Microbes and Infectious Diseases 2023; 4(4): 1088-1099

Microbes and Infectious Diseases

Journal homepage: https://mid.journals.ekb.eg/

Original article

Characterization of SARS-CoV-2 genetic diversity in Zambia: single center study

Emmanuel Zulu[1], Rodrick S Katete[1,2,3,*], Paul Simusika[1], Mable M. Mutengo[1], Freeman W. Chabala[1], Mpanga Kasonde[4]

1- *1Institute of Basic and Biomedical Sciences, Levy Mwanawasa Medical University, Lusaka, Zambia*
2- *Department of Biological Sciences, Faculty of Science, Technology and Innovation, Mzuzu University, Luwinga, Malawi*
3- *Department of Biological Sciences, School of Mathematics and Natural Sciences, Mukuba University, Kitwe, Zambia*
4- *Genome Sequencing and Bioinformatics Unit, Zambia National Public Health Institute (ZNPHI), Lusaka, Zambia.*

ARTICLE INFO

Article history:
Received 15 August 2023
Received in revised form 7 September 2023
Accepted 9 September 2023

Keywords:
COVID-19
Variants of concern
Mutation
Genome sequencing
Phylogenetic tree
Genomic surveillance

ABSTRACT

Background: The emergency of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2) variants has led to COVID-19 worldwide with substantial social and economic consequences. SARS CoV-2 S-glycoproteins are involved in viral entry to human cells. They are naturally the main target of host immune responses, and most vaccine designs are based on them. SARS-CoV-2 variants are categorized as variants of interest (VOI) or variants of concern (VOC) because they are deadly and highly transmissible, causing illness by overcoming the host immune system. **Material and Methods:** The COVID-19 patients' samples were isolated for genome sequencing to identify the mutations that alter the viral genotypic traits. Genome annotation and phylogenetic analysis were performed using MEGA 7.0. **Results:** Phylogenetic analysis revealed that the omicron variant of concern and sub-variants XBB, XBB.2.6, BQ.1, and BQ.1.1 are the most prevalent variants in Lusaka, Zambia. Analysis of the translated protein sequences in this study revealed D614G mutation in all the sequences. This mutation has previously been implicated in viral transmission, increasing the infectivity, replication efficacy, stability of virions, and virulence in its human host. **Conclusion:** Analysis of the SARS-CoV-2 genome provided crucial information on the variant and its source as it spreads in Lusaka from person to person. This research has a huge potential in genomic epidemiology, where genomic surveillance is employed to detect new mutations or SARS-CoV-2 variants, which can help the virus to spread rapidly, increase disease severity, or even evade vaccine-induced immunity.

Introduction

The Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2) comes from a family of coronaviruses that infect humans and other animals such as bats, civets, and camels [1]. In humans, this virus causes severe respiratory diseases and common colds. The SARS-CoV-2 comprises a large positive single-stranded RNA genome enclosed in a nucleocapsid surrounded by an envelope membrane of lipids and proteins [2,3]. The coronaviruses that infect humans belong to the beta-coronaviruses. These viruses can cause respiratory tract infections in mammals. Mutations can occur randomly due to antigenic drift and error-prone RNA polymerase, producing viruses that can better survive because of natural selection and can cause disease outbreaks in humans. This is a result of the transmission of the virus from species to species, which also leads to the emergence of different variants. The first coronavirus infection occurred in Guangdong Province, China, in 2002/2003, caused by the severe

DOI: 10.21608/MID.2023.229457.1592
* *Corresponding author:* Rodrick S Katete
E-mail address: *rskatete@gmail.com*

acute respiratory syndrome coronavirus (SARS-CoV) [4,5]. This virus was also a beta-coronavirus transmitted from bats to civets to humans. The second outbreak occurred in 2012 in Saudi Arabia, caused by the Middle East Respiratory Syndrome (MERS-CoV), also a beta-coronavirus. These viruses are transmitted from camels to humans; the main transmission route is eating camel meat or drinking camel milk [6]. The first outbreak of COVID-19 occurred in China Wuhan, in Hubei Province, in 2019 [7]. A new type of pneumonia emerged in Wuhan, the capital of Hubei Province in the Republic of China, and had spread throughout the country by the end of 2019 [7]. And by January 2020, it spread across all the provinces in China. The number of cases rose rapidly during this period, and the same cases were imported to other world regions. By 23rd March 2020, over 190 countries reported about 332,218 laboratory-confirmed cases and 14,510 deaths [8]. In Africa, the first case of COVID-19 was reported from Egypt around February 2020, prompting many African countries to prepare for the arrival of COVID-19 [9]. As the spread of the COVID-19 disease continued, Zambia was poised to receive its first case and assembled a response system that included the Public Health Operations Center (PHEOC). In March 2020, the first Zambian COVID-19 case was reported from a couple who had returned from France and were put on quarantine for about 14 days [10]. The disease has since spread throughout the country.

The SARS-CoV-2 genome is a positive-sense, single-stranded genome that is linear and composed of about 30,000 nucleotides (approximately 30kb) with a 5' cap and a poly-A tail [11]. This genome has eleven coding regions or genes and twelve potential gene products. The genetic variation in different geographical areas indicates genetic diversity in the virus, attributed to natural selection driven by long periods of an evolutionary arms race between the host immunity and the virus. The virus genome analysis illustrates that the emerging variants carry multiple mutations in the Spike (S) glycoprotein and the receptor-binding domain. There is a view that some of these mutations are biologically important and the result of adaptive evolution [12]. The SARS-CoV-2 S-glycoprotein impacts viruses by affecting transmissibility through interaction with host cell-binding receptors and contributing to immune escape via the receptor binding domain (RBD) **(Figure 1)**. The most important host protein involved in SARS-CoV-2 entry was identified as angiotensin-converting enzyme 2 (ACE2) and cell surface transmembrane protease serine 2 (TMPRSS2) [13]. Multiple emerging variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), increased infectiousness, severity, immune evasion, treatment failure, and reduced vaccine efficacy [14,15]. All variants are a product of random mutations, increasing abundance through person-to-person or species-to-species transmission. They are usually highly contagious, can evade immunity, and can spread rapidly in any given population.

The SARS-CoV-2 variant is defined as a variant of concern when mutations (e.g., N501Y, as shown in **Figure (1)**) in the spike protein's receptor binding domain (RBD) cause strong binding affinity to angiotensin-converting enzyme 2 (ACE2) and rapid transmission from person to person. The most common variants of concern are the British variant (B.1.1.7), the South African variant (B.1.351), and the Brazilian variant (P.1). These variants have the following mutations, E484K, which helps the virus evade the immune system by evading antibodies, P681H, which enhances virus entry into the cell, Y144del, which decreases antibody binding affinity, N501Y, which increases binding affinity to ACE2 receptor and D614G, which reduces S1 shedding, thereby increasing the infectivity, replication efficacy, viral transmission, stability of virions, and virulence in its human host **(Figure 1)** [17,18]. B.1.351 also has a K417N mutation that reduces the virus's susceptibility to antibody reactivity (Figure 1) [19]. The omicron variant B.1.1.529 has multiple mutations in the spike protein, so it is highly transmittable and can evade neutralization by monoclonal antibodies [20].

This research elucidated the genetic diversity of SARS-CoV-2 circulating in Zambia during the COVID-19 pandemic using genome sequencing to identify the mutations that alter the viral genotypic and phenotypic traits.

Material and methods

Study participants

This cross-section study quantitatively analyzed 51 SARS-CoV-2 genomes to detect mutations associated with viral pathogenicity. The study described SARS-CoV-2 genetic diversity from December 2022 to February 2023. It involved using nasopharyngeal swab samples from COVID-19 patients brought to the Zambia National Public

Health Reference Laboratory (ZNPHIL) from various facilities around Lusaka province for quantitative reverse transcription polymerase chain reaction (RT-qPCR) (Luna® Universal One-Step RT-qPCR Kit, New England Biolabs, Inc, Ipswich, Massachusetts, USA) analysis to confirm the COVID-19 cases. Samples obtained from the following categories were eligible for the study: hospitalized patients, healthcare workers facing patients, academic staff, residents, essential care workers, University students, patients in the emergency department, and outpatients who just tested positive for COVID-19. All the samples were anonymized and assigned distinct ID numbers. All participants provided verbal consent. Participation was entirely voluntary, and only non-invasive samples were used for sequencing. The inclusion criteria were COVID-19 samples that tested positive in both the rapid antigen test, SD Biosensor STANDARD Q COVID-19 Ag-RDT Kit (SD Biosensor; Republic of Korea) and the RT-qPCR test, regardless of vaccination status, was part of the study as there is a high likelihood of detecting mutations in samples from these patients, and sequences coverage of $\geq 80\%$ and no duplications or shorter sequences, and the exclusion criteria were COVID-19 positive samples with a cycle threshold of 30 or greater (Ct value > 30), these were considered to have a lower viral load hence less likely to display any mutations.

Data collection

Clinical data for COVID-19 cases were obtained from medical records from Zambia National Public Health Reference Laboratories in Lusaka Province. The data records were anonymously matched with the sample IDs for downstream analysis.

SARS-CoV-2 Genomic Isolation The SARS-CoV-2 genome was extracted from the patient's nasopharyngeal swab samples using the QIAseq DIRECT SARS-CoV-2 (Qiagen GmbH, Hilden, Germany). The extracted genomic RNA was quantified on a Qubit1 2.0 Fluorometer using the Qubit RNA HS assay kit (Life Technologies, Carlsbad, California, USA) according to the manufacturer's protocol. QuantStudio™ 5 Real-Time PCR System (Applied Biosystems, Foster City, USA) was used to detect the presence of the purified SARS-CoV-2 using AccuPower® SARS-CoV-2 Real-Time RT-PCR Kit (Bioneer, South Korea) of the fluorescence RdRp and E genes according to manufacturer's protocol.

Next generation sequencing

Library preparation

The assay library for sequencing the SARS-CoV-2 genome was prepared according to the manufacturer's protocol. Briefly, the whole genome was amplified using the CleanPlex® SARS-CoV-2 Research and Surveillance NGS Panel (Paragon Genomics Inc., Hayward, CA,USA). The cDNA library of the SARS-CoV-2 genome was generated from the extracted RNA using RT Primer Mix DP. The RT reaction mixture was cleaned and purified using the CleanMag® Magnetic Beads (Paragon Genomics Inc., Hayward, CA, USA). Two multiplex PCR (mPCR) reactions were performed using specific primers to amplify the whole SARS-CoV-2 genome by a 2-pool design. The mPCR was digested and purified so that the next set of primers, i5 and i7, could amplify the generated libraries for Illumina sequencing platforms (Illumina, San Diego, CA, USA). The libraries were visualized on agarose gel-electrophoresis where the fragment size of ~ 275 bp was obtained with the final concentration of 2.0 ng/µl as measured by the Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The cDNA samples obtained were stored at −20°C until further use. The library was bound to magnetic beads, so little fragments were washed away. The next step was quantification and normalization, where the 2 µl library pool was analyzed using a Qubit dsDNA HS assay kit (Thermo Fisher Scientific, Waltham, MA, USA), molarity was calculated, and the library pool was diluted to reach the required concentration of 1 nM. Then, a 5 µl sample volume was taken for denaturation and subsequently hybridized. A final volume of 500 µl was loaded into the Illumina Miniseq sequencing instrument (Illumina, Inc., San Diego, CA, USA), setting 2 x 150 cycles. During the analysis, the base calls were made by measuring the intensity of the signals presented during each sequencing cycle of the instrument.

Genome annotation and phylogenetic analysis

The FASTQ files of the Next Generation Sequences that were generated from MiSeq Local Run Manager (Illumina, San Diego, CA, USA) were further analyzed for quality measurements, re-alignment of indels, variant call review, trimming of adaptors, and determination of the consensus genome by mapping to reference sequence Wuhan-Hu-1/2019 EPI_ISL_402125. Whole genome

sequences were annotated by using the reference genome of Wuhan-Hu-1/2019 EPI_ISL_402125. The spike genome region was cut off from the entire genome because it was the area of interest for this study. After that, data set from the spike region (spike genome) only was created, which included 24 sequences generated from this research, 24 sequences generated from the National Centre for Bioinformatics Information (NCBI) dataset, and 2 reference sequences for the omicron variant, which originated from South Africa and Botswana, and 1 reference sequence which was the Wuhan sequence, Wuhan-Hu-1/2019 EPI_ISL_402125, thereby making a total of 51 sequences. Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to retrieve SARS-CoV-2 spike genome sequences most similar to the sequences generated from this study. Multiple sequence alignment was performed using the MUSCLE (MUltiple Sequence Comparison by Log Expectation) algorithms in MEGA 7.0 software (megasoftware.net) using default settings. Then, the alignment was inspected and cleaned. In the following alignment, a Maximum Likelihood (ML) phylogenetic tree was constructed using the MEGA 7.0 software using the bootstrap method with 1000 replicates. The branch support was estimated using the bootstrap values, and the ML tree was rooted on the Wuhan-Hu-1/2019 sequence. The PANGO lineage identification was performed using the pangolin on the Nextclade Web tool v.1.11.1 [21].

Diversity of the spike proteins

The SARS-CoV-2 spike protein is the primary target of the immune system and vaccine development. Mutation of the spike gene leads to different viral phenotypes. Therefore, the diversity or changes of the amino acids of the spike proteins were further analyzed because these mutations can help the virus with its antigenicity.

Ethical consideration / approval

The ethical approval for this study was obtained from Levy Mwanawasa Medical University Research Ethics Committee (approval number: REF. No. LMMU-REC 000014/22) and the ethical permission to conduct the study were obtained from the National Health Research Authority (NHRA) (Ref No: NHRA 0000014 /20/12/2022.)

Results

Multiple sequence alignment

Multiple sequence alignment of 51 sequences was performed using the MUSCLE algorithm in MEGA 7.0 software **(Figure 2)**. Therefore, a maximum likelihood (ML) phylogenetic tree was constructed in MEGA 7.0 software using 1000 bootstrap replications.

Phylogenetic analysis

The Phylogenetic analysis involved a downstream analysis of genomic sequences with the highest coverage of ≥80% and no duplications or shorter sequences [22,23,24]. A dataset of 51 spike genome sequences was created, which included 24 sequences generated from this study and 24 sequences retrieved from NCBI-BLAST. Also, reference sequences of omicron from South Africa and Botswana and the Wuhan-Hu-1-2019 sequences were included.

The lineages are labeled and highlighted in different colors **(Figure 3)**. The internal nodes display bootstrap values generated by the bootstrap method with 1000 replicates using MEGA 7.0. Phylogenetic analysis revealed the omicron as the SARS-CoV-2 variant of concern circulating in Lusaka, Zambia, as illustrated in the phylogenetic tree.

The evolutionary history was inferred using the Maximum Likelihood method based on the Tamura 3-Parameter model [25]. The tree with the highest log likelihood (-3695.50) is shown. The percentage of trees in which the associated taxa clustered is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.1000)). The rate variation model allowed some sites to be evolutionarily invariable ([+I], 49.32% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 51 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 2418 positions in the final dataset. Evolutionary studies were conducted in MEGA7 [26].

PANGO lineage classification

The following clade analysis clustered the 51 sequences into 15 PANGO as follows; Wuhan-Hu/2019 (1 sequence), BA.1 (2 sequences), BQ.1 (6 sequences), BQ.1.1 (4 sequences), BQ.1.1.44 (2 sequences), DB.1 (2 sequences), BA.4.1.8 (2 sequences), BA.5.2 (2 sequences), BF.19 (1 sequence), BN.1.3 (2 sequences), XBB (10 sequences), XBB.1 (2 sequences), XBB.2.6 (13 sequences), XBB.1.5.21 (1 sequence), and XBB.8.1 (1 sequence) **(Table 1)**.

Substitution mutation analysis in the protein sequence

After all the 51 nucleotide sequences were translated into proteins, a substitution mutational analysis was carried out at position 614, where it was observed that the amino acid D (Aspartate) in the Wuhan sequence was substituted by the amino acid G (Glycine) in all the omicron sequences thereby signifying a D614G mutation as illustrated in **Figure (4)**.

Table 1. The PANGO lineages classification of SARS-CoV-2 among the sequences

SN.	Clade	PANGO Lineage	Country of Origin for each Lineage	WHO Name	Number of Sequences
1.	19B	A.1	China, Wuhan (1)	Wuhan-Hu 1/2019	1
2	22E	BQ.1	Zambia, Lusaka-ZLU (3) USA, California-USA/CA (1) USA, Alabama-USA/NY (1) USA	omicron	6
3.	22B	DB.1	Zambia, Lusaka-ZLU (1) USA, Arizona-USA/AZ (1)	Omicron	2
4	22A	BA.4.1.8	Zambia, Lusaka-ZLU (1) USA, Colorado-USA/AZ (1)	Omicron	2
5.	22B	BA.5.2	Russia, Primorye-RUS (1) USA, Utah-USA/UT (1)	Omicron	2
6.	22B	BF.19	Zambia, Lusaka-ZLU (1)	Omicron	1
7.	22D	BN.1.3	Zambia, Lusaka-ZLU (1) USA, Arizona-USA/AZ (1)	Omicron	2
8.	22F	XBB	Zambia, Lusaka-ZLU (6) USA, Utah-USA/UT (1) USA, Michigan-USA/MI (1) Bahrain, BHR (1) India, IND (1)	Omicron	10
9.	22F	XBB.1	Brazil, BRA (1) USA/Utah-USA/UT (1)	Omicron	2
10.	22F	XBB.2.6	Zambia, Lusaka-ZLU (8) USA, Georgia-USA/GA (2) USA, Utah-USA/UT (3)	Omicron	13
11.	22F	XBB.1.5.21	USA, University of Washington-USA/WA-UV (1)	Omicron	1
12.	22F	XBB.8.1	USA, Utah-USA/UT (1)	Omicron	1
13.	21K	BA.1	South Africa, NCID (1) Botswana, BHP (1)	omicron	2
14.	22E	BQ.1.1.44	Zambia, Lusaka (1) USA, Utah-USA/UT (1)	omicron	2
15.	22E	BQ.1.1	Zambia, Lusaka-ZLU (2) USA, Minnesota (1) USA, Tennessee (1)	omicron	4

Figure 1. Structure of SARS-CoV-2 spike protein (PDB ID: 7y6d). The receptor-binding domain (RBD) binds to human angiotensin-converting enzyme 2 (ACE2). The ribbon diagram of SARS-CoV-2 was created using SwissPdb Viewer [16]

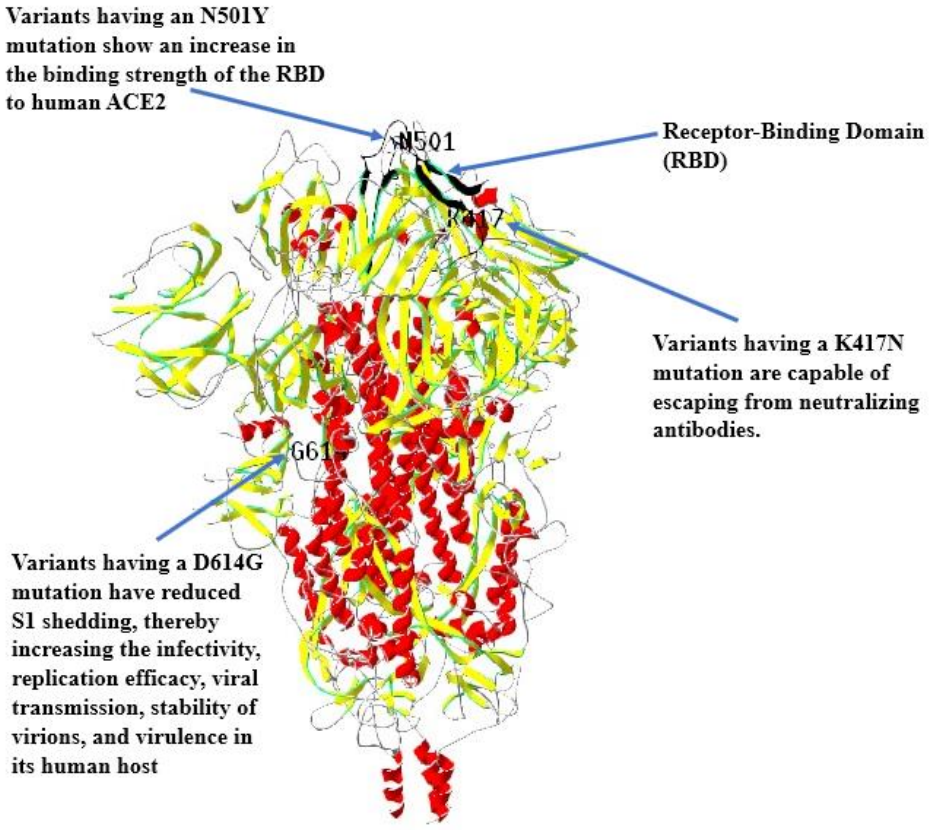


Figure 2. Multiple Sequence Alignment of SARS-CoV-2 showing aligned Zambian, South Africa, Botswana, Wuhan, USA, Russia, India, Bahrain, and Brazilian sequences. Mega 7.0 (MUSCLE Algorithm) program generated the multiple sequence alignment

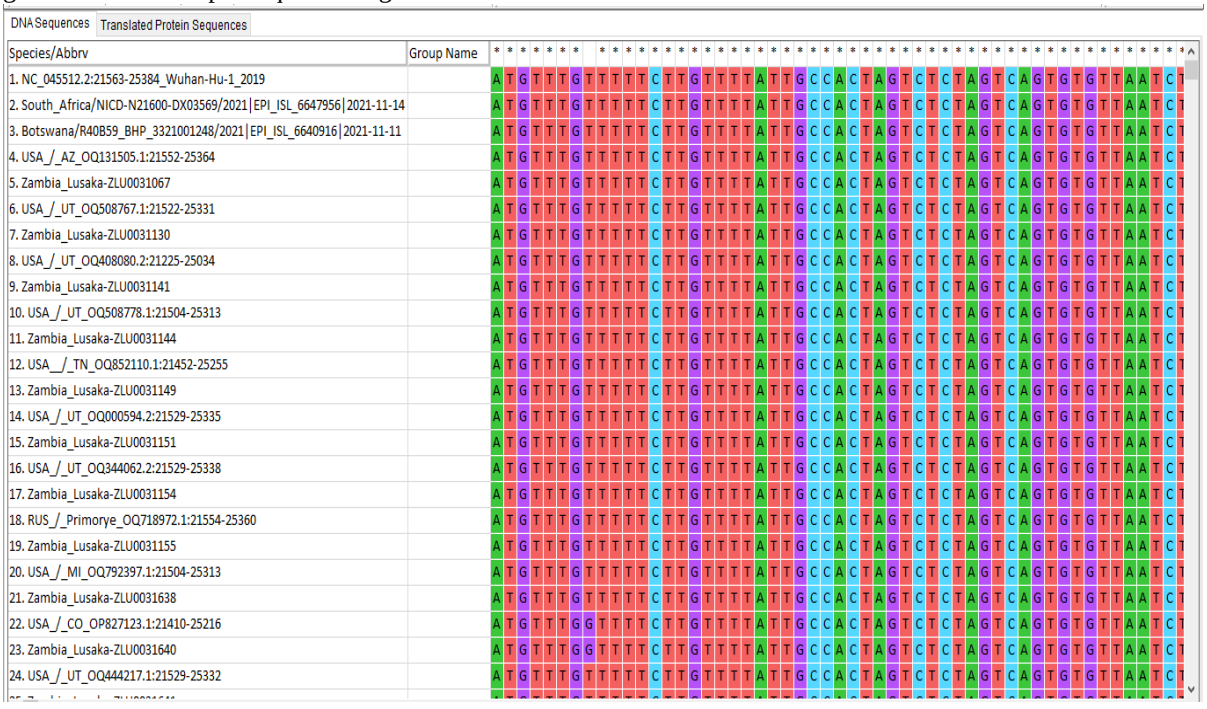


Figure 3. Neighbor-joining phylogenetic tree of the SARS-CoV-2 variants. Samples are colored by taxonomic affiliation to clades or subclades. Clades are assigned according to Nextclade nomenclature. The tree shows that the Zambian variants are more closely related to omicron variants commonly found in South Africa, Botswana, USA, Russia, India, Bahrain, and Brazil

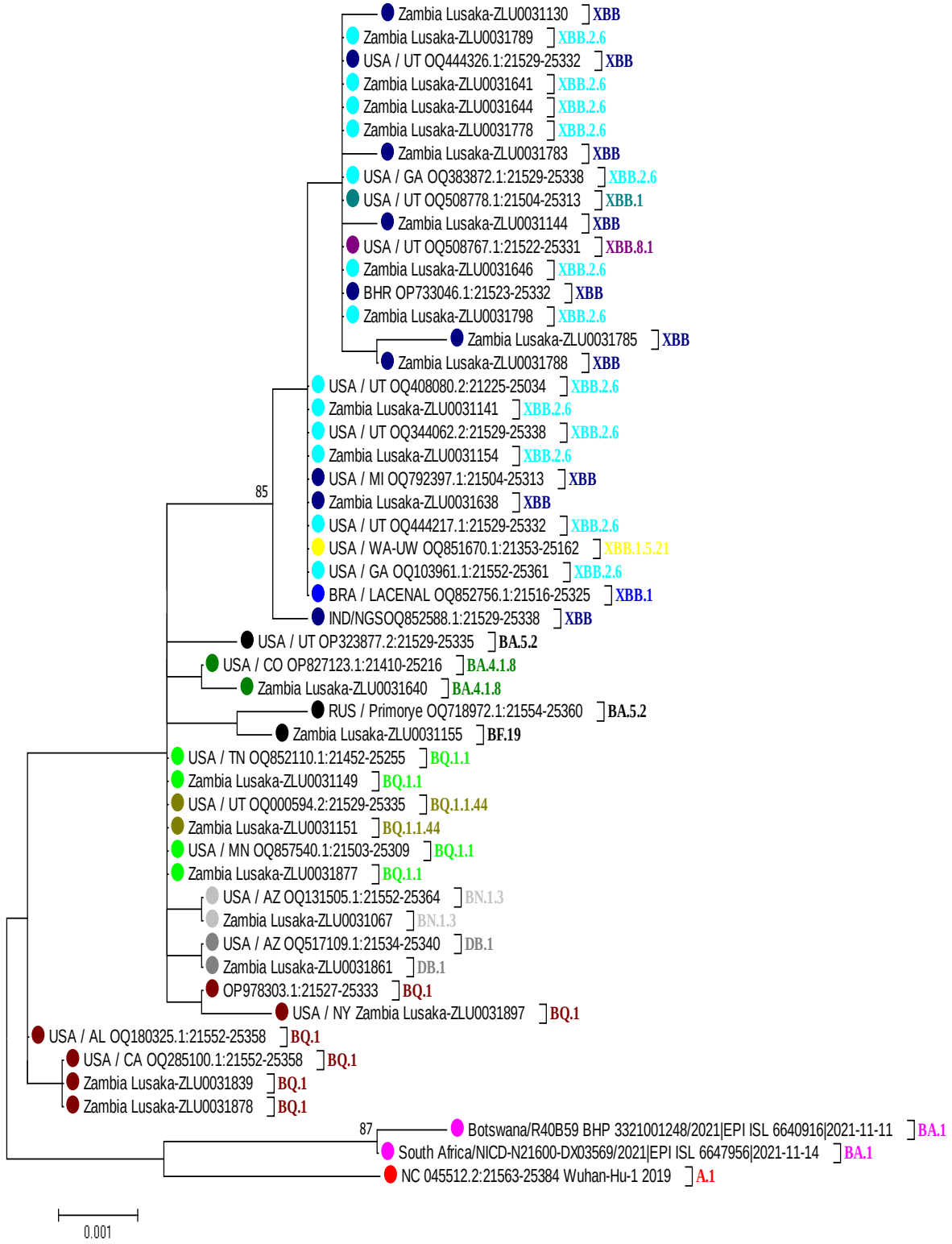
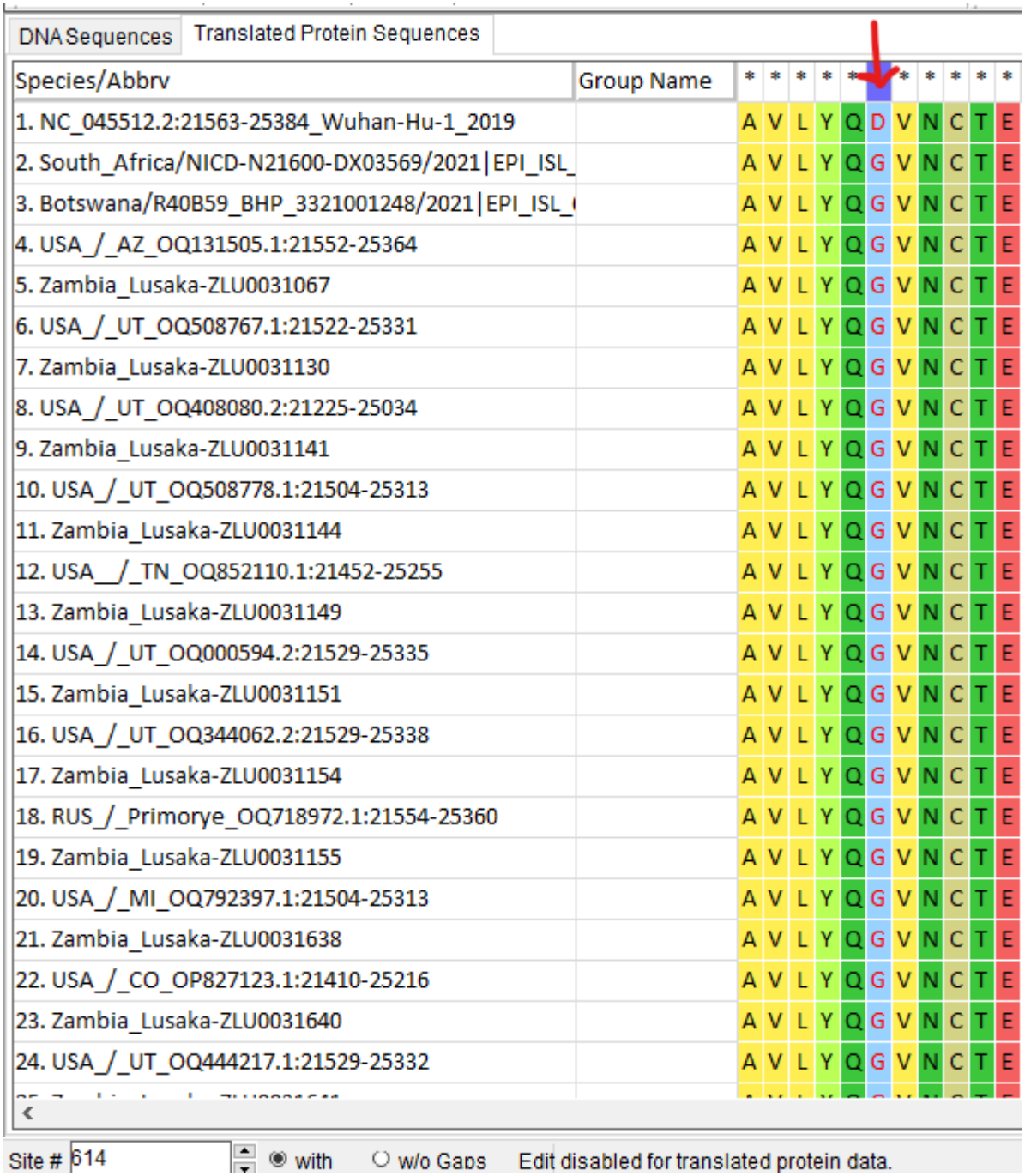


Figure. 4. Substitution Mutation Analysis in translated protein sequence of the SARS-CoV-2 showing the D614G mutation



Discussion

The study described SARS-CoV-2 genetic diversity in Lusaka, Zambia, from December 2022 to February 2023. The phylogenetic analysis involved a downstream analysis of 51 sequences which were from the spike genome region; this included 24 from the study, 24 from NCBI-BLAST, 2 were the omicron reference sequences, which were done in South Africa and Botswana, and 1 sequence was an out-group which was the Whan-hu-1 2019 sequence which was done in China during the first wave of COVID-19. All the sequences used for the phylogenetic analysis had better quality, with a genome coverage of 80% and above and longer sequences [22,23,24]. However, other sequences were not included for phylogenetic analysis due to sequence duplication after NCBI-BLAST, and some were shorter after cleaning. After that, multiple sequence alignment tools aligned the 51 spike

genome sequences using the MUSCLE algorithm in MEGA 7.0 Software. After this, the best-fit model was determined to be the Tamura-3 parameter model with gamma-distributed with invariant sites (T93 G+I). The evolutionary analysis among the sequences was estimated by the bootstrap method; a bootstrap of 1000 replicates was used to estimate the reliability of the tree. The nodes with bootstrap values greater than 70 were observed in the phylogenetic tree. This helped assess the nodes' reliability and identify which part of the tree was to be trusted. The sequence for the Wuhan-hu-1/2019 was the out-group for the phylogenetic tree, and the two sequences from South Africa and Botswana were the reference sequences for omicron. The maximum likelihood statistical method using a substitution method was used to assess the probability of particular mutations. Phylogenetic analysis revealed that all 24 genomes generated in the study belonged to the SARS-CoV-2 omicron variant. In comparison with other studies, it was observed that a phylogenetic tree gives a clear path of variant evolution, even if it lacks an indication of some variants being less widespread and does not detect intermediate variants [27,28]. Due to its increased viral transmissibility, the Omicron variant has presented alarming public health challenges during the COVID-19 pandemic. The other 24 genome sequences obtained from NCBI_BLAST were from different countries, such as various states in the USA, Russia, Brazil, and Bahrain in the Middle East. They showed a close relationship with the 24 sequences generated from the study, implying a possible multiple introduction or importation of the virus from different world regions. This was in agreement with various studies in other regions, which revealed that global human mobility is likely the crucial event in introducing new virus variants [29]. It was also observed that some Lusaka Zambia sequences clustered together, suggesting a local circulation of the variants within Lusaka Zambia. Phylogenetic analysis also revealed that all the sequences were the sub-variants of omicron BA.1, which supported the idea of the continued evolution of the omicron variant of concern. As observed, XXB and XBB.2.6 sub-variants of omicron were the latest at the time the study was conducted, around January to February 2023, all coming from the 22F clade (XBB and XBB.2.6 omicron variants), which had the higher prevalence.

Then, the PANGO analysis of 51 sequences was used to cluster the 51 sequences into the 15 PANGO lineages, as illustrated in **Table (1)**. The classification was used to identify lineages circulating in Lusaka, Zambia, the USA, Russia, Brazil, India, and Bahrain. It was observed that the PANGO lineage XBB (10 sequences) and XBB.2.6 (23 sequences) were the highest, followed by the BQ.1 (6 sequences) PANGO lineage. The PANGO classification is based on a nomenclature system that combines genetic and geographic information about the variant's dynamics. This was in agreement with what different latest studies have reported, that there were 16 subvariants of omicron identified from overseas, including XBB, BQ.1, and BQ.1.1. [30].

Mutation in the spike gene of SARS-CoV-2 has been found to play a vital role in viral transmission, replication efficacy, pathogenicity, and virulence in its human host [31]. Analysis of the translated protein sequences in this study revealed D614G mutations in all the sequences. Many studies report this mutation as associated with high viral load, infectivity, and transmissibility [32,33].

Limitation

The limitation of the study was that out of all the 116 samples initially tested positive for SARS-CoV-2 by RT-PCR and successfully sequenced genome, only 24 were used for phylogenetic downstream analysis. This was due to poor genome coverage; some were shorter and could not align properly because bioinformatic pipelines required computers with much higher operating systems. Some samples were not used because they had a Ct value > 30. This contributed to a small sample size that was less than the sample size estimated.

Conclusion

The study findings highlighted the circulation of the omicron variant of concern and sub-variants XBB, XBB.2.6, BQ.1, and BQ.1.1 as the most prevalent variants in Lusaka, Zambia. The phylogenetic analysis revealed that genome sequences generated from this study were closely related or similar to those generated from countries such as the USA, Russia, India, Brazil, and Bahrain in the Middle East. The sequences from Lusaka that formed a cluster suggested local spread, while those clustered with those from other countries suggested the international introduction of the variants in Lusaka, Zambia. The mutation analysis of the translated protein sequences shows a D614G mutation in all the sequences. This mutation is responsible for the higher viral transmissibility and

infectivity observed by many studies in the omicron variant [33].

Data availability

Some data supporting this study cannot be publicly shared for ethical or privacy reasons.

Conflict of interest

The authors declare no conflict of interest.

Declaration of funding

This research did not receive any specific funding.

Acknowledgment

The authors would like to thank colleagues in their respective universities for their moral support and encouragement. We thank the doctors and nurses of the Zambia National Public Health Reference Laboratory for their help in collecting the samples and COVID-19 data from symptomatic patients.

References

1- **Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al.** A new coronavirus associated with human respiratory disease in China. Nature 2020;579(7798), 265–269.

2- **V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V.** Coronavirus biology and replication: implications for SARS-CoV-2. Nature Reviews. Microbiology 2021;19(3), 155–170.

3- **Dërmaku-Sopjani M, Sopjani M.** Molecular Characterization of SARS-CoV-2. Current Molecular Medicine 2021;21(7), 589–595.

4- **Cherry JD, Krogstad P.** SARS: the first pandemic of the 21st century. Pediatric Research 2004;56(1), 1–5.

5- **Xu RH, He JF, Evans MR, Peng GW, Field HE, Yu DW, et al.** Epidemiologic clues to SARS origin in China. Emerging Infectious Diseases 2004;10(6), 1030–1037.

6- **Hui DS, Azhar EI, Kim YJ, Memish ZA, Oh MD, Zumla A.** Middle East respiratory syndrome coronavirus: risk factors and determinants of primary, household, and nosocomial transmission. The Lancet. Infectious Diseases 2018;18(8), e217–e227.

7- **Wu YC, Chen CS, Chan YJ.** The outbreak of COVID-19: An overview. Journal of the Chinese Medical Association: JCMA 2020;83(3), 217–220.

8- **Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, Duan G.** Virology, Epidemiology, Pathogenesis, and Control of COVID-19. Viruses 2020;12(4), 372.

9- **Khattab NM, Vermund SH, Hu Y.** How coronavirus disease 2019 entered Africa and the Middle East: a case study from Egypt. Transactions of the Royal Society of Tropical Medicine and Hygiene 2020;114(10), 715–717.

10- **Chipimo PJ, Barradas DT, Kayeyi N, Zulu PM, Muzala K, Mazaba ML, et al.** First 100 persons with COVID-19—Zambia, March 18–April 28, 2020. Morbidity and Mortality Weekly Report 2020;10;69(42):1547

11- **Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, Singh A, et al.** Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. Biochimica et Biophysica Acta. Molecular Basis of Disease 2020;1866(10), 165878.

12- **Aleem A, Akbar Samad AB, Vaqar S.** Emerging Variants of SARS-CoV-2 and Novel Therapeutics Against Coronavirus (COVID-19). In StatPearls. StatPearls Publishing; 2023.

13- **Beyerstedt S, Casaro EB, Rangel ÉB.** COVID-19: angiotensin-converting enzyme 2 (ACE2) expression and tissue susceptibility to SARS-CoV-2 infection. European Journal of Clinical Microbiology and Infectious Diseases: Official Publication of the European Society of Clinical Microbiology 2021;40(5), 905–919.

14- **Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, et al.** SARS-CoV-2 spike-

protein D614G mutation increases virion spike density and infectivity. Nature Communications 2020;11(1):6013.

15-**Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al.** The biological and clinical significance of emerging SARS-CoV-2 variants. Nature reviews. Genetics 2021;22(12), 757–773.

16-**Guex N, Peitsch MC.** SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 1997;18, 2714-2723.

17-**Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. (2021).** Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. bioRxiv: The Preprint Server for Biology 2021;01.25.428137.

18-**Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG. et al.** Spike mutation D614G alters SARS-CoV-2 fitness [published correction appears in Nature. 2021 Jul;595(7865):E1].Nature 2021 ;592 (7852) :116-121.

19-**Rayati Damavandi A, Dowran R, Al Sharif S, Kashanchi F, Jafari R.** Molecular variants of SARS-CoV-2: antigenic properties and current vaccine efficacy. Medical Microbiology and Immunology 2022;211(2-3), 79–103.

20-**Khan A, Waris H, Rafique M, Suleman M, Mohammad A, Ali SS.** The Omicron (B.1.1.529) variant of SARS-CoV-2 binds to the hACE2 receptor more strongly and escapes the antibody response: Insights from structural and simulation data. International Journal of Biological Macromolecules 2022; 200, 438–448.

21-**Aksamentov I, Roemer C, Hodcroft EB, Neher RA.** Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software 2021:6(67), 3773.

22-**Alpert T, Vogels CBF, Breban MI, Petrone ME, Yale IMPACT Research Team, Wyllie AL, et al.** Sequencing SARS-CoV-2 Genomes from Saliva. Preprint. medRxiv 2021; doi: 2021.06.21.21259289.

23-**Ntagereka PB, Oyola SO, Baenyi SP, Rono GK, Birindwa AB, Shukuru DW, et al.** Whole-genome sequencing of SARS-CoV-2 reveals diverse mutations in circulating Alpha and Delta variants during the first, second, and third waves of COVID-19 in South Kivu, east of the Democratic Republic of the Congo. Int J Infect Dis 2022;122:136-143.

24-**Maljkovic Berry I, Hang J, Fung C, Yang Y, Chibucos M, Pollio A, et al.** Genomic surveillance of SARS-CoV-2 in US military compounds in Afghanistan reveals multiple introductions and outbreaks of Alpha and Delta variants. BMC Genomics 2022;23(1):513.

25-**Tamura K, Kumar S.** Evolutionary distance estimation under heterogeneous substitution pattern among lineages Molecular Biology and Evolution 2002; 19:1727-1736.

26-**Kumar S, Stecher G, Tamura K.** MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. Molecular Biology and Evolution 2016;33(7):1870-1874.

27-**Caputo E, Mandrich L.** Structural and phylogenetic analysis of SARS-CoV-2 spike glycoprotein from the most widespread variants. Life 2022;16;12(8):1245.

28-**Pan Y, Wang L, Feng Z, Xu H, Li F, Shen Y, et al.** Characterization of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis. The Lancet 2023;401(10377):664-72.

29-**Yechezkel M, Weiss A, Rejwan I, Shahmoon E, Ben-Gal S, Yamin D.** Human mobility and

poverty as key drivers of COVID-19 transmission and control. BMC Public Health 2021;21(1):596.

30-**Castelán-Sánchez HG, Martínez-Castilla LP, Sganzerla-Martínez G, Torres-Flores J, López-Leal G.** Genome Evolution and Early Introductions of the SARS-CoV-2 Omicron Variant in Mexico. Virus evolution 2022;8(2).

31-**Kumar R, Srivastava Y, Muthuramalingam P, Singh SK, Verma G, Tiwari S, et al.** Understanding Mutations in Human SARS-CoV-2 Spike Glycoprotein: A Systematic Review and Meta-Analysis. Viruses 2023;15(4), 856.

32-**Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, et al.** The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv: The Preprint Server for Biology 2020;06.12.148726.

33-**Shi PY, Plante J, Liu Y, Liu J, Xia H, Johnson B, et al.** Spike mutation D614G alters SARS-CoV-2 fitness and neutralization susceptibility. Preprint. Res Sq. 2020;rs.3.rs-70482.

Zulu E, Katete R S, Simusika P, Mutengo M M, Chabala FW, Kasonde M. Characterization of SARS-CoV-2 genetic diversity in Zambia: single center study. Microbes Infect Dis 2023; 4(4): 1088-1099.