-2 infection among young individuals, while only eight samples were col-

lected from those who were aged above 50 years old, indicating a lower prevalence of the infection in the older age group. Other studies have postulated that younger groups may drive COVID-19 infection due to frequent interactions such as during playing or school settings [51,52]. Despite the younger age groups being drivers of COVID-19 infection, they are less affected by the severe outcomes of the disease [51]. However, it cannot be ruled out that the small number of samples analyzed in this study may have had an impact on the observed gender distribution and age of COVID-19 patients, as was observed in a previous study [34].

This present study found that there was high genetic diversity in the S gene. High genetic diversity in the S gene has also been reported in other studies [53–55]. The areas of high diversity indicate highly mutated genes. Consequently, viruses, including the SARS-CoV-2, use mutations for evolution, survival, fitness and pathogenesis [53,54]. Our study found a total of 4097 mutations from 69 unique SARS-CoV-2 genomes, of which 1925 were in the spike protein. Additionally, 47% of mutational events were biased towards the spike protein, with high mutation counts observed in the NSP3, NSP4, NSP5, NSP6, NSP12b, ORF3ab and M proteins. Consequently, 83 variant classes were also obtained in the spike protein. These mutational events and variants indicate high diversity in the spike protein, similar to reports from other studies [56–59]. Unfortunately, the increased mutations in the spike protein of the SARS-CoV-2 Omicron variant may lead to the evolution of invasive and adaptive variants [21,60]. Additionally, a high frequency of single nucleotide polymorphisms (SNPs) was also found in our study, especially A18163G, A24424T, A28271T, C23525T and C23854A. Single nucleotide variants (SNVs) have also been reported elsewhere [61]. Similar findings were reported in a previous study in South and Southeast Asia, where high-frequency substitution sites were mainly localized in the S gene, including A24424T and C23525T [62]. The high frequency of these SNPs in a population may indicate genetic diversity, population stratification, disease prognosis or a lower prevalence of diseases and a higher risk of severe COVID-19 in certain individuals [63,64]. In South Africa, the Omicron variant had an average of 50 mutations, of which approximately 30 mutations were in the spike protein and about 15 in the RBD which affect virus transmission and immune escape [65]. Our findings and those reported in other studies imply that the continuous evolution of the S gene due to mutations threatens vaccine efficacy because the spike protein has been the main target of vaccine development [66,67].

The present study found 83 variant classes, of which 43.4% were located in the receptor binding domain (RBD) of the spike protein. These findings have also been reported in other countries [68–71]. This is a significant finding as the RBD is crucial for the interaction between the virus and the host cell receptor, ACE2 [13,72]. The presence of mutations in the RBD suggests potential impacts on viral high infectivity and immune recognition [21,73–75]. The stronger interaction between the RBD and ACE2 can cause Omicron to evade the antibodies that are produced by the COVID-19 vaccines [28,76,77]. This may render vaccines less effective against the SARS-CoV-2 Omicron variant [27,78]. The current study also identified specific variants that were more frequently encountered, including G339D, K417N, S375F, S373P, Y505H, E484A, Q498R, N501Y, S477N, T478K, Q493R, S371L, D405N, T376A, S371F, R408S, G446S, G496S and N440K. These variants may have functional implications, such as altered binding affinity to the ACE2 receptor or immune evasion, leading to the increased infectivity of the variants [71,73]. It is worth noting that some variants were rare, and these included R346K, T547K, A372, L368I, N370S, T430A, A397A, T376, I410V, T385A, I434M, R403G, S399S, V367I, A475A, R454R and T333A. Although these variants were less frequently encountered, they may still have important implications for viral fitness, immune evasion and pathogenicity.

The phylogenetic analysis of our study indicated that Omicron evolved independently from the Delta variant in the Eastern province, which has important implications for our understanding of the virus's evolution and spread [79]. Notably, the Omicron variant emerged as the dominant strain from November 2021, while the Delta variant was scarcely detected. These findings resonate with findings from Puerto Rico, where the BA1 variant



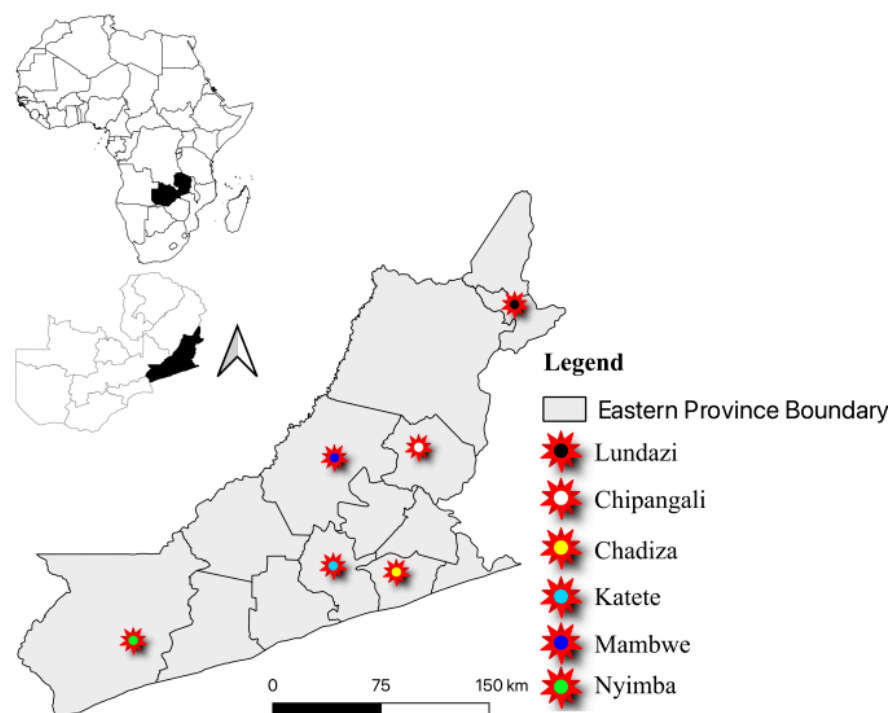
emerged replacing Delta as the dominant variant in December 2021, which was followed by increased transmission and a dynamic landscape of Omicron sublineage infections in the population [80]. Further studies have reported increased transmission and infection attributable to the emergence of the Omicron variant after the replacement of Delta in England, Finland and the United States of America [81–83]. Further, the spread of the Omicron variant in Eastern Province exhibited similarities to its spread in South Africa and Malawi [84,85]. This highlights the interconnectedness of regional transmission dynamics and emphasizes the importance of collaborative efforts in addressing the evolving landscape of SARS-CoV-2 variants [78].

We are aware that our study was conducted in one province of Zambia, thereby limiting the generalization of the findings to the rest of the country. However, our study suggests the co-circulation of some Omicron subvariants, which raises the risk of the generation of recombinants with potentially devastating mutations. We believe that this is the first genomic study on the SARS-CoV-2 in Eastern Province, Zambia, emphasizing the need for more comprehensive nationwide genomic studies to understand the evolution of the SARS-CoV-2 in Zambia. Additionally, the relative proximity of the Eastern province to Malawi poses a serious concern with cross-border transmission. Overall, the findings from this study contribute valuable insights into viral evolution patterns and highlight important considerations for genetic surveillance efforts aimed at monitoring changes in SARS-CoV-2 variants.

#### 4. Materials and Methods

##### 4.1. Study Area and Design

We conducted a cross-sectional study from November 2021 to February 2022 in the Chadiza, Nyimba, Katete, Chipangali, Mambwe and Lundazi Districts of the Eastern Province of Zambia (Figure 11). The Eastern Province of Zambia has a population of over 2.4 million people [86]. There is a high burden of HIV and Malaria in this region, and the proximity of the province to Malawi and Mozambique raises serious concerns about the cross-border spread of communicable diseases such as COVID-19.



**Figure 11.** Map of Eastern Province showing the study sites of Chadiza, Nyimba, Katete, Chipangali, Mambwe and Lundazi Districts.

#### 4.2. Sample Collection and RNA Extraction

The criteria for next-generation sequencing (NGS) sample submission to the Churches Health Association of Zambia (CHAZ) laboratory were COVID-19 samples positive for the SARS-CoV-2 on the rapid antigen test kit. A total of 192 SARS-CoV-2 respiratory samples (i.e., nasal and throat swabs) positive for the SARS-CoV-2 antigen on the rapid test kit were included in this study. The samples were collected during routine surveillance and cluster outbreaks between November 2021 and February 2022. Sample collection was carried out using a flocked nasopharyngeal swab and stored on ice in virus transport media (COPAN Diagnostics, Inc., Murrieta, CA, USA) before being transported to the CHAZ Laboratory in Lusaka. Anonymized patient forms accompanying the samples were used to collect patient metadata including age, gender, place of residence, sampling date and clinical symptoms. These data were then entered into the DISA Laboratory Information and Management system (Laboratory System Technologies Ltd., Johannesburg, South Africa).

Viral RNA was extracted from nasopharyngeal swabs using the MagMAX viral isolation kit (Applied Biosystems, Foster City, CA, USA) on an automated Kingfisher Flex 96 Deep-well magnetic particle processor (ThermoFisher Scientific, Waltham, MA, USA) according to the manufacturer's recommendation. Briefly, in a class II biosafety cabinet, a binding bead mix was prepared and aliquoted into 275 µL per sample well to which 200 µL of sample was added. Further, 20 µL of proteinase K/MS2 solution was added to the sample/binding bead mix and shaken at 1050 rpm for 2 min. Following the removal of the supernatant, samples were washed twice with wash buffer, and RNA was eluted in 50 µL of elution buffer for downstream processing.

#### 4.3. SARS-CoV-2 Genome Detection by RT-qPCR

To confirm SARS-CoV-2 samples for whole genome sequencing and eliminate low copy number samples (sample cycle threshold (Ct) > 30), screening for the SARS-CoV-2 genome was conducted using the TaqPath COVID-19 CE-IVD RT-qPCR assay (ThermoFisher Scientific, Waltham, MA, USA) in a 25 µL reaction mix containing 15 µL of TaqPath qPCR master mix and 10 µL of extracted RNA.

#### 4.4. cDNA Synthesis and Amplification of SARS-CoV-2

First-strand cDNA synthesis for samples was achieved through random hexamer priming using the First Strand cDNA master mix (Illumina) according to the manufacturer's recommended protocol. Briefly, in a 96-well PCR plate, 8.5 µL of random hexamers were added to an equal volume of extracted RNA and denatured on an ABI 7500 real-time thermal cycler for 3 min at 65 °C. Ten µL of First Strand Mix and 1 µL of Reverse Transcriptase were then added to the denatured sample. cDNA synthesis was achieved with the following cycling conditions: 5 min at 25 °C, 10 min at 50 °C and 5 min at 80 °C. SARS-CoV-2 genome amplification was conducted using the ARTIC network.

V4 primer pools ARTIC V4 (<https://github.com/artic-network/primer-schemes>; accessed on 7 October 2021). The primer pool amplification employed two reactions per sample, i.e., COVIDseq Primer Pool 1 (CPP1) and COVIDseq Primer Pool 2 (CPP2). The reaction components for each reaction consisted of 12.5 µL of Illumina PCR Master Mix, 3.5 µL of either CPP1 or CPP2, 5 µL of first-strand cDNA synthesis and 3.9 µL of nuclease-free water. The thermoprofile was as follows: holding stage at 98 °C for 3 min, followed by 35 cycles of 98 °C for 15 s and 63 °C for 5 min. For each run, a single positive control (TaqPath COVID-19 Control; ThermoFisher, Carlsbad, CA, USA) and negative no template control (nuclease-free water) were included to serve as indicators of extraneous nucleic acid contamination. PCR amplicons for each sample were then combined by transferring 10 µL from each well of the CPP1 and CPP2 into a new well.

#### 4.5. Library Preparation, Illumina Sequencing and Genome Assembly

Library preparation was performed using the Illumina COVIDSeq kit (Illumina Inc., San Diego, CA, USA) on the automated Hamilton robotic instrument (Hamilton, NV, USA).

Pooled PCR products were processed for tagmentation and adapter ligation using the Illumina COVIDSeq Kit with IDT Illumina-PCR indexes. Pooling and library clean-up were performed as per the protocol provided by the manufacturer (Illumina Inc.). Pooled libraries were quantified on the Qubit 4.0 fluorometer (Invitrogen Inc., Waltham, MA, USA) using the Qubit dsDNA High Sensitivity kit. The pooled library was normalized to a 4 nM concentration. The library was further diluted to a final loading concentration of 1 nM using a resuspension buffer and sequenced (301 paired-end) on the Illumina NextSeq 2000 (Illumina, San Diego, CA, USA) platform.

To assemble SARS-CoV-2 whole genomes, the Illumina DRAGEN DNA pipeline was used to analyze sequence reads prepared using the ARTIC gene panel assay (<https://github.com/artic-network/primer-schemes>; accessed on 7 October 2021). The DRAGEN pipeline uses a kmer reference database to match kmers from the sequencing read to kmers from the SARS-CoV-2 reference genome (Wuhan Hu-1, accession no. NC\_045512). The kmer reference list is created by splitting the SARS-CoV-2 in 32bp kmers while any kmers that contain cross-reactivity are eliminated. Variant calling and consensus sequence generation were then performed for each sample using the DRAGEN COVID-19 pipeline. Consensus sequences were initially annotated using Glimmer in the Geneious software using the SARS-CoV-2 reference genome (NC\_045512). Annotations were verified for reliability using genome annotation transfer utility (GATU) software [87].

Maximum likelihood phylogenetic analysis was implemented in IQ TREE [41] based on the best nucleotide substitution model in ModelFinder [42]. Phylogenetic tree reliability was evaluated by 10,000 ultrafast bootstrap replicates [43]. Tree editing and annotation were performed in the Interactive Tree of Life (iTOL) [88] software available at <https://itol.embl.de> (accessed on 7 May 2023).

#### 4.6. Genetic Diversity and Mutations of SARS-CoV-2 Genomes

Analysis was performed in recombination detection program version 4 (RPD4) using a 200-base pair (bp) window at a 20-bp step and the Kimura two-parameter model on a nucleotide alignment generated by the MAFFT software. The analysis of synonymous and non-synonymous mutations was performed according to the complete Wuhan HU-1 reference strain (accession No. NC\_045512.2). Mutation analysis in this study was performed on 69 complete unique sequences using the Coronapp, available at <http://giorgilab.unibo.it/coronannotator/> (accessed on 7 May 2023). We conducted a comparative analysis of amino acid sequences of the complete spike protein to detect mutations and variants in the spike protein. The analysis was performed on the complete S protein of 69 unique sequences using the Coronapp, available at <http://giorgilab.unibo.it/coronannotator/> (accessed on 7 May 2023). We further analyzed the Lineage, Clade Assignment and Recombination Analysis in the present study.

**Author Contributions:** Conceptualization, D.M.S.; methodology, D.M.S. and J.S.; software, H.M.C.; validation, G.K., S.M. (Sody Munsaka), E.S. and H.M.C.; formal analysis, H.M.C.; investigation, D.M.S.; resources, K.S.; data curation, K.M., R.C. and E.S.; writing—original draft preparation, D.M.S.; writing—review and editing, D.M.S., J.S., F.C., L.M., S.M. (Steward Mudenda), D.C., M.K., M.N., V.D., R.C., K.S., G.K., S.M. (Sody Munsaka) and E.S.; visualization, D.M.S. and E.S.; supervision, S.M. (Sody Munsaka) and E.S.; project administration, D.M.S. and K.S.; funding acquisition, D.M.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Zambia Biomedical Research Ethics Committee (approval number REF 3574-2022). The study was of no risk to participants as only archived deidentified samples were used in the study. All study data were restricted to the investigators and confidentiality was strictly adhered to.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data supporting the findings from this study are available from the corresponding author on request.

**Acknowledgments:** The authors would like to acknowledge the management and staff of the Churches Health Association for their support during the conduct of the study.

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

## References

- Swelum, A.A.; Shafi, M.E.; Albaqami, N.M.; El-Saadony, M.T.; Elsify, A.; Abdo, M.; Taha, A.E.; Abdel-Moneim, A.M.; Al-Gabri, N.A.; Almainan, A.A.; et al. COVID-19 in Human, Animal, and Environment: A Review. *Front. Vet. Sci.* **2020**, *7*, 564334. [\[CrossRef\]](#)
- Santaniello, A.; Perruolo, G.; Cristiano, S.; Agognon, A.L.; Cabaro, S.; Amato, A.; Dipineto, L.; Borrelli, L.; Formisano, P.; Fioretti, A.; et al. SARS-CoV-2 Affects Both Humans and Animals: What Is the Potential Transmission Risk? A Literature Review. *Microorganisms* **2023**, *11*, 514. [\[CrossRef\]](#)
- Spernovasilis, N.; Tsiodras, S.; Poulakou, G. Emerging and Re-Emerging Infectious Diseases: Humankind's Companions and Competitors. *Microorganisms* **2022**, *10*, 98. [\[CrossRef\]](#)
- Akhter, S.; Akhtar, S. Emerging coronavirus diseases and future perspectives. *VirusDisease* **2020**, *31*, 113. [\[CrossRef\]](#)
- da Costa, V.G.; Moreli, M.L.; Saivish, M.V. The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. *Arch. Virol.* **2020**, *165*, 1517. [\[CrossRef\]](#)
- Pollard, C.A.; Morran, M.P.; Nestor-Kalinowski, A.L. The COVID-19 pandemic: A global health crisis. *Physiol. Genom.* **2020**, *52*, 549–557. [\[CrossRef\]](#)
- Jeanne, L.; Bourdin, S.; Nadou, F.; Noiret, G. Economic globalization and the COVID-19 pandemic: Global spread and inequalities. *GeoJournal* **2023**, *88*, 1181. [\[CrossRef\]](#)
- Telenti, A.; Arvin, A.; Corey, L.; Corti, D.; Diamond, M.S.; García-Sastre, A.; Garry, R.F.; Holmes, E.C.; Pang, P.S.; Virgin, H.W. After the pandemic: Perspectives on the future trajectory of COVID-19. *Nature* **2021**, *596*, 495–504. [\[CrossRef\]](#)
- Chen, Y.; Liu, Q.; Zhou, L.; Zhou, Y.; Yan, H.; Lan, K. Emerging SARS-CoV-2 variants: Why, how, and what's next? *Cell Insight* **2022**, *1*, 100029. [\[CrossRef\]](#)
- Mahilkar, S.; Agrawal, S.; Chaudhary, S.; Parikh, S.; Sonkar, S.C.; Verma, D.K.; Chitalia, V.; Mehta, D.; Koner, B.C.; Vijay, N.; et al. SARS-CoV-2 variants: Impact on biological and clinical outcome. *Front. Med.* **2022**, *9*, 995960. [\[CrossRef\]](#)
- Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367*, 1444–1448. [\[CrossRef\]](#)
- Callaway, E. Delta coronavirus variant: Scientists brace for impact. *Nature* **2021**, *595*, 17–18. [\[CrossRef\]](#)
- Starr, T.N.; Greaney, A.J.; Hilton, S.K.; Ellis, D.; Crawford, K.H.; Dingens, A.S.; Navarro, M.J.; Bowen, J.E.; Tortorici, M.A.; Walls, A.C.; et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182*, 1295–1310.e20. [\[CrossRef\]](#)
- Volz, E.; Mishra, S.; Chand, M.; Barrett, J.C.; Johnson, R.; Geidelberg, L.; Hinsley, W.R.; Laydon, D.J.; Dabrera, G.; O'Toole, Á.; et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **2021**, *593*, 266–269. [\[CrossRef\]](#)
- Brown, K.A.; Gubbay, J.; Hopkins, J.; Patel, S.; Buchan, S.A.; Daneman, N.; Goneau, L.W. S-Gene Target Failure as a Marker of Variant B.1.1.7 among SARS-CoV-2 Isolates in the Greater Toronto Area, December 2020 to March 2021. *JAMA* **2021**, *325*, 2115–2116. [\[CrossRef\]](#)
- Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.D.; Mishra, S.; Crispim, M.A.; Sales, F.C.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [\[CrossRef\]](#)
- Wise, J. COVID-19: The E484K mutation and the risks it poses. *BMJ* **2021**, *372*, n359. [\[CrossRef\]](#)
- Ramanathan, M.; Ferguson, I.D.; Miao, W.; Khavari, P.A. SARS-CoV-2 B.1.1.7 and B.1.351 spike variants bind human ACE2 with increased affinity. *Lancet Infect. Dis.* **2021**, *21*, 1070. [\[CrossRef\]](#)
- Di Giacomo, S.; Mercatelli, D.; Rakhimov, A.; Giorgi, F.M. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. *J. Med. Virol.* **2021**, *93*, 5638–5643. [\[CrossRef\]](#)
- Viana, R.; Moyo, S.; Amoako, D.G.; Tegally, H.; Scheepers, C.; Althaus, C.L.; Anyaneji, U.J.; Bester, P.A.; Boni, M.F.; Chand, M.; et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **2022**, *603*, 679–686. [\[CrossRef\]](#)
- Harvey, W.T.; Carabelli, A.M.; Jackson, B.; Gupta, R.K.; Thomson, E.C.; Harrison, E.M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S.J.; et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424. [\[CrossRef\]](#)
- Shrestha, L.B.; Foster, C.; Rawlinson, W.; Tedla, N.; Bull, R.A. Evolution of the SARS-CoV-2 omicron variants BA.1 to BA.5: Implications for immune escape and transmission. *Rev. Med. Virol.* **2022**, *32*, e2381. [\[CrossRef\]](#)
- Maqsood, T.; Shahzad, K.; Naz, S.; Simsek, S.; Afzal, M.S.; Ali, S.; Ahmed, H.; Cao, J. A Cross-Sectional Study on the Association between Risk Factors of Toxoplasmosis and One Health Knowledge in Pakistan. *Front. Vet. Sci.* **2021**, *8*, 1390. [\[CrossRef\]](#)
- Ren, W.; Zhang, Y.; Rao, J.; Wang, Z.; Lan, J.; Liu, K.; Zhang, X.; Hu, X.; Yang, C.; Zhong, G.; et al. Evolution of Immune Evasion and Host Range Expansion by the SARS-CoV-2 B.1.1.529 (Omicron) Variant. *MBio* **2023**, *14*, e00416-23. [\[CrossRef\]](#)
- Karim, S.S.A.; Karim, Q.A. Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic. *Lancet* **2021**, *398*, 2126–2128. [\[CrossRef\]](#)



26. Park, A.K.; Kim, I.H.; Lee, C.Y.; Kim, J.A.; Lee, H.; Kim, H.M.; Lee, N.J.; Woo, S.; Lee, J.; Rhee, J.; et al. Rapid Emergence of the Omicron Variant of Severe Acute Respiratory Syndrome Coronavirus 2 in Korea. *Ann. Lab. Med.* **2023**, *43*, 211. [\[CrossRef\]](#)
27. Dhama, K.; Nainu, F.; Frediansyah, A.; Yatoo, M.I.; Mohapatra, R.K.; Chakraborty, S.; Zhou, H.; Islam, M.R.; Mamada, S.S.; Kusuma, H.I.; et al. Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies. *J. Infect. Public Health* **2023**, *16*, 4–14. [\[CrossRef\]](#)
28. Carabelli, A.M.; Peacock, T.P.; Thorne, L.G.; Harvey, W.T.; Hughes, J. SARS-CoV-2 variant biology: Immune escape, transmission and fitness. *Nat. Rev. Microbiol.* **2023**, *21*, 162–177.
29. Cameroni, E.; Bowen, J.E.; Rosen, L.E.; Saliba, C.; Zepeda, S.K.; Culap, K.; Pinto, D.; VanBlargan, L.A.; De Marco, A.; di Iulio, J.; et al. Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **2021**, *602*, 664–670. [\[CrossRef\]](#)
30. Mwenda, M. Detection of B.1.351 SARS-CoV-2 Variant Strain—Zambia, December 2020. *MMWR Morb. Mortal. Wkly. Rep.* **2021**, *70*, 280–282. [\[CrossRef\]](#)
31. Tosta, S.; Moreno, K.; Schuab, G.; Fonseca, V.; Segovia, F.M.; Kashima, S.; Elias, M.C.; Sampaio, S.C.; Ciccozzi, M.; Alcantara, L.C.; et al. Global SARS-CoV-2 genomic surveillance: What we have learned (so far). *Infect. Genet. Evol.* **2023**, *108*, 105405. [\[CrossRef\]](#)
32. AlKalamouni, H.; Abou Hassan, F.F.; Bou Hamdan, M.; Page, A.J.; Lott, M.; Matthews, M.; Ghosn, N.; Rady, A.; Mahfouz, R.; Araj, G.F.; et al. Genomic surveillance of SARS-CoV-2 in COVID-19 vaccinated healthcare workers in Lebanon. *BMC Med. Genom.* **2023**, *16*, 14. [\[CrossRef\]](#)
33. Simulundu, E.; Mupeta, F.; Chanda-Kapata, P.; Saasa, N.; Changula, K.; Muleya, W.; Chitanga, S.; Mwanza, M.; Simusika, P.; Chambaro, H.; et al. First COVID-19 case in Zambia—Comparative phylogenomic analyses of SARS-CoV-2 detected in African countries. *Int. J. Infect. Dis.* **2021**, *102*, 455–459. [\[CrossRef\]](#)
34. Katowa, B.; Kalonda, A.; Mubemba, B.; Matoba, J.; Shempela, D.M.; Sikalima, J.; Kabungo, B.; Changula, K.; Chitanga, S.; Kasonde, M.; et al. Genomic Surveillance of SARS-CoV-2 in the Southern Province of Zambia: Detection and Characterization of Alpha, Beta, Delta, and Omicron Variants of Concern. *Viruses* **2022**, *14*, 1865. [\[CrossRef\]](#)
35. Akande, O.W.; Carter, L.L.; Abubakar, A.; Achilla, R.; Barakat, A.; Gumed, N.; Guseinova, A.; Inbanathan, F.Y.; Kato, M.; Koua, E.; et al. Strengthening pathogen genomic surveillance for health emergencies: Insights from the World Health Organization's regional initiatives. *Front. Public Health* **2023**, *11*, 1146730. [\[CrossRef\]](#)
36. Martin, D.P.; Murrell, B.; Khoosal, A.; Muhire, B. Detecting and Analyzing Genetic Recombination Using RDP4. *Methods Mol. Biol.* **2017**, *1525*, 433–460.
37. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [\[CrossRef\]](#)
38. Muhire, B.M.; Varsani, A.; Martin, D.P. SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. *PLoS ONE* **2014**, *9*, e108277. [\[CrossRef\]](#)
39. Mercatelli, D.; Triboli, L.; Fornasari, E.; Ray, F.; Giorgi, F.M. Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations. *J. Med. Virol.* **2021**, *93*, 3238–3245. [\[CrossRef\]](#)
40. Martin, D.P.; Varsani, A.; Roumagnac, P.; Botha, G.; Maslamoney, S.; Schwab, T.; Kelz, Z.; Kumar, V.; Murrell, B. RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **2021**, *7*, 87. [\[CrossRef\]](#)
41. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [\[CrossRef\]](#)
42. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; Von Haeseler, A.; Jermini, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [\[CrossRef\]](#)
43. Hoang, D.T.; Chernomor, O.; Von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [\[CrossRef\]](#)
44. Graudenzi, A.; Maspero, D.; Angaroni, F.; Piazza, R.; Ramazzotti, D. Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. *iScience* **2021**, *24*, 102116. [\[CrossRef\]](#)
45. Van Dorp, L.; Acman, M.; Richard, D.; Shaw, L.P.; Ford, C.E.; Ormond, L.; Owen, C.J.; Pang, J.; Tan, C.C.; Boshier, F.A.; et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **2020**, *83*, 104351. [\[CrossRef\]](#)
46. Wang, L.; Berger, N.A.; Kaelber, D.C.; Davis, P.B.; Volkow, N.D.; Xu, R. COVID infection rates, clinical outcomes, and racial/ethnic and gender disparities before and after Omicron emerged in the US. *medRxiv* **2022**. [\[CrossRef\]](#)
47. Mseka, U.L.; Mandolo, J.; Nyoni, K.; Divala, O.; Kambalame, D.; Mapemba, D.; Kamzati, M.; Chibwe, I.; Henrion, M.Y.; Manda, K.; et al. Omicron B.1.1.529 variant infections associated with severe disease are uncommon in a COVID-19 under-vaccinated, high SARS-CoV-2 seroprevalence population in Malawi. *eClinicalMedicine* **2023**, *56*, 101800. [\[CrossRef\]](#)
48. Taboada, B.; Zárate, S.; Iša, P.; Boukadida, C.; Vazquez-Perez, J.A.; Muñoz-Medina, J.E.; Ramírez-González, J.E.; Comas-García, A.; Grajales-Muñoz, C.; Rincón-Rubio, A.; et al. Genetic Analysis of SARS-CoV-2 Variants in Mexico during the First Year of the COVID-19 Pandemic. *Viruses* **2021**, *13*, 2161. [\[CrossRef\]](#)
49. Bwire, G.M. Coronavirus: Why Men are More Vulnerable to COVID-19 Than Women? *Sn Compr. Clin. Med.* **2020**, *2*, 874. [\[CrossRef\]](#)
50. Miao, Y.; Ren, Y.; Ren, T. Clinical Characteristics Profile of COVID-19 Patients with Omicron Variant Admitted in a Tertiary Hospital, Central China. *Int. J. Gen. Med.* **2023**, *16*, 2365. [\[CrossRef\]](#)



51. Statsenko, Y.; Al Zahmi, F.; Habuza, T.; Almansoori, T.M.; Smetanina, D.; Simiyu, G.L.; Neidl-Van Gorkom, K.; Ljubisavljevic, M.; Awawdeh, R.; Elshekhali, H.; et al. Impact of Age and Sex on COVID-19 Severity Assessed from Radiologic and Clinical Findings. *Front. Cell Infect. Microbiol.* **2022**, *11*, 777070. [\[CrossRef\]](#)
52. Monod, M.; Blenkinsop, A.; Xi, X.; Hebert, D.; Bershan, S.; Tietze, S.; Baguelin, M.; Bradley, V.C.; Chen, Y.; Coupland, H.; et al. Age groups that sustain resurging COVID-19 epidemics in the United States. *Science* **2021**, *371*, eabe8372. [\[CrossRef\]](#)
53. Rahimi, A.; Mirzazadeh, A.; Tavakolpour, S. Genetics and genomics of SARS-CoV-2: A review of the literature with the special focus on genetic diversity and SARS-CoV-2 genome detection. *Genomics* **2021**, *113*, 1221–1232. [\[CrossRef\]](#)
54. Rehman, S.U.; Shafique, L.; Ihsan, A.; Liu, Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens* **2020**, *9*, 240. [\[CrossRef\]](#)
55. da Silva, B.W.; Felix, P.T. Analysis of the genetic diversity of SARS-CoV-2 genomes carrying the Omicron B.1.1.529 mutation. *medRxiv* **2022**. [\[CrossRef\]](#)
56. Soratto, T.A.; Darban, H.; Bjerkner, A.; Coorens, M.; Albert, J.; Allander, T.; Andersson, B. Four SARS-CoV-2 Genome Sequences from Late April in Stockholm, Sweden, Reveal a Rare Mutation in the Spike Protein. *Microbiol. Resour. Announc.* **2020**, *9*, e00934-20. [\[CrossRef\]](#)
57. Yin, C. Genotyping coronavirus SARS-CoV-2: Methods and implications. *Genomics* **2020**, *112*, 3588–3596. [\[CrossRef\]](#)
58. Pachetti, M.; Marini, B.; Benedetti, F.; Giudici, F.; Mauro, E.; Storici, P.; Masciovecchio, C.; Angeletti, S.; Ciccozzi, M.; Gallo, R.C.; et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* **2020**, *18*, 179. [\[CrossRef\]](#)
59. Lu, Y.; Wang, W.; Liu, H.; Li, Y.; Yan, G.; Franzo, G.; Dai, J.; He, W.T. Mutation and codon bias analysis of the spike protein of Omicron, the recent variant of SARS-CoV-2. *Int. J. Biol. Macromol.* **2023**, *250*, 126080. [\[CrossRef\]](#)
60. Ou, J.; Lan, W.; Wu, X.; Zhao, T.; Duan, B.; Yang, P.; Ren, Y.; Quan, L.; Zhao, W.; Seto, D.; et al. Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. *Signal Transduct. Target. Ther.* **2022**, *7*, 138. [\[CrossRef\]](#)
61. Saldivar-Espinoza, B.; Garcia-Segura, P.; Novau-Ferré, N.; Macip, G.; Martínez, R.; Puigbò, P.; Cereto-Massagué, A.; Pujadas, G.; Garcia-Vallve, S. The Mutational Landscape of SARS-CoV-2. *Int. J. Mol. Sci.* **2023**, *24*, 9072. [\[CrossRef\]](#)
62. Yu, Q.; Tong, X.; Zuo, L.; Tao, X.; Xu, Z.; Li, X.; Liu, H.; Guan, W.; Liu, D.; Liu, H.; et al. Genomic Surveillance of SARS-CoV-2 Variants That Emerged in South and Southeast Asia during Early 2022. *Viruses* **2023**, *15*, 1355. [\[CrossRef\]](#)
63. Pavlova, N.I.; Bocharov, A.A.; Alekseev, V.A.; Diakonova, A.T.; Dodokhov, V.V.; Kurtanov, K.A. Frequency of the Risk A Allele of rs17713054 Localized in the 3p21.31 COVID-19 Risk Locus in the Yakut Population. *Int. J. Biomed.* **2022**, *12*, 155–159. [\[CrossRef\]](#)
64. Suzuki, T.; Aizawa, K.; Shibuya, K.; Yamanaka, S.; Anzai, Y.; Kurokawa, K.; Nagai, R. Prevalence of Asymptomatic SARS-CoV-2 Infection in Japan. *JAMA Netw. Open* **2022**, *5*, e2247704. [\[CrossRef\]](#)
65. Lundberg, A.L.; Lorenzo-Redondo, R.; Ozer, E.A.; Hawkins, C.A.; Hultquist, J.F.; Welch, S.B.; Prasad, P.V.; Oehmke, J.F.; Achenbach, C.J.; Murphy, R.L.; et al. Has Omicron Changed the Evolution of the Pandemic? *JMIR Public Health Surveill.* **2022**, *8*, e35763. [\[CrossRef\]](#)
66. Ke, H.; Chang, M.R.; Marasco, W.A. Immune Evasion of SARS-CoV-2 Omicron Subvariants. *Vaccines* **2022**, *10*, 1545. [\[CrossRef\]](#)
67. Planas, D.; Bruel, T.; Staropoli, I.; Guivel-Benhassine, F.; Porrot, F.; Maes, P.; Grzelak, L.; Prot, M.; Mougari, S.; Planchais, C.; et al. Resistance of Omicron subvariants BA.2.75.2, BA.4.6, and BQ.1.1 to neutralizing antibodies. *Nat. Commun.* **2023**, *14*, 824. [\[CrossRef\]](#)
68. da Costa, C.H.S.; de Freitas, C.A.B.; Alves, C.N.; Lameira, J. Assessment of mutations on RBD in the Spike protein of SARS-CoV-2 Alpha, Delta and Omicron variants. *Sci. Rep.* **2022**, *12*, 8540. [\[CrossRef\]](#)
69. Adhikari, P.; Jawad, B.; Podgornik, R.; Ching, W.Y. Mutations of Omicron Variant at the Interface of the Receptor Domain Motif and Human Angiotensin-Converting Enzyme-2. *Int. J. Mol. Sci.* **2022**, *23*, 2870. [\[CrossRef\]](#)
70. Peka, M.; Balatsky, V. The impact of mutation sets in receptor-binding domain of SARS-CoV-2 variants on stability of the RBD-ACE2 complex. *Future Virol.* **2023**, *18*, 225–242. [\[CrossRef\]](#)
71. Chakraborty, C.; Bhattacharya, M.; Sharma, A.R.; Mallik, B. Omicron (B.1.1.529)—A new heavily mutated variant: Mapped location and probable properties of its mutations with an emphasis on S-glycoprotein. *Int. J. Biol. Macromol.* **2022**, *219*, 980–997. [\[CrossRef\]](#)
72. Davies, N.G.; Abbott, S.; Barnard, R.C.; Jarvis, C.I.; Kucharski, A.J.; Munday, J.D.; Pearson, C.A.; Russell, T.W.; Tully, D.C.; Washburne, A.D.; et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **2021**, *372*, eabg3055. [\[CrossRef\]](#)
73. Alenquer, M.; Ferreira, F.; Lousa, D.; Valério, M.; Medina-Lopes, M.; Bergman, M.L.; Gonçalves, J.; Demengeot, J.; Leite, R.B.; Lilue, J.; et al. Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies. *PLoS Pathog.* **2021**, *17*, e1009772. [\[CrossRef\]](#)
74. Ching, W.Y.; Adhikari, P.; Jawad, B.; Podgornik, R. Effect of Delta and Omicron Mutations on the RBD-SD1 Domain of the Spike Protein in SARS-CoV-2 and the Omicron Mutations on RBD-ACE2 Interface Complex. *Int. J. Mol. Sci.* **2022**, *23*, 10091. [\[CrossRef\]](#)
75. Patil, S.; Alzahrani, K.J.; Banjer, H.J.; Halawani, I.F.; Alzahrani, H.; Altayar, M.A.; Albogami, S.; Angeles, R.F.; Hassan, A.A.; Bhandi, S.; et al. Receptor binding domain of SARS-CoV-2 from Wuhan strain to Omicron B.1.1.529 attributes increased affinity to variable structures of human ACE2. *J. Infect. Public Health* **2022**, *15*, 781–787. [\[CrossRef\]](#)

76. Shah, M.; Woo, H.G. Omicron: A Heavily Mutated SARS-CoV-2 Variant Exhibits Stronger Binding to ACE2 and Potently Escapes Approved COVID-19 Therapeutic Antibodies. *Front. Immunol.* **2022**, *12*, 830527. [\[CrossRef\]](#)
77. Mannar, D.; Saville, J.W.; Zhu, X.; Srivastava, S.S.; Berezuk, A.M.; Tuttle, K.S.; Marquez, A.C.; Sekirov, I.; Subramaniam, S. SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex. *Science* **2022**, *375*, 760–764. [\[CrossRef\]](#)
78. Tong, C.; Shi, W.; Zhang, A.; Shi, Z. Tracking and controlling the spatiotemporal spread of SARS-CoV-2 Omicron variant in South Africa. *Travel. Med. Infect. Dis.* **2022**, *46*, 102252. [\[CrossRef\]](#)
79. Eisen, J.A. Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* **1998**, *8*, 163–167. [\[CrossRef\]](#)
80. Santiago, G.A.; Volkman, H.R.; Flores, B.; González, G.L.; Charriez, K.N.; Huertas, L.C.; Van Belleghem, S.M.; Rivera-Amill, V.; Major, C.; Colon, C.; et al. SARS-CoV-2 Omicron Replacement of Delta as Predominant Variant, Puerto Rico. *Emerg. Infect. Dis.* **2023**, *29*, 855–857. [\[CrossRef\]](#)
81. Robles-Escajeda, E.; Mohl, J.E.; Contreras, L.; Betancourt, A.P.; Mancera, B.M.; Kirken, R.A.; Rodriguez, G. Rapid Shift from SARS-CoV-2 Delta to Omicron Sub-Variants within a Dynamic Southern U.S. Borderplex. *Viruses* **2023**, *15*, 658. [\[CrossRef\]](#)
82. Vauhkonen, H.; Nguyen, P.T.; Kant, R.; Plyusnin, I.; Erdin, M.; Kurkela, S.; Liimatainen, H.; Ikonen, N.; Blomqvist, S.; Liitsola, K.; et al. Introduction and Rapid Spread of SARS-CoV-2 Omicron Variant and Dynamics of BA.1 and BA.1.1 Sublineages, Finland, December 2021. *Emerg. Infect. Dis.* **2022**, *28*, 1229–1232. [\[CrossRef\]](#)
83. Sharma, D.; Notarte, K.I.; Fernandez, R.A.; Lippi, G.; Gromiha, M.M.; Henry, B.M. In silico evaluation of the impact of Omicron variant of concern sublineage BA.4 and BA.5 on the sensitivity of RT-qPCR assays for SARS-CoV-2 detection using whole genome sequencing. *J. Med. Virol.* **2023**, *95*, e28241. [\[CrossRef\]](#)
84. Chibwana, M.G.; Thole, H.W.; Anscombe, C.; Ashton, P.M.; Green, E.; Barnes, K.G.; Cornick, J.; Turner, A.; Witte, D.; Nthala, S.; et al. Different clinical features in Malawian outpatients presenting with COVID-19 prior to and during Omicron variant dominance: A prospective observational study. *PLOS Glob. Public Health* **2023**, *3*, e0001575. [\[CrossRef\]](#)
85. Simwanza, J.; Hines, J.Z.; Sinyange, D.; Sinyange, N.; Mulenga, C.; Hanyinza, S.; Sakubita, P.; Langa, N.; Nowa, H.; Gardner, P.; et al. COVID-19 Vaccine Effectiveness during a Prison Outbreak when Omicron was the Dominant Circulating Variant–Zambia, December 2021. *Am. J. Trop. Med. Hyg.* **2022**, *107*, 1055–1059. [\[CrossRef\]](#)
86. Zambia Cenral Statistical Office; Government of the Republic of Zambia. Census of Population and Housing–2022. 2022. Available online: <https://www.zamstats.gov.zm/population-by-province/> (accessed on 1 February 2024).
87. Tcherepanov, V.; Ehlers, A.; Upton, C. Genome annotation transfer utility (GATU): Rapid annotation of viral genomes using a closely related reference genome. *BMC Genom.* **2006**, *7*, 150. [\[CrossRef\]](#)
88. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [\[CrossRef\]](#)

**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.