GENETIC DIVERSITY OF SARS-COV-2 IN SOUTHERN PROVINCE OF ZAMBIA

By

Ben Katowa

A dissertation submitted in partial fulfilment of the requirement for the degree of Master of Science in Infectious Diseases at the University of Zambia

THE UNIVERSITY OF ZAMBIA

LUSAKA

2022

Declaration

I Ben Katowa hereby declare that this research work is my original work as a requirement for the Master of Science in Infectious Diseases at the University of Zambia, Lusaka, Zambia. It has not been submitted either wholly or partially for any degree, diploma or other qualification at this or any other university.

Sign: _____ Date: _____

Certificate of Approval

Thesis Title: Genetic Diversity of SARS-CoV-2 in Southern Province of Zambia

This thesis of **Ben Katowa** has been approved in in partial fulfilment of the requirement for the degree of Master of Science in Infectious Diseases at the University of Zambia.

……	……..	……
Internal Examiner	Signature	Date
……	……..	……..
Internal Examiner	Signature	Date
……	…..……..	……..
External Examiner	Signature	Date
……	…..……..	……..
Chairman of the Board	Signature	Date

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), has spread globally with catastrophic effects on the public health system. Four main Variants of Concern (VOCs), namely Alpha, Beta, Delta and Omicron, have been associated with waves of high transmission throughout the world. This study aimed at characterizing the SARS-CoV-2 variants that were circulating in Southern Province, Zambia. Nasopharyngeal swabs samples were collected between December 2020 and April 2022 from eight districts in Southern Province. A total of 198 samples were transported to Macha Research Trust (MRT) for Whole Genome Sequencing (WGS). Detection of the SARS-CoV-2 genome was done using the CDC protocol (Lu et al., 2020). The Oxford Nanopore and Nextseq illumina platforms were used for WGS. From the 198 samples collected, a total of 40 whole genomes were sequenced. Genetic analysis of 40 SARS-CoV-2 whole genomes revealed the circulation of Alpha (B.1.1.7), Beta (B.1.351), Delta (AY.116) and four Omicron lineages (BA.1, BA 1.1, BA 1.1.4 and BA2). From the seven lineages 42 percent, 15 percent, 13 percent, 10 percent, 10 percent, seven percent and three percent belonged to BA.1, B.1.351, BA.2, BA 1.1., AY.116, B.1.17 and BA 1.1.4 lineages, respectively. Across the 40 genomes analysed, a total of 292 mutations were observed from the original Wuhan/Hu 1/2019/EPI 1SL 402125 reference genome, of which 281 (96.2 percent) were found in the coding region of the genome, including 182 missense mutations, 66 synonymous mutations, 23 deletions, 9 insertions, 1 stop codon, and 11 mutations in the non-coding region. The Spike region had 82 mutations followed by the NSP3 region with 42 mutations. Phylogenetic analysis showed evidence of local transmission and possible multiple introductions of SARS-CoV-2 VOCs in Zambia from different European and African countries. This study stresses the need for continued genomic monitoring of SARS-CoV-2 circulation in Zambia to inform public health mitigation measures, particularly in strategically positioned regions such as the Southern Province which could be at increased risk of introduction of novel VOCs.

Dedication

This thesis is dedicated to my parents and family for their support throughout my studies. I would also like to dedicate this thesis to my supervisors for the unwavering support, guidance and extreme patience during my studies.

Acknowledgements

First and foremost I would like to thank the Africa Centre of Excellence for Infectious Disease of Humans and Animals (ACEIDHA) project (grant number P151847) funded by the World Bank, the Science and Technology Research Partnership for Sustainable Development (SATREPS) (grant number JP22jm0110019), the Japan Program for Infectious Diseases Research and Infrastructure (grant number JP21wm0125008, JP20wm0225003), Agency for Medical Research and Development (AMED) and the European and Developing Countries Clinical Trials Partnership (EDCTP2) programme under the PANDORA-IDNET Consortium (EDCTP Reg/Grant RIA2016E-1609) for supporting my research.

I would also like to thank my supervisor Dr Simulundu. He has provided me with unprecedented mentorship throughout the process of my studies. Together with Drs. Muleya and Mubemba, they supported my research endeavors, gave me opportunities to grow and develop skills in order to achieve my goals and gave me great advice not only on my school work, but also on life in general. I am proud and honored to have such a great team guiding me. I would also like to thank Dr. Changula and Dr. Chitanga for their continued guidance and mentorship throughout this process. I would also like to say a special thanks to Annie and Japhet who tirelessly helped and worked the long nights with me throughout this process. I would also want to thank Passwell, Mutinta, Limonty, Twig, and the Macha Research Trust for their support during this process

My family has been incredibly supportive and patient through this process. My parents always told me ‘an education will take you to places you had never dreamed of, only if you work hard enough’. I thank my family for pushing me from beginning. Lastly, I would like to thank my wife, Hapi Kamatha Katowa and our kids; Simone, Maurice and Ryan Kenji. You guys have been very supportive, encouraging and comforting to me when I’m struggling. You always keeping me busy and “trying to get your share of my time.” You really have been my pillar of hope and strength.

Table of contents

List of Tables

Appendix

List of figures

List of Abbreviations

ACE-2	Angiotensin Converting Enzyme -2
cDNA	Complementary Deoxyribo-Nucleic Acid
CHAZ	Churches Health Association of Zambia
CI	Confidence interval
COVID-19	Corona Virus Disease 2019
Ct	Cycle Threshold
DNA	Deoxyribonucleic acid
E	Envelope
M	Membrane
MERS-CoV	Middle East Respiratory Syndrome Corona Virus
MoH	Ministry of Health
MRT	Macha Research Trust
N	Nucleocaspid
NSP	Non-Structural Protein
ORF	Open Reading Frame
PCR	Polymerase chain reaction
POE	Point Of Entry
RBD	Receptor Binding Domain
RdRp	RNA dependant RNA Polymerase
RNA	Ribonucleic acid
S	Spike

SARS-CoV	Severe Acute Respiratory Syndrome Corona Virus
SARS-CoV-2	Severe Acute Respiratory Syndrome Corona Virus Type 2
SNP	Single Nucleotide Polymorphism
VOCs	Variant of Concern
WGS	Whole Genome Sequencing

CHAPTER ONE

INTRODUCTION

1.1 Background

The Corona Virus Disease 2019 (COVID-19) pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) has resulted in 647, 025, 583 confirmed cases and 6, 638, 728 deaths in the world as of November 30th, 2022. With a total of 12, 708, 137 confirmed cases and 258, 079 deaths in Africa as of November 30th, 2022 (World Health Organization, 2022a; Worldometer 2022 [accessed November 30, 2022]). The first COVID-19 case in Africa was reported in Egypt on 14 February 2020 (Kapata et al., 2020) followed by Algeria and Nigeria, with their first cases being reported on 25 February 2020 and 27 February 2020, respectively (World Health Organization, 2020a). By the end of March 2020, the majority of African countries had reported their first cases, and most of these cases were brought into Africa by travellers from Europe (Wilkinson et al., 2021; Loemb'e et al., 2020). The SARS-CoV-2 spread fast to most of Africa, prompting most African nations to implement harsh control measures including flight bans, school closures, and nationwide lockdowns (Morobe et al., 2022; Zambia National Public Health Institute, 2020a). Low levels of SARS-CoV-2 were observed as the control measures went into force, with little intra-country transmission and sporadic international viral transmission across neighbouring countries, perhaps due to the road and rail connections between these countries (Wilkinson et al., 2021; Morobe et al., 2022; Zambia National Public Health Institute, 2020a). According to reports, only 4 percent of the COVID-19 disease burden in the world is found in Africa; however, this is likely due to under reporting and the true disease burden may be considerably higher (Gill et al., 2022; Wilkinson et al., 2021).

Zambia also implemented control measures to slow the spread of the SARS-CoV-2, but as of November 30th, 2022, there had been 333, 746 confirmed cases and 4,019 deaths from COVID-19 (Zambia National Public Health Institute, 2022; Worldometer, 2022[accessed November 30, 2022]). Zambia has had four waves of high transmission: the first wave occurred from July to September 2020 and the second wave occurred from December 2020 to April 2021 and was dominated by the Beta variant of SARS-CoV-2, while the Delta variant dominated the third wave

from May to September 2021 (Reuters, 2022; Centers for Disease Control and Prevention, 2021). In Zambia, the fourth pandemic wave was dominated by the Omicron variant, with cases reaching their peak in early January 2022 before declining to low levels (World Health Organization, 2022a).

As of November 10th, 2022, Southern Province had 37 260 documented cases and 375 fatalities since the first case was detected in May 2020 (Zambia National Public Health Institute, 2022). Southern Province is a major tourist destination and shares three international boarders making it the principal trade route for most of the import and export from southern Africa.

1.2 Problem statement

The COVID-19 pandemic has continued to wreak havoc on global public health. With the continuously rising disease burden and death toll, genomic monitoring of SARS-CoV-2 has become an important tool in the fight against COVID-19. Currently in Zambia, there is limited information on the genomic epidemiology of SARS-CoV-2 in the country. Studies by Simulundu et al., 2020 and Mwenda et al., 2020 both highlight the detection of SARS-CoV-2 variants in the country, but are of limited scope. In Southern Province, there is no information on the SARS-CoV-2 variants circulating within the province. Therefore, there is need to identify and monitor the SARS-CoV-2 variants circulating in Southern province to provide timely and accurate information to guide decision makers in curbing the negative impact of COVID-19.

1.3 Study rationale

While SARS-CoV-2 variants of concern (VOCs) have emerged in the country (Mwenda, et al., 2020), there is a potential of some variants going undetected due to the limited genomic surveillance system in the country (Gill et al., 2022; Wilkinson et al., 2021). For Southern Province, there is currently no information on the genomic surveillance in the province. Effective implementation of control measures to decrease SARS-CoV-2 transmission depends on the prompt identification of these variants. Whole genome sequence analysis of the circulating variants provides critical information such as transmissibility, drug resistance, vaccine effectiveness and disease pathogenicity (World Health Organization, 2022b). Therefore, increased genomic surveillance is essential, especially at ports of entry (POE), to quickly identify SARS-CoV-2

variants and implement measures to curb their spread. Furthermore, in Southern Province this information is crucial to mitigating the spread of the SARS-CoV-2, as the area has high human movement due to trade and tourism and this has been exemplified by border towns recording high COVID-19 cases in the country (Zambia National Public Health Institute, 2020a).

Additionally, making these whole-genome sequences available in public databases will inform the wider research community in developing effective vaccines and diagnostics, among others (Yang et al., 2020; World Health Organization, 2020b).

1.4 Research Question

What are the genetic characteristics of the SARS-CoV-2 variants circulating in the Southern Province of Zambia?

1.5 Objectives

1.5.1 General objective

To determine the genetic diversity of SARS-CoV-2 circulating in Southern Province of Zambia.

1.5.2 Specific objectives

1. To determine whole genome sequences of SARS-CoV-2 detected in Southern Province of Zambia.
2. To determine the SARS-CoV-2 lineages and variants circulating in Southern Province.
3. To determine the evolutionary relatedness of SARS-CoV-2 lineages and variants circulating in Southern Province.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

2.1.1 Taxonomy

The SARS-CoV-2 is a large positive-sense RNA virus that is 29 903 bases long, and it belongs to the genus *Betacoronaviruses* (Lone & Ahmad, 2020: Su et al., 2016). It is the seventh coronavirus known to cause disease in humans. Other human-infective coronaviruses are Middle East Respiratory Syndrome Corona Virus (MERS-CoV), Severe Acute Respiratory Syndrome Corona Virus (SARS-CoV) and the Human Corona Viruses (229E, OC43, NL63 and HKU1) (Su et al., 2016). The SARS-CoV-2 is said to be highly related to corona viruses found in bats, with about 96.2 percent genome similarity with the RaTG13 virus found in horseshoe bats (*Rhinophus affinis*) and 93.3 percent genome similarity with RmYN02 virus found in *Rhinophus malayanus* another horseshoe bat (Holmes et al., 2021; Hu et al., 2021). In addition the SARS-CoV-2 genome has a 79 percent and 50 percent genome similarity with SARS-CoV and MERS-CoV respectively (WHO-China Study, 2021)

2.1.2 Coronavirus Genomic Structure

The Coronavirus genome encodes for four structural proteins. These are the Envelope protein (E) which is involved in viral shaping and assembly; the Membrane (M) protein; a protein that interacts with the other structural proteins; the Nucleocapsid (N) protein, an RNA binding protein and the Spike (S) protein, a structural protein that binds to the host receptors. The genome also encodes for 16 non-structural proteins (NSP 1 - 16) which are embeded within two Open Reading Frames (ORFs) (Singh & Yi, 2021; Amoutzias et al., 2022). The 16 NSPs encode for enzyme and functional proteins that are responsible for the transciption and replication of the virus (Amoutzias et al., 2022). NSPs 1–11 are considered to coordinate the expression of the entire genome. The NSP1 region encodes for a protein that reduces type 1 interferon production, resulting in an increase in viral replication. (Lin et al., 2021; Naqvi et al., 2020). NSPs 12–16 encode the principal enzymes involved in RNA synthesis and these include; the RNA dependent RNA polymerase enzyme(RdRp) which is responsible for viral replication (NSP 12); a helicase enzyme used to

separate the double stranded RNA into single strands during replication (NSP 13); an exonuclease (a proofreading enzyme) that is used to reduce SNP errors during replication (NSP 14); an endoribonuclease (NSP 15) and finally a methyl-transferase (NSP 16) (Amoutzias et al., 2022; Naqvi et al., 2020). A defective RdRp enzyme has been associated with causing more mutations within the viral genome which could leading to development of new variants (Amoutzias et al., 2022; Saberiyan et al., 2022).

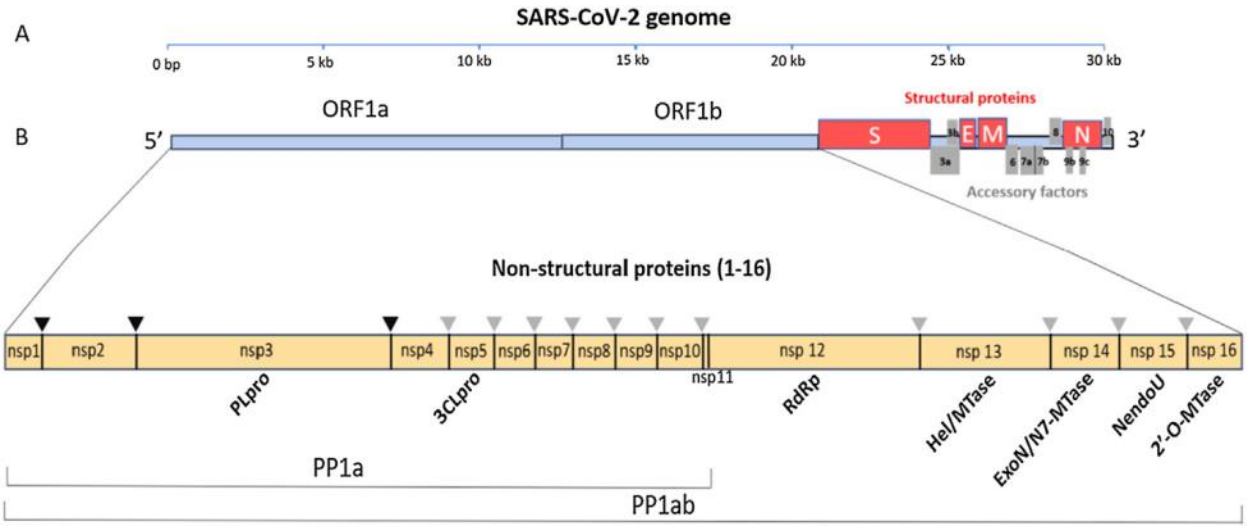

Figure 1: SARS-CoV-2 genome structure highlighting the genome length of the virus (A) and the structural and non-structural protein that are encoded by the SARS-CoV-2 genome (B).

2.2 SARS-CoV-2 Epidemiology

2.2.1 Origin and Spread

The SARS-CoV-2, known to cause the COVID-19 disease, was first identified in Wuhan, China. An outbreak of pneumonia-like illnesses was observed among individuals associated with a wet animal market (Rothan & Byrareddy, 2020)

The COVID-19 has continued to rampage the world with over 600 million cases and over 6 million deaths reported. The United States of America, India, France Brazil and Germany have recorded the highest number of COVID-19 cases (Worldometer, 2022). In Africa, South Africa, Morocco, Tunisia and Egypt have had the highest number of cases (WHO African Region, 2022; Worldometer, 2022).

The first case in Zambia was reported on March 18, 2020, resulting in the closure of educational institutions and enforcing of COVID-19 control measures. On the 2nd of April, 2020, Zambia recorded its first death and during the same month, the town of Kafue was put on lockdown (Zambia National Public Health Institute, 2020b). Subsequently Nakonde, became a COVID-19 hotspot, with 76 new cases recorded and the total number of cases reaching 1000, all within the month of May 2020 (Zambia National Public Health Institute, 2020b). The country then saw a significant increase in the number of cases, with over 12 000 cases being recorded by the end of August 2020 (Reuters, 2022). The B.1.351 variant was identified in Zambia in December 2020, and was followed by an increase in the number of cases, with over 57 000 cases recorded between January and February 2021 (Mwenda et al., 2020; Reuters, 2022). In April 2021, the first set of vaccines were received in Zambia while in June and July another upsurge of cases was recorded. This was the third wave in the country which was dominated by the Delta variant (Reuters, 2022). In December 2021, the Omicron variant was detected in Zambia. Vaccine campaigns were then accelerated, with 4.5 million people vaccinated in Zambia by June 2022 (Centers for Disease Control and Prevention, 2021).

2.2.2 Clinical symptoms, pathogenesis

Transmission of SARS-CoV-2 has been shown to occur through contact with an infected individual mainly through air droplets from the infected person (Rothan & Byrareddy, 2020) and there has been evidence of transmission through wastewater (Shaheen et al., 2022). SARS-CoV-2 infections begin with the virus attaching to the angiotesin converting enzyme-2 (ACE2) receptors in the respiratory tract (Srivastava et al., 2021) followed by viral replication once the virus attaches itself to host ribosomes (Naqvi et al., 2020). Following infection with SARS-CoV-2, symptoms present approximately 5.2 days after infection (Rothan & Byrareddy, 2020). The length of the incubathion period as well as symptom presentation vary according to individuals and is dependant on the age and immune status of the patient (Li et al., 2020). Symptoms of COVID-19 include but are not limited to coughing, fever, headache, sore throat, chest pains, shortness of breath and diahrroea (Chen et al., 2020; Huang et al., 2020). Severe symptoms include; acute respiratory distress syndrome (ARDS), hypoxia, sepsis and dysponoea (Chen et al., 2020; Huang et al., 2020). SARS-CoV-2 infections have also shown a change in blood cell levels, blood biochemistry and coagulation function (Chen et al., 2020; Huang et al., 2020; Li et al., 2021).

SARS-CoV-2 infections have been shown to be more severe in adults above the age of 60 as compared to children and adolescents (Li et al., 2021). SARS-CoV-2 seems to infect more men than females and is most severe in individuals with other cormorbidities (Chen et al., 2020; Huang et al., 2020).

2.2.3 Prevention and Control

The major COVID-19 prevention and control measures include; maintaining social distance, wearing of masks in public, washing of hands as often as possible, isolation of patients, improved ventilation and vacination (European Centre for Disease Prevention and Control, 2020). On the other hand, stringent control measures include; country or city lockdowns, curfews, banning of gatherings of any kind, point of entry screening, closing of airports and quarantine of travellers (Kapata et al., 2020; Zambia National Public Health Institute, 2020a; Loemb'e et al., 2020)

2.2.4 Disease Detection

Detection of SARS-CoV-2 is mainly done through real time quantitative polymerase chain reaction test, which will detect the SARS-CoV-2 RNA (Lu et al., 2020; Chen et al., 2020). Another detection method is the rapid detection antigen test which uses monoclonal antibodies targeting the nucleocapsid protein of the virus. The rapid test has recently become the point of care test due to its ease of use (World Health Organization, 2020c). Another test method is the serological method that tests for antibodies angainst the SARS-CoV-2 virus (World Health Organization, 2020c)

2.3 Variants of Concern (VOCs)

Variants of concern are strains that are known to have an increased transmissibility and an ability to cause more severe disease (World Health Organization, 2022b). Studies have shown that mutations in the genome may affect the virus differently. The main area of concern has been mutations that occur within the S region. The S region encodes for the proteins that help the SARS-CoV-2 virus attach to the host cells i.e. the Receptor Binding Domain (RBD) which attaches to the angiotensin-converting enzyme-2 (ACE2) receptors found in human cells (Haddad et al., 2021; Amoutzias et al., 2022). Mutations in the RBD have been associated with improved transmisibility of the virus, ability to evade the host's immunity and increase in pathogenicity (Srivastava et al.,

2021; Amoutzias et al., 2022). The S region has also been a target for vaccine development (Kober et al., 2020).

To date, five VOCs namely Alpha, Beta, Delta, Gamma and Omicron have been detected globally, while only four VOC (except gamma) have been detected in Africa (World Health Organization, 2022b). The first VOC, designated Alpha (B.1.1.7) was detected in September 2020 in the United Kingdom (UK). This variant is characterised by nine mutations in the S region, increased transmissibility and increased risk of hospitalisation of about 40 to 60 percent and increased risk of death of about 30 to70 percent when compared to the original Wuhan-1 strain (Saberiyan et al., 2022; Volz et al., 2021) The second VOC was the Beta (B.1.351 lineage) variant which was first detected in South Africa in October 2020 and became the most common variant in many African countries (World Health Organization, 2022b; Tegally et al., 2021). This VOC is characterized by mutations in the S protein, including in the RBD—K417N, E484K, and N501Y (Tegally et al., 2021; Saberiyan et al., 2022). In addition, the Beta variant is known to cause severe disease in young and healthy individuals and is also known to evade antibody neutralization (Tegally et al., 2021). This variant was replaced by the highly transmissible Delta (B.1.617.2 lineage) variant which was initially detected in India in December 2020 and spread worldwide among vaccinated as well as unvaccinated individuals (World Health Organization, 2022b; Singanayagam et al., 2021). This variant seeded the third wave of the pandemic in 2021 and was introduced in Africa in June 2021. The Omicron variant is characterised by over 30 mutations in the S protein, including a set of mutations previously observed in other VOCs and novel mutations. It was first reported in South Africa in November 2021 and became the driver of the fourth global wave of SARS-CoV-2 (Wilkinson et al., 2021; Telenti et al., 2022). The Gamma (P.1 lineage) variant was detected in December 2020 in Manaus Brazil. It is characterised by ten mutations in the S protein, which have been associated with increased transmissibility and immune evasion. Some of the mutations in the S protein of the gamma have variant have also been identified in the Beta and Delta variants i.e. E484K and N501Y (Banho et al., 2022).

2.4 SARS-CoV-2 Phylogenetics

With the rapid evolution of the SARS-CoV-2 and the constant emergence of VOCs, phylogenetics is needed to understand the evolutionary relatedness of these variants and assess the molecular

epidemiology of the pathogen (Singh et al., 2021). There are many different ways to carryout phylogenetics and for SARS-CoV-2 the most common method has been based the maximum likelihood method (Munjal et al., 2019). This method uses probability to consider the nucleotide variation in each sequence, and then tries to find the best fitting model that has the highest probablity under a given evolutionary model (Singh et al., 2021). SARS-CoV-2 phylogenetic trees are usually rooted to a single ancestor i.e. the Wuhan-1 strain, which is used to determine both the evolutionary rate of the virus and determine groups or clades of that virus. Clades are groups of viruses in a phylogenetic tree that cluster together and have similar characteristics. The evolutionary rate of the phylogenetic tree represents a change in the genomic sequence of SARS-CoV-2 genome and this has been demonstrated to be about 10^{-4} substitutions per site per year (Singh & Yi, 2021; Munjal et al., 2019; Singh et al., 2021).

Sallam et al. (2021) used phylogenetic analysis to assess the origin and spread of SARS-CoV-2 mutations in the Middle East and North Africa. Their study showed the temporal prevalence change of the D614G mutation in certain countries in the Middle East and North Africa. While Giandhari et al. (2020*)* used phylogenetic analysis to determine the early transmission dynamics of SARS-CoV-2 in South Africa and was able to determine both the origins and introductions of SARS-CoV-2 into Kwa-Zulu Natal, South Africa.

Currently, there is limited information regarding the detection of SARS-CoV-2 variants and their epidemiology in the country. Two studies have shown the detection of the B.1.1 variant (Simulundu et al., 2021) and the B.1.351 variant (Mwenda et al., 2021). Whole-genome sequencing (WGS) and phylogenetic analyses were used in this study to describe the genetic characteristics of SARS-CoV-2 in Zambia's Southern Province.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Study Area

The study was conducted in the Southern Province of Zambia. The province shares borders with three neighboring countries, namely Zimbabwe, Botswana and Namibia (Figure2). The region is a major tourist destination with Livingstone District being the tourist capital of the country and it is a major trade entry point for most goods from Southern Africa. These socioeconomic activities place Southern Province at an increased risk of introduction and spread of novel variants.

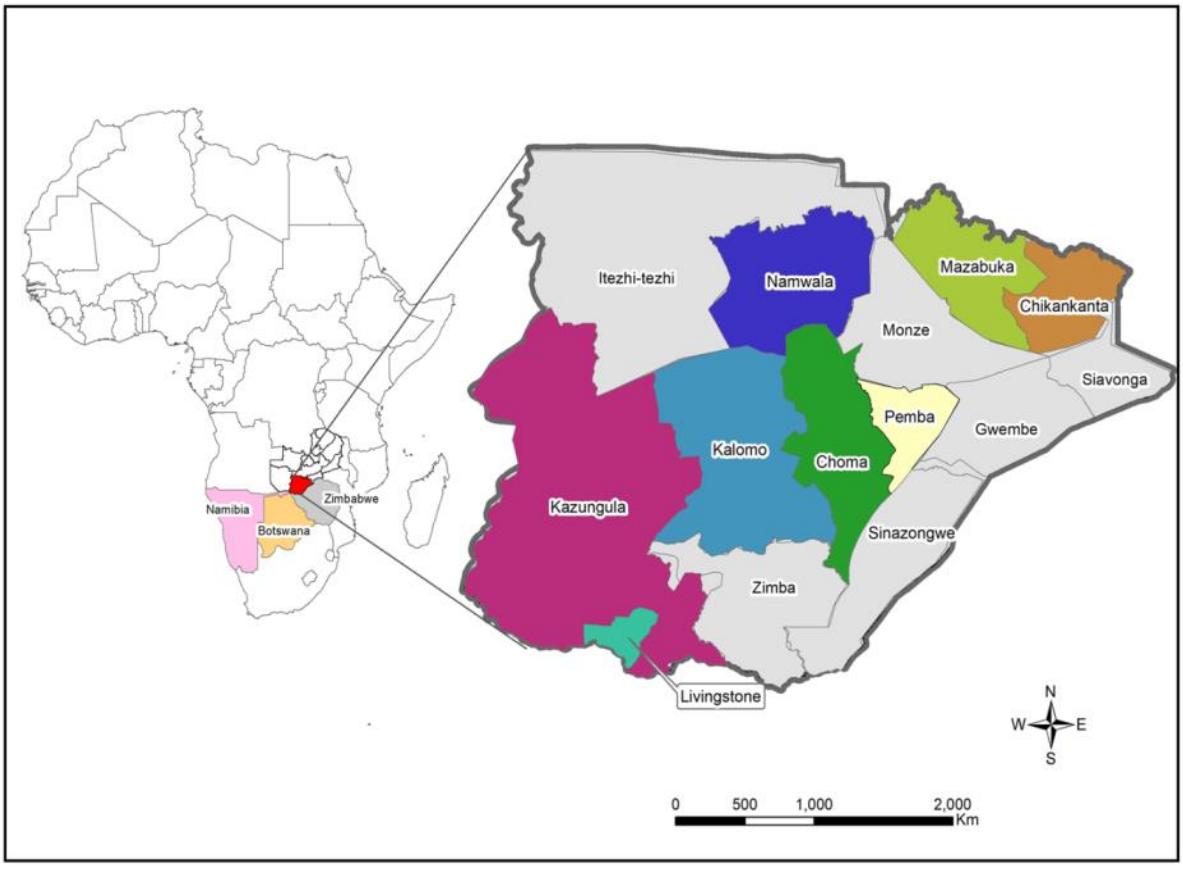

Figure 2: Map showing the location of study sites in Southern Province. The map of Africa depicts Zambia with its neighbouring countries that share the border with Southern Province. The insert map shows the Southern Province of Zambia. The maps were generated using Quantum Geographic Information System (QGIS) version 3.10.

3.2 Study Design

This study was a cross-sectional study that employed a simple random sampling method. Positive samples from the routine diagnosis of COVID-19 were randomly selected for whole genome sequencing (WGS). WGS was carried out at two sites, Churches Health Association of Zambia (CHAZ) (Illumina NextSeq sequencing) and Macha Research Trust (MRT) (MinION sequencing).

3.3 Sample size

The sample size refers to the number of samples that were included in the study and these samples were representative of the study population.

Formula for sample size calculation:

$n = Z^2 * P(1-P)/d^2$

Z is Z-score; the number of standard deviations a given fraction is away from the mean. The desired level of confidence in this case is 95 percent which has the Z-score equal to 1.96.

P is expected prevalence proportion; the expected prevalence proportion is the percent of the population that has the specific characteristic of interest

d is the margin of error; a fraction that told us how the expected result reflected the entire population. The smaller the margin of error, the closer one is to having the precise answer at a given confidence level.

n is the sample size; the number of individuals that will be drawn from the population (N) to give a representative conclusion of a specific characteristic of interest.

For this study, 95 percent disease prevalence was used, with a marginal error of 0.05 and a Z-score of 1.96. The sample size required was 73 participants (Niang et al., 2006).

$n = Z^2 * P(1-P)/d^2$

$n = (1.96)^2 * 095(1-0.95)/(0.05)^2$

$n = 3.8416 * 0.95\ (0.05)/0.0025$

n= 3.8416 * 0.0475/0.0025

n = 0.182476/0.0025

n= 72.9904

n= 73 participants

3.4 Sampling technique

Sample collection was done through the Ministry of Health (MoH) COVID-19 surveillance system under the guidance of the Zambia National Public Health institute. Nasopharyngeal swabs were collected between December 2020 and April 2022 from eight districts in the Southern Province (Figure 2) through routine surveillance and screening for Influenza-like illnesses and targeted surveillance of cluster outbreaks. The samples were then brought to the Macha Research Trust (MRT) laboratory, which is one of the Zambian MoH approved COVID-19 testing laboratories for samples collected in Southern Province. The samples were transported to MRT using the COVID-19 routine courier service provided by MoH. Demographic data for each sample was obtained from the information sheets sent with each sample or emailed electronically to the testing laboratory. The primary information on each sheet included the name, age, sex, date of collection, place of collection and details of the care giver.

3.5 Inclusion and exclusion criteria

3.5.1 Inclusion criteria

All positive samples that were collected through routine surveillance and targeted surveillance of cluster outbreaks and had a Ct value of less than or equal to 30 with complete metadata or partial metadata were included in the study. Only samples that were collected from within Southern Province were included. Samples that had only one critical metadata missing i.e. date of collection, place of collection, age or sex, were also included.

3.5.2 Exclusion Criteria

Samples that were not collected from routine and targeted surveillance were excluded. All negative samples and positive samples with a Ct value greater than 30, and with incomplete metadata or missing two or more critical metadata i.e. date of collection, place of collection, age and sex, were

excluded from the study. Samples that had low genome coverage were excluded from the study. Samples that were not collected from within Southern Province were also excluded

3.6 RNA extraction

Viral RNA was extracted from nasopharyngeal swabs using two methods; the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) or the MagMax kit (Thermo Fisher Scientific, Waltham, MA, USA) on an automated KingFisher Flex platform (Thermo Fisher Scientific, USA) according to manufacturer specifications and protocols. Briefly, using the QIAamp, 560 microliters of AVL Buffer with carrier RNA were added to 1.5ml tubes. 14 microliters of sample were added to each respective tube and incubated for ten minutes. Another 560 microliters of 99 percent molecular grade ethanol was added to each sample, and this solution was then added to spin columns and centrifuged for one minute. After two washes, 60 microliters of RNA were eluted from each sample and stored at -80 degrees Celsius.

With regards to the kingfisher system, a total of 100 microliters of sample was added to the KingFisher 96 deep well plate together with 20 microliters of Proteinase K and 100 microliters of lysis buffer. After a 15 minutes incubation, 200 microliters of binding buffer and 20 microliters of magnetic beads were added to each sample. The samples were then washed three times and 100 microliters of RNA was eluted from each sample and stored at -80 degrees Celsius.

3.7 Real time Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR)

Real time quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR) was done using a modified Centre for Disease Control and Prevention (CDC) SARS-CoV-2 qRT-PCR assay (Lu et al., 2020). Briefly, two separate 20 microliters reactions were set with ten microliters Quanti-Nova Reverse Transcriptase Master mix, 0.2 microliters of Quanti-Nova Probe Reverse Transcriptase mix, 3.8 microliters of water and one microliter of primer probe set (IDT, Coralville, IA), with five microliters of RNA extract. Amplification was performed on a StepOne plus Applied BioSystems instrument (Waltham, Massachusetts, USA). Cycling conditions used were 10 minutes at 45˚C, 5 minutes at 95˚C then 45 cycles of 5 seconds at 95˚C and 30 seconds at 60˚C. A positive test was considered when sample amplification crossed a pre-set threshold with a Cycle

threshold (Ct) value of less than 37, but only samples that had a Ct value of 30 or less were considered for WGS.

3.8 Whole genome sequencing

3.8.1 Library Preparation and sequencing on the Oxford Nanopore MinION platform

cDNA synthesis reaction was performed on the samples using 5X Lunascript Reverse Transcriptase kit (Invitrogen, USA), following the manufacturer's instructions. Library preparation was done using the ARTIC protocol version 3 (Quick, 2020). Briefly, two microliters of 5X Lunascript master mix (New England Biolabs, Massachusetts USA) and eight microliters of RNA template were incubated at 25°C for 2 minutes then 55°C for 10 minutes and 95°C for 10 minutes for cDNA synthesis. Whole genome amplification of SARS-CoV-2 was done using a multiplex PCR with custom designed primers (Simulundu et al., 2020). The PCR products were cleaned using AMPure XP beads (Beckman Coulter) and DNA quantification was done using a Qubit fluorometer (Thermo Fisher Scientific). End-repair on the amplified samples was done using NEBNext Ultra II End Repair Module (New England BioLabs). Native barcode expansion kits 1-12 and 13-24 were used in combination with a Ligation Sequencing Kit (SQK-LSK109) (Oxford Nanopore Technologies). Subsequently genomic sequencing was done using the minION 1MkB. The RAMPART (v1.0.6) software package was used to monitor sequencing performance in real-time, with runs proceeding until a minimum of approximately 200-fold coverage was achieved across all amplicons. At this point, the run was terminated and the resulting reads were base called using Guppy (4.0.14). Consensus sequence generation was done using the ARTIC bioinformatics pipeline (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html).

3.8.2 Illumina NextSeq Sequencing

Libraries for samples were prepared using the Illumina COVIDSeq kit on the automated Hamilton robotic instrument. Briefly, cDNA synthesis was done using random hexamer through RT-PCR. Targeted genome amplification to amplify fragments of the whole genome was done using the COVIDSeq primers (which contains two sets of primers i.e. primer 1 and primer 2). Two separate plates were prepared with ten microliters of master mix and five microliters of sample, with each plate containing a different primer set. Samples were then pooled and cleaned using AMPure XP beads (Beckman Coulter). After successful library clean-up and pooling, pooled samples were

quantified and normalized using a Qubit dsDNA HS Assay kit by diluting the sample library to a final loading concentration of 1nM. Thereafter, 25µl was loaded on the NextSeq 2000 (Illumina) SARS-CoV-2 genomic sequencing. A customized version of the DRAGEN DNA pipeline was used to perform Kmer-based detection of SARS-CoV-2. The Nextseq 2000 (Illumina) then aligned the reads to a reference genome, called variants and generated consensus genome sequences. The NextSeq 2000 (Illumina) optionally performs lineage/clade analysis using Pangolin and NextClade.

3.9 Data Analysis

3.9.1 Phylogenetic Analysis

Whole-genome sequences were annotated using the reference genome of hCoV-19/Wuhan/Hu-1/2019|EPI_ISL_402125 (GISAID, 2020). Audacity *Instant* (GISAID, 2022) was used to retrieve SARS-CoV-2 whole-genome sequences from the GISAID database that were most similar to the sequences generated in this study. A total of 140 sequences were retrieved from the GISAID database and were used for phylogenetic analysis. Reference sequences of VOCs detected in Southern Africa and other parts of the world were included as well. Sequences isolated before or within the same period and belonging to the same lineage as those characterized in this study were also used. Reference sequences from the GISAID database with more than 10 percent of the nucleotide sequence missing were excluded from the analysis. A total of 180 sequences were used for analysis. The fastA files for each sequence were combined in a single multi FastA file using Geneious Prime v2022.0.1 (https://www.geneious.com). Multiple sequence alignment was performed using the FFT-NS-2 algorithm available in the multiple sequence alignment programme (MAFFT), using the default settings (https://mafft.cbrc.jp/alignment/server/index.html (Katoh et al., 2019). Alignment inspection, gaps trimming and conversion of the alignment into a phylip file were done using Geneious Prime v2022.0.1 (https://www.geneious.com). Following alignment and conversion, a maximum likelihood (ML) phylogenetic tree was constructed using the PhyML Online server (www.atgc-montpellier.fr/phyml/) (Guindon et al., 2005) utilizing the smart model selection (SMS) and the Bayesian Information Criterion (LeFort et al., 2017). Branch support was estimated through the Shimodaira Hasegawa (SH) approximate likelihood ratio test (SH-aLRT) (LeFort et al., 2017) and the ML tree was then rooted using TempEst v1.5.3 (Rambaut et al., 2016),

which also estimated the best-fitting root of this phylogeny using the heuristic residual mean squared function aimed at minimizing the variance of root-to-tip distances. The resultant ML tree file was edited using Interactive Tree of Life (iTOL) v5, an online tool for phylogenetic tree display and annotation (Letunic & Bork, 2021)

3.9.2 Classification of SARS-CoV-2 Lineages and Identification of SNP

PANGO lineage classification was done using Pangolin v3.1.16 (https://pangolin.cog-uk.io/). Identification of single nucleotide polymorphisms (SNPs) was performed using the coronapp (http://giorgilab.unibo.it/coronannotator/). SNPs were identified by comparing the sample nucleotides at each position with the reference genome at each position of interest. These variations were then exported to a vcf file and visualized in Microsoft Excel

3.10 Ethical Consideration

Ethical approval to conduct this study was obtained from ERES Converge with Ref. No. 2020-Aug-008. All samples were de-identified to maintain participant confidentiality. No other identifiable information was collected for the study.

CHAPTER FOUR

RESULTS

4.1 Overview

A total of 198 samples were collected between December 2020 and April 2022 for whole genome sequencing, from which 74 were negative for SARS-CoV-2, 51 had Ct values greater than 30 and 33 had low genome coverage. From the 40 samples that were successfully sequenced with a genome length of 29,756bp to 29,889bp long, seven SARS-CoV-2 lineages were identified which belonged to four VOCs. These sequences were obtained from eight districts in Southern Province. A total of 292 SNPS were identified from the 40 samples of which 281 were found in the coding region of the genome. Phylogenetic analysis showed the sequences clustering into their respective VOC clades with most the sequences from this study clustering with samples from Europe and other African countries.

4.2 Demographics

A total of 198 samples were collected between December 2020 and April 2022 for whole genome sequencing, from which 74 were negative for SARS-CoV-2, 51 had Ct values greater than 30 and 33 had low genome coverage. A total of 40 samples were successfully sequenced of which 13 samples were sequenced at MRT and 27 were sequenced at the CHAZ laboratory.

Demographic data were analysed for all the 198 samples and the majority of the samples (104/198; 52.5 percent) were from females (Table 1). The mean age of the participants was 28 (range: 0–82). The age group 15 – 50 years had the highest number patients. The data set for gender and age were not available for one and five samples, respectively (Table 1).

Table 1: Characteristics of 198 SARS-CoV-2 patients in Southern Province

Parameters	Sample Distribution n (%), n =198
Age Group	
0–14 Years	7 (3.5)

15–50 Years	171 (86.4)
>50 Years	15 (7.6)
Unknown	5 (2.5)
Gender	
Female	104 (52.5)
Male	93 (47.0)
Unknown	1 (0.5)

From the 40 samples that were successfully sequenced 26 (65 percent) were males and 14 (35 percent) were females. Five participants were below the age of 15 and 31 were between the ages of 15-50. Two were above the age of 50 and three had no information on age (Table 2).

4.3 SARS-CoV-2 Lineage Distribution in Southern Province

The 40 genomes detected in this study were distributed into seven lineages, namely B.1.1.7 (Alpha), B.1.351 (Beta), AY.116 (Delta), and Omicron (BA.1, BA.1.1, BA.1.14, and BA.2) (Table 2: Figure 3A). The largest number of the sequences (n = 17, 42.5 percent) belonged to lineage BA.1/GRA (Table 2: Figure 3A). All lineage AY.116 sequences originated from Choma District, whereas the six B.1.351 lineage was from Choma (n = 2), Namwala (n = 2), Kalomo (n = 1) and Mazabuka (n = 1) districts (Table 2: Figure 3B). The Alpha variant (B.1.1.7) viruses were found in Namwala, Pemba, and Chikankata districts (Table 2: Figure 3B). Of the Omicron variants, 11/27 (40.7 percent) were from Livingstone, 9/27 (33.3 percent) from Chikankata, 4/27 (14.8 percent) from Kazungula, 2/27 (7.4 percent) from Choma and 1/27 (3.7 percent) from Namwala (Table 2: Figure 3B). Most of lineage BA.1 viruses were detected in Chikankata and Livingstone districts where 7/27 (25.9 percent) viruses of this lineage were found in each district (Table 2: Figure 3B). Lineage BA.1.1 was detected in Chikankata and Kazungula districts whereas B.1.14 was only detected in Livingstone (Table 2: Figure 3B). Three of the five BA.2 lineage viruses were detected in Livingstone whereas the other two were detected in Choma and Kazungula districts (Table 2: Figure 3B)

Table 2: Characteristics of 40 SARS-CoV-2 whole genomes from Southern Province, Zambia

Parameter	Sample Distribution n(%)

Sex	
Male	26 (65)
Female	14 (35)
Age	
0-15	5 (12.5)
15-50	31 (77.5)
>50	2 (5)
Unknown	2(5)
Year of Collection	
2020	1 (2.5)
2021	26 (65)
2022	7 (17.5)
Unknown	6 (15)
District	
Chikankata	10 (25)
Choma	8 (20)
Kalomo	1 (2.5)
Kazungula	4 (10)
Livingstone	11 (27.5)
Mazabuka	1 (2.5)
Namwala	4 (10)
Pemba	1 (2.5)

Lineages	
B.1.1.7	3 (7.5)
B.1.351	6 (15)
AY. 116	4 (10)
BA. 1	17 (42.5)
BA. 1.1	4 (10)
BA. 1.1.4	1(2.5)
BA. 2	5 (12.5)
Waves	
First	0 (0)
Second	8 (20)
Third	5 (12.5)
Fourth	27 (67.5)

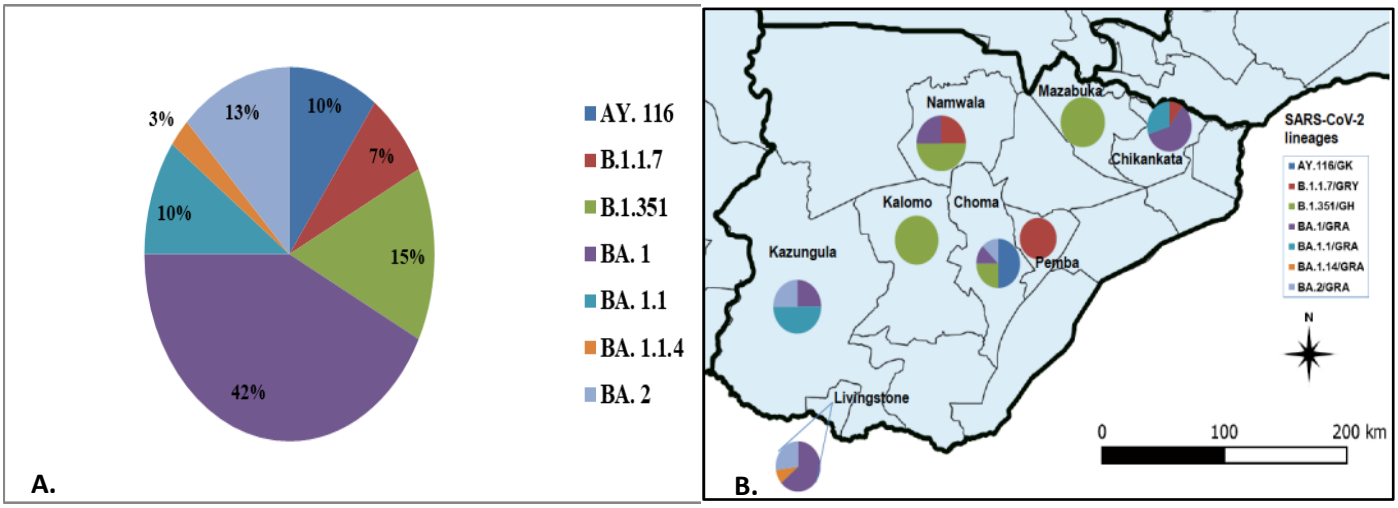

Figure 3: (A) Percentages of different lineages of SAR-CoV-2 detected in Southern Province. (B) Distribution of the different SAR-CoV-2 lineages detected in different districts of Southern Province.

4.3 Phylogenetic Analysis

Phylogenetic analysis showed the clustering of sequences according to their respective clades (Figure 4). In the Delta clade, four Southern Province sequences (Zambia/SP250/2021|EPI ISL 6761088, Zambia/SP253/2021|EPI ISL 6762977, Zambia/SP251/2021|EPI ISL 6761106, and Zambia/SP252/2021|EPI ISL 6761100) could be found and separated into two groups of which two formed a distinct cluster with a Zambian isolate whereas the other two clustered with sequences from Angola, Eswatini, and Zambia (Figure 4). With regards to the beta clade, six sequences generated in this study could be found. Two of the sequences (Zambia/SP30/2021|EPI ISL 6760973 and Zambia/SP87/2021|EPI ISL 6764745) clustered with isolates from Zambia, Zimbabwe, England, and the Democratic Republic of Congo (DRC) while the other two formed a distinct cluster with sequences from Zambia (Figure 4). The last two sequences (Zambia/SP11/2021|EPI_ISL_6760905 and Zambia/SP10/2021|EPI_ISL_6760707) belonged to separate clusters with the first sequence closely clustering with Zambian sequences and the other with sequences from Malawi, Eswatini, and Botswana (Figure 4). In the Alpha clade, three Southern Province sequences namely, Zambia/SP32/2021|EPI_ISL_6761015, Zambia/SP37/2020|EPI_ISL_6761027, and Zambia/SP172/2021|EPI_ISL_6761052 formed a distinguishable cluster with sequences from England and Zambia (Figure 4). The Omicron clade

separated into two clusters (Figure 4) with the majority (22/27; 81.5 percent) and minority (5/27; 18.5 percent) of the Zambian sequences belonging to the BA.1 sub-lineage and BA.2 lineage, respectively. Phylogenetic analysis further showed that the Omicron sequences from this study were mainly closely related to sequences from European and other African countries (Figure 4).

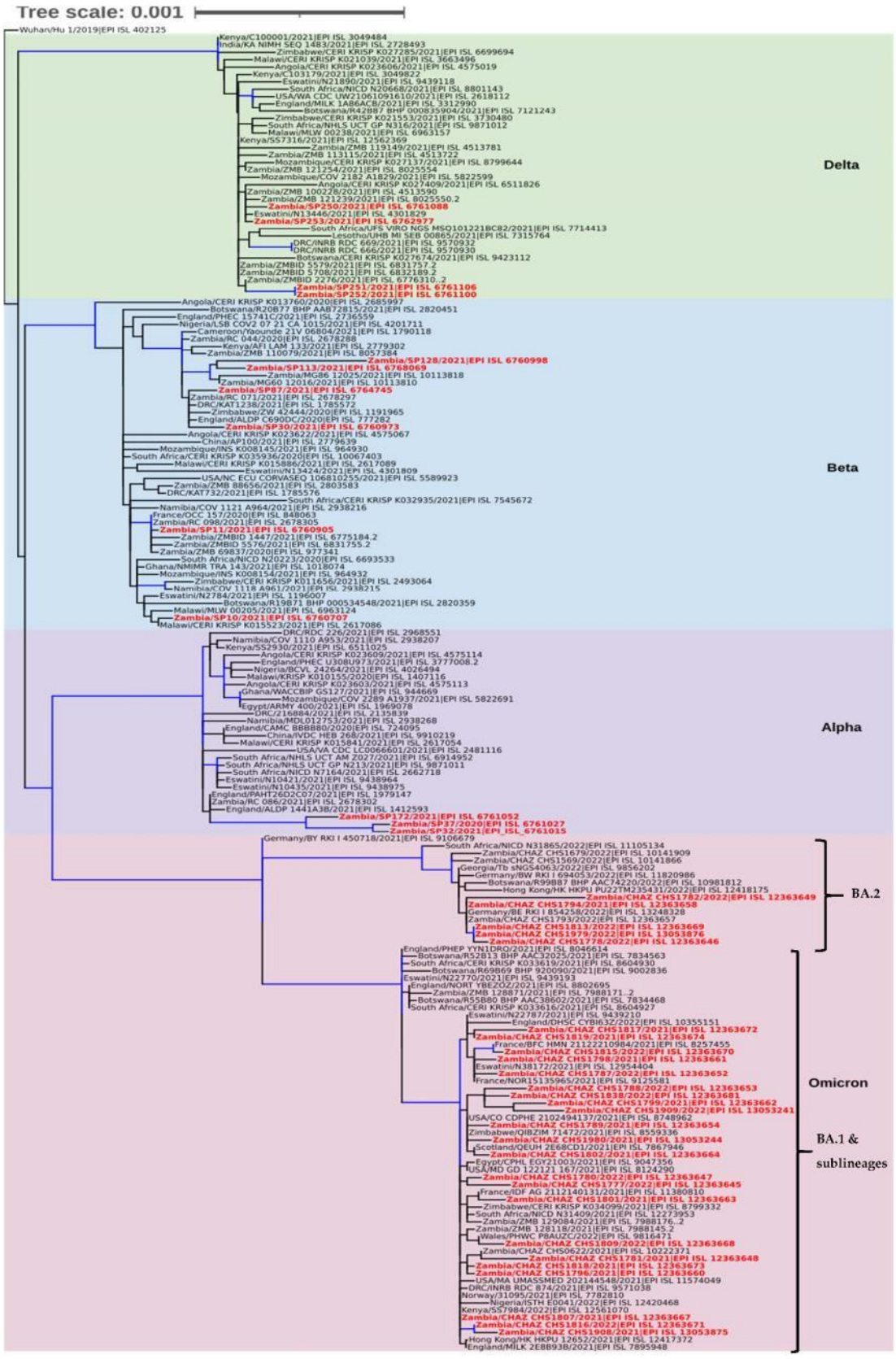
Tree scale: 0.001
Wuhan/Hu 1/2019|EPI ISL 402125
Delta
Beta
Alpha
BA.2
Omicron
BA.1 &
sublineages

Figure 4. Phylogenetic analysis of SARS-CoV-2 genomes from Zambia and other countries. The genomes generated in this study are presented in red whereas the shaded areas indicate the clades of variants of concern. Each sequence was named with the country name first followed by the isolate name and then the GISAID accession number. The tree branches highlighted in blue indicate tree branches that had a strong maximum likelihood ratio greater than 0.9, whereas the tree scale represents the nucleotide substitutions per site.

4.4 Single Nucleotide Polymophism and Amino Acid mutation

A total of 292 SNP mutations were identified from the 40 genomes studied when compared to the Wuhan/Hu 1/2019|EPI ISL 402125 reference genome (GISAID, 2020) (Table 3). A total of 281 (96.2 percent) mutations were detected in the coding regions of the genomes, of which 64.8% (182/281) were missense (mutations that cause a change in the amino acid at that position) mutations, 23.5% (66/281) were synonymous (mutation that codes for the same amino acid) mutations, 8.2% (23/281) were deletions, 3.2% (9/281) insertions and one was a stop codon (0.4%) gained on the ORF8 region (Table 3)

The S protein was the most mutated gene with 82 mutations (Table 3) whereas the second mutated gene was the NSP3 protein with 42 mutations (Appendix 1). From the 82 mutations in the S protein, 65/82 were missense mutations, 3/82 synonymous mutations, 10/82 deletions, and 4 insertions (Table 3). A large deletion of 26 nucleotides was observed on position 29734 of the 3′UTR of the four sequences from the Omicron clade (Appendix 1).

Table 3: Distribution of Mutation along different genomic regions of the SARS-CoV-2 sequences detected in Southern Province

Genome Segment	Missense Mutation	Synonymous Mutation	Deletion	Insertion	Others	Total Mutation
Coding Region						
ORF1ab	74	48	9	3	0	134
Spike	65	3	10	4	0	82
ORF3a	5	4	0	0	0	9

Envelope	5	0	0	0	0	5
Membrane	5	2	0	0	0	7
ORF6	2	2	0	0	0	4
ORF7a	2	0	0	0	0	2
ORF7b	4	1	2	2	0	9
ORF8	4	3	1	0	1	9
Nucleocapsid	16	3	1	0	0	20
Non-coding Region						
5′UTR	0	0	0	0	4	4
3′UTR	0	0	0	0	7	7
Total	182	66	23	9	12	292

The most common amino acid (AA) mutation was the D614G substitution on the S protein and P314L substitution on the NSP12b (RNA-dependant RNA-polymerase (RdRp)) protein which occurred in all the sequences studied (Appendix 1). The second most common AA mutation (39/40; 97.5%) was the F106F substitution on the NSP3, followed by the K417N (31/40; 77.5%) substitution on the S protein and the T492I substitution on the NSP4. Which was followed by P681H (30/40; 75%) and N501Y (29/40; 72.5%) substitutions on the S protein. A total of 67.5 percent (27/40) showed other amino acid substitution in the S protein including T95I, G339D, S373P, S375F, H655Y, N679K, N764K, D796K, Q954H, and D1146D (Appendix 1).

The Omicron variant had the highest number of mutations when compared to the Wuhan-Hu-1 SARS-CoV-2 reference genome out of all the VOCs (Appendices 2-5). The Omicron variant had 60 AA mutations which included six deletions and four insertions (Appendix 5). From the 60 mutations identified, 22 were found in the RBD of the S protein i.e. G339D, R346K, Y369Y, S371L, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446S, T470A, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H while the N501Y was found in the Alpha variant. S325P, I326K, V327A, K417N, E484K and N501Y were found in the Beta variant and L452R and T478K in the Delta variant (Table 4).

Table 4: Spike protein mutations in different SARS-CoV-2 variants compared to the wild-type (Wuhan-Hu-1 strain)

SARS-CoV-2 Variants	**Spike Mutations**
Wuhan-Hu-1 (wild-type)	-
Alpha (B.1.1.7)	delH69, delY145, **N501Y**, A570D, D614G, P681H, T716I, T874I, S982A, D1118H
Beta (B.1.351)	L18F, D80A, D215G, ΔL242, T307P, N317F, **S325P, I326K, V327A, K417N, E484K, N501Y**, D614G, A701V, A1087S
Delta (AY.116)	T19R, T95I, G142D, delE156, **L452R, T478K,** D614G, P681R, D950N
Omicron (BA.1, BA.1.1, BA.1.14, BA.2)	T19I, delL24, delA67, delA67, delI68, T95I, delG142, G142D, V193L, Y200C, insI210, delN211, N211K, L212C, V213G, insS214, insV213, insR214, insV213, insR214, R214R, A243S, L244S, **G339D, R346K, Y369Y, S371L, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446S, T470A, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H,** T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F, V1104L, D1127G, D1146D, V1264L

CHAPTER FIVE

5.0 DISCUSSION

The 40 SARS-CoV-2 genomes generated from this study belonged to four SARS-CoV-2 VOCs (i.e. Alpha, Beta, Delta, and Omicron variants). These VOCs are a public health challenge because of their increased viral transmissibility and disease severity (Amoutzias et al., 2022; Gomez et al., 2021). The VOCs were detected in eight different districts of the Southern Province. The majority of the Omicron variants were detected in Chikankata and Livingstone districts, with Livingstone having more sub-variants detected. It is possible that Livingstone, being a border town, a tourist capital and a major transportation link to Zambia's neighbouring countries, could be at an increased risk of the introduction of novel VOCs. Chikankanta was undergoing the construction of a new hydro-electric power station (Power Technology, 2020) and this may have led to increased visitation from international contractors as well as local gatherings and thus increasing the transmission and importation of new variants during the construction process. This is supported by the findings of Gan & Koh (2021) who observed high levels of transmission among construction workers in Singapore, before the application of interventions. Three of the four variants of concern were detected in both Choma and Namwala with the exception of the Alpha and Delta variants respectively. This could be explained by both towns being in close proximity to Macha Research Trust, making it easy for them to deliver samples. Choma is also the provincial capital of Southern Province and this may lead to a higher influx of people travelling to Choma to conduct governmental and administrative duties as well as conduct business.

Omicron (BA.1) and Beta (B.1.351) lineages were the two most common among the sequences in this study. The Beta lineage predominated in the second wave while the Delta and Omicron dominated the third and fourth waves. The findings of this study support those of other authors who reported the predominance of Beta (B.1.351), Delta (B.1.617.2), and Omicron (BA.1) variants in the second, third, and fourth waves of the pandemic in Africa, respectively (Wilkinson et al., 2021; Tegally et al., 2021; Viana et al., 2022), in addition, the viral transmissibility improved with each variant as the mutations in RBD increased in each variant (Table 4). Furthermore, Zambia had a sharp rise in confirmed cases and fatalities following the detection of the Beta and the Delta variants (Reuters, 2022; Mwenda et al., 2020).

The S protein plays a crucial role in viral infection and pathogenesis because of its role in host cell receptor recognition, viral attachment, and entry (Morobe et al., 2022; Isabel et al., 2020). The D614G mutation which has been associated with a high viral load, infectivity, and transmissibility (Zhang et al., 2020) was also observed among all the samples that were studied. Other authors have also reported an increase in the abundance of the D614 mutation worldwide (Isabel et al., 2020; Sallam et al., 2021).All the sequences from this study also contained the P314L mutation in the NSP12/RdRp region. This mutation is associated with the production of a dysfunctional enzyme that generates errors during RNA synthesis, increasing the chances of mutations occurring which may lead to the emergence of new variants. A high co-occurrence of these mutations has been reported around the globe (Flores-Alanis et al., 2021; Obeid et al., 2021). It is also suggested that the co-occurrence of the D614G and NSP12_P314L mutations may enhance viral entry and replication, respectively (Haddad et al., 2021).

The N501Y mutation, which is present in the RBD of the S protein, is shared by the Alpha, Beta, and Omicron variants. It is known to cause an increased binding affinity of the RBD for the ACE2 receptor, raising the viral transmission rate (Gomez et al., 2021). This mutation was detected in all Alpha, Beta, and 20 Omicron sequences in this study. Furthermore, the K417N and E484K mutations in the S protein, common to all Beta variants (Tegally et al., 2021) were also detected in this study's sequences. These mutations have been associated with increased binding affinity to the ACE receptors (Tegally et al., 2021). Other mutations in this study included the Q27 stop in the ORF8 in all three Alpha variants. This mutation has been known to truncate or inactivate the ORF8 protein, facilitating build-up of additional mutations in other regions (Gomez et al., 2021; Saberiyan et al., 2022). Further, eight mutations namely D614G, D950N, F157Δ, L452R, P681R, R158Δ, T19R, and T478K were detected in the S protein in the four sequences of the Delta variant identified in this study. These mutations are identical to those detected in the Indian Delta variants (B.1.617.2) (Kang et al., 2021).

The close phylogenetic relatedness of sequences generated in this study with those from European and other African countries supports the idea of possible multiple introductions of the virus from different regions (Wilkinson et al., 2021). Additionally, phylogenetic analysis showed that some of the study's sequences clustered with other Zambian sequences, which may indicate that these

viruses circulated locally. Interestingly, sequences obtained in this study within the Alpha variant clade were phylogenetically distinguishable and were detected in three different districts. This may also suggest independent introductions, particularly from Europe, as these sequences were closely related to isolates from England. The study sequences in the Alpha clade also displayed a longer branch length as compared to the other sequences in this clade indicating continued evolution as the virus circulated. Surprisingly, the Alpha variant has not been associated with any COVID-19 wave in Zambia. This observation may suggest that the Alpha variant has no selective advantage over the other VOCs such as the Beta and Delta variants (Sun et al., 2022; Hirabara et al., 2022) Although some of the Beta and Delta variants were closely related to isolates from Europe and Zambia, others showed a close relationship to isolates from Eswatini, DRC, Malawi, and Zimbabwe, suggesting an intra-continental transmission which could be associated to poor implementation of COVID-19 control measures thus permitting the variants to spread within the region. Phylogenetic analysis also revealed that Omicron variants separated into two major clusters, BA.1 and BA.2, signifying the continued evolution of this VOC.

Sequences of the Omicron variant obtained in this study were highly mutated, having 149 mutations across the 27 sequences. The findings are consistent with the findings of Saxena et al. (2022), who detected more mutations in Omicron variants than the Delta variant. When the S protein mutations of the VOCs in this study were compared to the hCoV-19/Wuhan/Hu-1/2019|EPI_ISL_402125 (GISAID, 2020), Omicron was highly mutated with 58 SNP mutations and 22 amino acid mutations in the RBD. These mutations are crucial as they are known to increase the overall risk of reinfection and partial resistance to existing vaccines (Geddes, 2021). In addition to mutations in the S protein, several substitutions and deletions in other genomic regions are also present in all the SARS-CoV-2 variants in this study. Moreover, mutations have an adverse impact on the pathogenicity of SARS-CoV-2 and the development of diagnostic assays, antivirals, and vaccines (Amoutzias et al., 2022; Gomez et al., 2021). Therefore, monitoring of mutations and characterization of their roles in virulence-related conditions in SARS-CoV-2 is very vital in the control and prevention of the spread of the virus.

The phylogenetic analysis and characterization of the SARS-CoV-2 genomes in Southern Province provided information on the distribution, transmission and the identification of mutations. These findings are cardinal in preventing further transmission of the COVID-19 pandemic.

Understanding the distribution of the SARS-CoV-2 variants and the different lineages is vital for policy makers to know what control measures to apply in each region. In addition, some of the mutations that were identified in this study (such as D614G, N501Y, E484K and P314L) have been associated with increased viral transmission, increased binding affinity to the RBD, immune evasion and enhanced viral replication. Despite the study only being able to sequence 20% of the samples collected, it was able to identify and characterize the SARS-CoV-2 variants. This information is critical in coming up with the right control measures and developing the right vaccines to fight the pandemic. Phylogenetic analysis showed that there were multiple introductions of the COVID-19 pandemic into Southern Province from Europe and Africa.

CHAPTER SIX
CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

The findings highlighted the circulation of four VOCs in the Southern Province of Zambia namely Alpha (B.1.1.7), Beta (B.1.351), Delta (AY.116), and Omicron (BA.1, BA.1.1, BA.1.14 and BA.2). Phylogenetic analysis showed that the genomes in this study were closely related to genomes from Europe and Southern Africa indicating intra- and intercontinental introductions of the virus into the country. Additionally, some sequences that clustered with Zambian sequences may signify local transmission of the virus. SNP analysis also showed high rates of viral mutation and evolution which may lead to the emergence of novel SARS-CoV-2 variants, whose full potential on infection and pathogenicity have not been fully understood.

6.2 Recommendations

There is need for improved SARS-CoV-2 genomic surveillance in the country not only at points of entry but the whole country. Studies to monitor and evaluate the impact of these SARS-CoV-2 variants will help prevent on-going as well as future outbreaks. There is also need to strengthen the genomic capacity of the country by investing in the country's scientific needs. The Government should also invest in health care infrastructure and research decentralization to help cover rural areas where electricity and transport are a challenge.

REFERENCE

Amoutzias, G. D., Nikolaidis, M., Tryfonopoulou, E., Chlichlia, K., Markoulatos, P.& Oliver, S. G., 2022. The Remarkable Evolutionary Plasticity of Coronaviruses by Mutation and Recombination: Insights for the COVID-19 Pandemic and the Future Evolutionary Paths of SARS-CoV-2. *Viruses,* 14(78).

Banho, C. A., Sacchetto, L., Campos, G. R. F., Bittar, C., Possebon, F. S., Ullman, L. S., Marques, B. de C., da Silva, G. C. D., Moraes, M. M., Parra, M. C. P., Negri, A. F., Boldrin, A. C., Barcelos, M. D., dos Santos, T. M. I. L., Milhim, B. H. G. A., Rocha, L. C., Dourado, F. S., dos Santos, A. L., Ciconi, V. B., Patuto, C., Versiani, A. F., da Silva, R. A., de Oliviera-Lobl, E. E., Hernandes, V. M., Zini, N., Pacca, C. C., Estofolete, C. F., Ferreira, H. L., Rahal, P., Araujo jr, J. P., Cohen, J. A., Kerr, C. C., Althouse, B. M., Vasilakis, N. & Nogueira, M. L., 2022. Imapct of SARS-CoV-2 Gamma lineage introduction and COVID-19 vaccination on the epidemiological landscape of a Bazilian city. *Communications Medicine,* 2(41).

Centers for Disease Control and Prevention, 2021. *Global Health-Stories: Taking on COVID-19 in Zambia.* [Online], Available at: www.cdc.gov/globalhealth/stories/2021/taking-on-covid-19-in-zambia.html, [Accessed 10 September 2022].

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X. & Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet,* Volume 395, pp. 507-513.

European Centre for Disease Prevention and Control, 2020. *Guidance on infection prevention and control of coronavirus disease (COVID-19)in migrant and refugee reception and detention centres in the EU/EEA and the United Kingdom,* Stalkhom: European Centre for Disease Prevention and Control.

Flores-Alanis, A., Cruz-Rangel , A., Rodriguez-Gomez, F., Gonzalez, J., Torres-Guerrero, C. A., Delgado, G., Cravioto, A. & Morales-Espinosa, R., 2021. Molecular Epidemiology

Surveillance of SARS-CoV-2 : Mutations and Genetic Diversity One Year after Emerging. *Pathogens* , 10(2).

Gan, W. H. & Koh, D., 2021. COVID-19 and Return-To-Work for the Construction Sector: Lessons from Singapore. *Safety and Health at Work,* Volume 12, pp. 277 - 281.

Geddes, L., 2021. *From Alpha to Omicron: Everything you need to know about SARS-CoV-2 variant of Concern; Gavi Vaccine Alliance.* [Online], Available at: www.gavi.org/vaccineworks/alpha-omicron everything-you-need-know-about-coronavirus-variants-concern, [Accessed 16 September 2022].

Giandhari, J., Pillay, S., Wilkinson, E., Tegally, H., Sinayskiy, I., Schuld, M., Luorenco, J., Chimukangaara, B., Lessells, R., Moosa, Y., Gazy, I., Fish, M., Singh, L., Khanyile, K. S., Fonseca, V., Giovanetti, M., Alcantara Jr., L. C., Petruccione, F. & de Oliviera, T., 2021. Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. *International Journal of Infectious Diseases,* Volume 103, pp. 234-241.

Gill, C. J., Mwananyanda, L., MacLeod, W., Kwenda, G., Pieciak, R., Etter , L., Bridges, D. J., Chikoti, C., Chirwa , S., Chimonga, C., Forman, L., Katowa, B., Lapidot, R., Lungu, J., Matoba, J., Mwiinga, G., Mubemba, B., Mupila, Z., Muleya, W., Mwenda, M., Ngoma, B., Nkazwe, R., Nzara, D., Pawlak, N., Pemba, L., Saasa, N., Simulundu, E., Yankonde, B. & Thea, D. M., 2022. Sustained high prevalence of COVID-19 deaths from a systematic post-mortem study in Lusaka, Zambia: one year later. *MedRxiv.*

GISAID, 2020. *Pandemic Coronavirus causing Covid-19.* [Online], Available at: www.epicov.org/epi3/frontend#1739f3, [Accessed 15 September 2022].

GISAID, 2020. *Pandemic Coronavirus causing Covid-19.* [Online], Available at: www.epicov.org/epi3/frontend#1739f3, [Accessed 15 September 2022].

Gomez, C. E., Perdiguero, B. & Esteban, M., 2021. Emerging SARS-CoV-2 Variants and Impact in Global Vaccination Programs against SARS-CoV-2/COVID-19. *Vaccine,* 9(3).

Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O., 2005. PHYML Online-a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acid Research,* 33(2), pp. W557-W559.

Haddad, D., John, S. E., Mohammad, A., Hammad, M. M; Hebbar , P., Channanath, A., Nizam , R., Al-Qabandi, S., Al Madhoun, A., Alshukry, A., Ali, H., Thanaraj, T. A. & Al-Mula, F., 2021. SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLoS One,* Volume 16.

Hirabara, S. M., Serdan, T. D. A., Gorjao, R., Masi, L. N., Pithon-Curi, T. C., Corvas, D. T., Curi, R. & Durigon, E. L., 2022. SARS-CoV-2 Variants: Differences and Potential of Immune Evasion. *Front Cell Infect Microbiol,* 11(781429).

Holmes, E. C., Goldstein, S. A., Rasmussen, A. L., Robertson, D. L., Crits-Christoph, A., Wertheim, J. O., Anthony, S. J., Barclay, W. S., Boni, M. F., Doherty, P. C., Farrar, J., Geoghegan, J. L., Jiang, X., Leibowitz, J. L., Neil, Stuart J.D., Skern, T., Weiss, S. R., Worobey, M., Andersen, K. G., Garry, R. F. & Rambaut, A., 2021. The origins of SARS-CoV-2: A critical review. *Cell,* Volume 184, pp. 4848 - 4856.

Huang, C., Wang, Y., Li, X., Ren, L., Zhoa, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J. & Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirusin Wuhan, China. *The Lancet,* Volume 395, pp. 497-506.

Hu, B., Guo, H., Zhou, P. & Shi, Z. L., 2021. Characteristics of SAR-CoV-2 and COVID-19. *Nature ,* Volume 19, pp. 141-154.

Holmes, E. C., Goldstein, S. A., Rasmussen, A. L., Robertson, D. L., Crits-Christoph, A., Wertheim, J. O., Anthony, S. J., Barclay, W. S., Boni, M. F., Doherty, P. C., Farrar, J., Geoghegan, J. L., Jiang, X., Leibowitz, J. L., Neil, Stuart J.D., Skern, T., Weiss, S. R., Worobey, M., Andersen, K. G., Garry, R. F. & Rambaut, A., 2021. The origins of SARS-CoV-2: A critical review. *Cell,* Volume 184, pp. 4848 - 4856.

Isabel, S., Grana-Miraglia, L., Guttierrez , J. M., Bundavolv-Torma, C., Groves, H. E., Isabel, M. R., Eshaghi, A., Patel, S. N., Gubbay, J. B., Poutanen, T., Guttman, D. S. & Poutanen, S. M., 2020. Evolutionary and structural analysis of SARS-CoV-2 D614G spike protein mutation now documented worldwide. *Scientific Reports* , 10(14031).

Kang, N., Liu, Y., Bi, F., Wang, J., Chen, H., wang, J., Liang, Z., Chen, M., Wei, G., Li, A., Li , Y., Lu, B., Jiang, Z., Tan, Y., Zhong , G., Lin, M. & Ju, Y., 2021. Detection of Variants of B.1.617 Lineage from Five Returning Chinese Nationals at a Guangxi-Vietnam Border Port Guangxi Zhuang Autonomous Region, China,April 2021. *China CDC Weekly*, 3(30), pp. 653-655.

Kapata, N., Ihekweazu, C., Ntoumi, F., Raji, T., Chanda-Kapata, P., Mwaba, P., Mukonka, V., Bates, M., Tembo, J., Corman, V., Mfinanga, S., Asogun, D., Elton, L., Arruda, L. B., Thomason, M. J., Mboera, L., Yavlinsky, A., Haider, N., Simons, D., Hollmann, L., Lule, S. A., Veas, F., Hamid, M. M. A., Dar, O., Edwards, S., Vairo, F., McHugh, T. D., Drosten, C., Kock, R., Ippolito, G., & Zumla, A., 2020. Is Africa prepared for tackling the COVID-19 (SARS-CoV-2) epidemic. Lessons from past outbreaks, ongoing pan-African public health efforts, and implications for the future. *International Journal of Infectious Diseases*, Volume 93, pp. 233 - 236.

Katoh, K., Rozewicki, J. & Yamada, K. D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, Volume 20, pp. 1160-1166.

Kober, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Patridge, D. G., Evans, C. M., Freeman, T. M. & de Silva, T. I., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of COVID-19 Virus. *Cell*, 182(4), pp. 812-827.

LeFort, V., Longueville, J. E. & Gascuel, O., 2017. SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, 34(9), pp. 2422-2424.

Letunic, I. & Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acid Research,* 49(W1), pp. W293-W296.

Li, F., Li, Y. Y., Liu, M. J., Fang, L. Q., Dean, N. E., Wong, G. W. K., Yang, X. B., Longini, I., Halloran, E. M., Wang, H. J., Liu, P. L., Pang, Y. H., Yan, Y. Q., Liu, S., Xia, W., Lu, X. X., Liu, Q., Yang, Y. & Xu, S. Q., 2021. Household transmission of SARS-CoV-2 and risk factors for susceptibility and infectivity in Wuhan: a retrospective observational study. *The Lancet Infectious Disease,* Volume 21, pp. 617-628.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S.M., Lau, E. H.Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen , X., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Lui, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T.Y., Wu, J. T., Goa, G. F., Crowling, B. J., Yang, B., Leung, G. M. & Feng, Z., 2020. Early Transmission Dynamics in Wuhan, China, of Novel coronavirus-Infected Pneumonia. *The New England Journal of Medicine,* Volume 382, pp. 1199-1207.

Lin, J. W., Tang, C., Wei, H. C., Du, B., Chen, C., Wang, M., Zhou, Y., Yu, M., Cheng, L., Kuivanen, S., Oganda, N. S., Levanov, L., Zhao, Y., Li, C., Zhou, R., Li, Z., Zhang, Y., Sun, K., Wang, C., Chen, L., Xiao, X., Zheng, X., Chen, S., Zhou, Z., Yang, R., Zhang, D., Xu, M., Song, J., Wang, D., Li, Y., Lei, S. K., Zeng, W., Yang, Q., He, P., Zhang, Y., Zhou, L., Coa, L., Luo, F., Liu, H., Yang , L., Ye, F., Zhang, M., Li, M., Fan, W., Li, X., Li, K., Ke, B., Xu, J., Yang, H., He, S., Pan, M., Yan, Y., Zha, Y., Jiang, L., Yu, C., Liu, Y., Xu, Z., Li, Q., Jiang, Y., Sun, J., Hong, W., Wei, H.,; Lu, Guangwen; V. O., Luo, Y., Wei, Y., Connor, T., Tan, W., Snijder, E. J., Smura, T., Li, W., Geng, J., Ying, B. & Chen, L., 2021. Genomic Monitoring of SARS-CoV-2 Uncovers an NSP1 deletionvariant that modulates type 1 interferon response. *Cell Host &Microbe,* Volume 29, pp. 489-502.

Loemb'e, M. M., Tshangela, A., Salyer, S. J., Varma, J. K., Ouma, A. E. O., Nkegasong, J. N., 2020. COVID-19 in Africa: the spread and response. *Nature Medicine,* Volume 26, pp. 999 - 1003.

Lone, S. A. & Ahmad, A., 2020. COVID-19 Pandemic- an African Perspectiive. *Emerging Microbes &Infections* , 9(1), pp. 1300 - 1308.

Lu, X., Wang, L., Sakthivel, S. K., Whitaker, B., Murray, J., Kamili, S., Lynch, B., Malapati, L., Burke, S. A., Harcourt, J., Tamin, A., Thornburg, N. J., Villanueva, J. M. & Lindstrom, S., 2020. US CDC Real-Time Reverse Transcription PCR Panel for Detection of Severe Acute Respirotory Syndrome Coronavirus 2. *Emerging Infectious Diseases,* 26(8), pp. 1654 - 1666.

Morobe, J. M., Pool, B., Marie, L., Didon, D., Lambisia, A. W., Makori, T., Mohammed, K. S., de Laurent, Z. R., Ndwiga, L., Mburu, M. W., Moraa, E. & Murunga, N., 2022. Genomic Epidemiology of SARS-CoV-2 in Seychelles, 2020-2021. *Viruses,* 14(1318).

Munjal, G., Hanmandlu, M. & Srivastava, S., 2019. Phylogenetics Algorithims and Applications. *Ambient Communications and Computer Systems* , Volume 904, pp. 187-194.

Mwenda, M., Saasa, N., Sinyangwe, N., Busby, G., Chipimo, P. J., Hendry, J., Kapona, O., Yingst, S., Hines , J. Z., Minchella, P., Simulundu, E., Changula, K., Nalubamba, K. S., Sawa, H., Masahiro, K., Yamagishi, J., Kapin'a, M., Kapata , N., Fwoloshi, S., Zulu, P., Mulenga , L. B., Agolory, S., Mukonka, V. & Bridges , D. J., 2020. Detection of B.1.351 SARS-CoV-2 Variant Strain-Zambia. *Morbidity and Mortality Weekly Report,* 70(8), pp. 208-282.

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., Atif, S. M., Hariprasad, G., Hasan, G. M. & Hassan, M. I., 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and Ttherapies: Structural genome approach. *Molecular Basis of Disease,* 1866(165878).

Niang, L., Winn, T. & Rusli, B., 2006. Practical Issues in Calculating the Sample Size for Prevalence Studies. *Archives of Orofacial Sciences,* Volume 1, pp. 9 - 14.

Obeid, D. A., Alsanea, M. S., Alnemari, R. T., Almaghrabi, R. S., Alhadheq, F. M., Alahideb, B. M. & Alhamlan, F. S., 2021. SARS-CoV-2 genetic diversity and variants of concern in Saudi Arabia. *Journal of Infection in Developing Countries,* 15(12).

Power Technology, 2020. *Kafue Gorge Lower (KGL) Power Station.* [Online], Available at: www.power-technology.com/projects/kafue-gorge-lower-kgl-power-station/, [Accessed 20 September 2022].

Quick, J., 2020. *Protocols.io.* [Online], Available at: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye?step=5, [Accessed 20 November 2021].

Rambaut, A., Lam, T. T., Carvalho, L. M. & Pybus, O. G., 2016. Exploring the temporal structure of Heterochronous sequences using TempEst (Formaly Path-O-Gen). *Virus Evolution*, 2(vew007).

Reuters, 2022. *REUTERS COVID-19 TRACKER.* [Online], Available at: https://graphics.reuters.com/world-coronavirus-tracker-maps/countries-and-territories/zambia/, [Accessed 08 September 2022].

Rothan, H. A. & Byrareddy, S. N., 2020. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of Autoimmunity,* 1024(33).

Saberiyan, M., Karimi, E., Khademi, Z., Movahhed, P., Safi, A. & Mehri-Ghahfarrokhi, A., 2022. SARS-CoV-2: phenotype, genotype and characterization of different variants. *Cellular & Molecular Biology Letters,* 27(50).

Sallam, M., Ababneh, N. A., Dababseh, D., Bakri, F. G. & Mahafzah, A., 2021. Temporal increase in D614G mutation of SARS-CoV-2 in the Middle East and North Africa. *heliyon,* 7(1).

Saxena, S. K., Kumar, S., Ansari, S., Paweska, J. T., Maurya, V. K., Tripathi, A. K. & Abdel-Moneim, A. S., 2022. Characterizationof the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. *J Med virol,* Volume 94, pp. 1738-1744.

Shaheen, M. N., Elmahdy, E. M. & Shahein, Y. E., 2022. The first detection of SARS-CoV-2 RNA in urban wastewater in Giza, Egypt. *Journal of Water & Health,* 20(8), pp. 1212-1222.

Simulundu, E., Mupeta, F., Kapata-Chanda, P., Saasa, N., Changula, K., Muleya, W., Chitanga, S., Mwanza, M., Simusika, P., Chambaro, H., Mubemba, B., Kajihara, M., Chanda, D., Mulenga, L., Fwoloshi, S., Shibemba, A. L., Kapaya, F., Zulu, P., Musonda, K., Monze, M., Sinyangwe, N., Mazaba, M. L., Kapin'a, M., Chipimo, P. J., Hamoonga, R., Simwaba, D.,

Ngosa , W., Morales, A. N., Kayeyi, N., Tembo, J., Bates, M., Orba, Y., Sawa, H., Takada, A., Nalubamba, K. S., Malama, K., Mukonka, V., Zumla , A. & Kapata, N., 2020.'First COVID-19 case in Zambia — Comparative phylogenomic analyses of SARS-CoV-2 detected in African countries. *International Journal of Infectious Disease,* 102(2021), pp. 455-459.

Singanayagam, A., Hakki, S., Dunning, J., Madon, K. J., Crone, M. A., Koycheva, A., Derqui-Fernandez, N., Barnett, J. L., Whitfield , M. G., Varro, R., Charlett, A., Kundu, R., Fenn, J., Cutajar, J., Quinn, V., Conibear, E., Barclay, W., Freemont, P. S., Taylor, G. P., Ahmad, S., Zambon, M., Ferguson, N. M. & Lalvani, A., 2021. Community transmission and Viral Kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK:a prospective, longitudinal, cohort study. *The Lancet,* 22(2), pp. 183 - 195.

Singh, D. & Yi, S. V., 2021. On the origin and evolution of SARS-CoV-2. *Experimental and Molecular Medicine,* Volume 53, pp. 537 - 547.

Singh, J., Pandit, P., McArthur, A. G., Banerjee, A. & Mossman, K., 2021. Evolutionary trajectory of SARS-CoV-2 and Emerging variants. *Virology Journal,* 18(166).

Srivastava, S., Banu, S., Singh, P., Sowpati, D. T. & Mishra, R. K., 2021. SARS-Cov-2 genomics: An Indian Perspective on sequencing Viral Variants. *Journal of Bioscience,* 46(22).

Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., Liu, W., Bi, y. & Gao, G. F., 2016. Epidemiology, Genetic Recombination and Pathogenesis of Corona Viruses. *Trends in Microbiology*, 24(6), pp. 490 - 502.

Sun, C., Xie, C., Bu, G.L., Zhong, L.Y. & Zeng, M.S., 2022. Molecular characteristics, immune evasion and impact of SARS-CoV-2 variants. *Signal Transduction and Targeted Therapy*, 7(202).

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., Gottenberg, A. V., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A. J., Engelbrecht, S., Zyl, G. V., Preier, W., Petruccione, F., Sigal, A., Hardie, D., Marias, G., Hsiao, N. Y., Korsman, S., Davies , M. A., Tyers, L.,

Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisotani-Dehkordi, A., Godzik, A., Wibmer, C. K., Sewell, B. T., Lourenco, J., Jr Alcantara, L. C., Pond, S. L. K., Weaver, S., Martin, D., Lessells, R. J., Bhiman, J. N., Williamson, C. & de Oliviera , T., 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* , 592(7854), pp. 438 - 443.

Telenti, A., Hadcroft, E. B. & Robertson, D. L., 2022. The Evolution and Biology of SARS-CoV-2 Variants. *Cold Spring Habour Perspective in Medicin,* 12(5).

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., Choga, W. T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., Giovanetti, M., Hardie, D., Hill, V., Hsiao, N. Y., Iranzadeh, A. & de Oliviera, T., 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in Southern Africa. *Nature,* Volume 603, pp. 679-686.

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg , L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, A., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Goncalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D. K. & Gaythorpe, K., The COVID-19 Genomics UK (COG-UK) consortium, Ferguson, N. M., 2021.'Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature,* Volume 593, p. 266 269.

WHO African Region, 2022. *COVID-19 (WHO African Region).* [Online], Available at: https://who.map.arcgis.com/apps/dashboards/0c9b3a8b68d0437a8cf28581e9c063a9, [Accessed 28 September 2022].

WHO-China Study, 2021. *WHO_convened Global Study of Origins of SAR_CoV-2: China Part,* Hong Kong: World Health Organization.

Wilkinson, E., Giovanetti, M., Tegally, H., San, J. E., Lessells, R., Cuadros, D., Martin, D. P., Rasmussen, D. A., Zekri, A. R. N., Sangare, A. K., Ouedraogo, A. S., Sesay, A. K.; Priscilla, A., Kemi, A. S., Olubusuyi, A. M., Oluwapelumi, A. O. O., Hammami, A., Amuri, A. A., Sayed, A., Oua, A. E. O., Elargoubi, A., Ajayi, N. A. & de Oliviera, T. et

al., 2021. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science,* Issue 374, pp. 423 - 431.

World Health Organization, 2020a. *A Second COVID-19 Case is Confirmed in Africa.* [Online], Available at: www.afro.who.int/news/second-covid-19-case-confirmed-africa, [Accessed 13 September 2022].

World Health Organization , 2020b. *.afro.who.int.* [Online], Available at: https://www.afro.who.int/sites/default/files/COVID-19%20situation%20reports/COVID-19-Bulletin_18%20Feb.pdf, [Accessed 05 April 2021].

World Health Organization, 2020c. *SARS-CoV-2 antigen-detecting rapid diagnostic tests: An implementation guide,* Geneva: World Health Organization.

World Health Organization, 2022a. *WHO Coronavirus (COVID-19) Dashboard.* [Online], Available at: https://covid19.who.int/, [Accessed 19 September 2022].

World Health Organization, 2022b. *Tracking SARS-CoV-2 Variants.* [Online], Available at: www.who.int/activities/tracking-SARS-CoV-2-variants, [Accessed 10 September 2022].

Worldometer, 2022. *COVID-19 coronavirus pandemic.* [Online], Available at: www.worldometers.info/coronaviruses/#countries, [Accessed 15 September 2022].

Yang, X., Dong, N., Chan, E. W. C. & Sheng, C., 2020. Genetic cluster analysis of SARS-CoV-2 and the identification of those responsible for the major outbreaks in various countries. *Emerging Microbes and Infections,* Volume 9, pp. 1287-1299.

Zambia National Public Health Institute, 2020a. *Zambia Situation Report No. 44,* Lusaka: Zambia National Public Health Institute.

Zambia National Public Health Institute, 2020b. *Zambia Situation Report No. 60,* Lusaka: Zambia National Public Health Institute.

Zambia National Public Health Institute, 2022. *Zambia COVID-19.* [Online], Available at: https://nsdi-mlnr.maps.arcgis.com/apps/dashboard/c08c4cce115244f7ba472a458e1483f7, [Accessed 22 September 2022].

Zhang, L., Jackson, C. C., Mou, H., Ojha, A., Peng, H., Quilan, B. D., Rangarajan, E. S., Pan, A., Vanderheiden, A., Suthar, M. S., Li, W., Izard, T., Rader, C., Farzan, M. & choe, H., , 2020. SARS-CoV-2 Spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications* , 11(6013).

APPENDIX

Appendix 1: SNP and AA mutation table

No.	Position	Genome Region	No. of Mutations	Reference Nucleotide	Sample Nucleotide	AA Change	Frequency	Mutation Type
1	174	5'UTR	4	*G*	T	174	6	extragenic
2	210			*G*	T	210	4	extragenic
3	222			*C*	T	222	1	extragenic
4	241			*C*	T	241	31	extragenic
5	343	NSP1	3	.	C	V26	2	Insertion
6	478			*C*	T	I71I	1	Synonymous
7	670			*T*	G	S135R	5	Missense
8	913	NSP2	5	*C*	T	S36S	3	Synonymous
9	926			*T*	C	F41L	2	Missense
10	940			*G*	A	K45K	1	Synonymous
11	1059			*C*	T	T85I	6	Missense
12	2692			*A*	T	T629T	2	Synonymous
13	2790	NSP3	42	*C*	T	T24I	5	Missense
14	2832			*A*	G	K38R	20	Missense
15	2903			*A*	G	I62V	1	Missense
16	3037			*C*	T	F106F	39	Synonymous
17	3177			*C*	T	P153L	3	Missense
18	3267			*C*	T	T183I	3	Missense
19	4150			*T*	C	V477V	2	Synonymous
20	4181			*G*	T	A488S	4	Missense
21	4184			*G*	A	G489S	5	Missense
22	4255			*G*	T	P512P	2	Synonymous
23	4276			*C*	T	Y519Y	2	Synonymous

24	4321			*C*	T	A534A	5	Synonymous
25	4540			*C*	T	Y607Y	2	Synonymous
26	4576			*A*	T	T619T	2	Synonymous
27	4579			*T*	A	L620L	3	Synonymous
28	4655			*C*	T	R646W	1	Missense
29	5161			*T*	A	F814L	1	Missense
30	5230			*G*	T	K837N	6	Missense
31	5260			*T*	A	T847T	1	Synonymous
32	5386			*T*	G	A889A	22	Synonymous
33	5388			*C*	A	A890D	3	Missense
34	5866			*C*	T	F1049F	1	Synonymous
35	5953			*G*	A	L1078L	2	Synonymous
36	5986			*C*	T	F1089F	3	Synonymous
37	6106			*T*	C	Y1129Y	1	Synonymous
38	6402			*C*	T	P1228L	4	Missense
39	6513			*GTT*	.	S1265	21	Deletion
40	6568			*C*	T	D1283D	1	Synonymous
41	6954			*T*	C	I1412T	3	Missense
42	6968			*C*	T	L1416L	1	Synonymous
43	7069			*A*	G	E1450E	1	Synonymous
44	7071			*G*	T	G1451V	1	Missense
45	7086			*C*	T	T1456I	2	Missense
46	7124			*C*	T	P1469S	4	Missense
47	7252			*G*	C	L1511F	1	Missense
48	7292			*T*	C	G1524G	1	Synonymous
49	7787			*G*	C	A1690P	1	Missense
50	7923			*C*	T	S1735F	1	Missense
51	8035			*G*	T	M1772I	2	Missense
52	8290			*C*	T	L1857L	1	Synonymous

53	8393			*G*	A	A1892T	22	Missense
54	8499			*C*	T	A1927V	2	Missense
55	8723	NSP4	9	*A*	C	I57L	1	Missense
56	8986			*C*	T	D144D	4	Synonymous
57	9053			*G*	T	V167L	4	Missense
58	9344			*C*	T	L264F	5	Missense
59	9424			*A*	G	V290V	3	Synonymous
60	9474			*C*	T	A307V	1	Missense
61	9534			*C*	T	T327I	5	Missense
62	9866			*C*	T	L438F	5	Missense
63	10029			*C*	T	T492I	31	Missense
64	10181	NSP5	7	*A*	G	I43V	1	Missense
65	10198			*C*	T	D48D	2	Synonymous
66	10323			*A*	G	K90R	6	Missense
67	10447			*G*	A	R131R	5	Synonymous
68	10449			*C*	A	P132H	27	Missense
69	10691			*A*	G	I213V	4	Missense
70	10892			*A*	G	T280A	1	Missense
71	11083	NSP6	9	*G*	T	L37F	1	Missense
72	11109			*C*	T	A46V	1	Missense
73	11201			*A*	G	T77A	4	Missense
74	11286			*TGTCTGGTT*	.	L105	22	Deletion
75	11288			*TCTGGTTTT*	.	S106	14	Deletion
76	11332			*A*	G	V120V	4	Synonymous
77	11537			*A*	G	I189V	20	Missense
78	11653			*C*	T	L227L	2	Synonymous
79	11713			*T*	C	S247S	1	Synonymous

80	11941	NSP7	1	*C*	T	V33V	1	Synonymous
81	12406	NSP8	4	*C*	T	N105N	1	Synonymous
82	12452			*C*	T	P121S	2	Missense
83	12473			*C*	T	K127K	1	Synonymous
84	12557			*A*	C	I156L	2	Missense
85	12786	NSP9	2	*C*	T	T34I	1	Missense
86	12880			*C*	T	I65I	5	Synonymous
87	13195	NSP10	2	*T*	C	V57V	22	Synonymous
88	13339			*T*	A	N105K	2	Missense
89	13592	NSP12b	11	*C*	T	T42I	2	Missense
90	13756			*A*	G	I97V	1	Missense
91	14408			*C*	T	P314L	40	Missense
92	14676			*C*	T	P403P	3	Synonymous
93	15240			*C*	T	N591N	22	Synonymous
94	15279			*C*	T	H604H	3	Synonymous
95	15451			*G*	A	G662S	4	Missense
96	15714			*C*	T	L749L	5	Synonymous
97	15939			*T*	C	D824D	1	Synonymous
98	15960			*C*	T	A831A	1	Synonymous
99	16176			*T*	C	T903T	3	Synonymous
100	16406	NSP13	6	*T*	G	V57G	2	Missense
101	16466			*C*	T	P77L	4	Missense
102	16636			*G*	T	A134S	1	Missense
103	17280			*G*	T	V348V	1	Synonymous
104	17410			*C*	T	R392C	5	Missense
105	17999			*C*	T	T588I	4	Missense
106	18163	NSP14	2	*A*	G	I42V	27	Missense
107	19220			*C*	T	A394V	4	Missense
108	19684	NSP15	8	*G*	T	V22L	1	Missense

109	19803			*C*	T	R61R	4	Synonymous
110	19884			*C*	T	Y88Y	1	Synonymous
111	19955			*C*	T	T112I	5	Missense
112	19999			*G*	T	V127F	1	Missense
113	20055			*A*	G	E145E	5	Synonymous
114	20094			*A*	G	K158K	2	Synonymous
115	20132			*C*	T	A171V	2	Missense
116	20741	NSP16	23	*A*	G	Q28R	1	Missense
117	21082			*GA*	CT	E142L	2	Missense
118	21110			*C*	T	T151I	4	Missense
119	21512			*A*	C	N285T	4	Missense
120	21517			*AG*	CA	R287Q	4	Missense
121	21520			*G*	C	V288L	4	Missense
122	21523			*GTT*	TCG	V289S	4	Missense
123	21528			*T*	C	I290I	4	Synonymous
124	21532			*A*	T	S292C	4	Missense
125	21534			.	A	S292	2	Insertion
126	21535			*G*	.	S292	1	Deletion
127	21536			.	TC	D293	1	Insertion
128	21537			*T*	A	D293E	1	Missense
129	21537			*T*	.	D293	1	Deletion
130	21539			*T*	A	V294D	2	Missense
131	21542			*T*	.	L295	1	Deletion
132	21543			*TGTT*	AACG	V296T	1	Missense
133	21543			*TGTTA*	.	L295	1	Deletion
134	21547			*AA*	CT	N297L	1	Missense
135	21547			*AA*	.	V296	1	Deletion
136	21549			*C*	G	N297K	1	Missense
137	21550			*A*	.	N297	1	Deletion

138	21551			*A*	T	N298I	1	Missense
139	21614	Spike	82	*C*	T	L18F	1	Missense
140	21618			*C*	G	T19R	9	Missense
141	21618			*C*	T	T19I	5	Missense
142	21633			*TACCCCCTG*	.	L24	5	Deletion
143	21762			*C*	.	A67	22	Deletion
144	21764			*A*	.	A67	22	Deletion
145	21765			*TACATG*	.	H69	3	Deletion
146	21767			*CATG*	.	I68	22	Deletion
147	21801			*A*	C	D80A	6	Missense
148	21846			*C*	T	T95I	26	Missense
149	21987			*G*	A	G142D	9	Missense
150	21987			*GTGTTTATT*	.	G142	22	Deletion
151	21993			*ATT*	.	Y145	3	Deletion
152	22029			*AGTTCA*	.	E156	4	Deletion
153	22139			*G*	C	V193L	1	Missense
154	22161			*A*	G	Y200C	1	Missense
155	22193			.	T	I210	16	Insertion
156	22194			*ATT*	.	N211	2	Deletion
157	22195			*T*	G	N211K	16	Missense
158	22197			*TA*	GC	L212C	16	Missense
159	22200			*T*	G	V213G	5	Missense
160	22201			.	AGC	S214	16	Insertion
161	22202			.	A	V213	1	Insertion
162	22203			.	A	R214	1	Insertion
163	22204			*T*	A	R214R	16	Synonymous
164	22206			*A*	G	D215G	6	Missense

165	22286			*CTTGCTTTA*	.	L242	4	Deletion
166	22289			*G*	T	A243S	1	Missense
167	22293			*T*	C	L244S	1	Missense
168	22481			*A*	C	T307P	1	Missense
169	22511			*AA*	TT	N317F	1	Missense
170	22535			*T*	C	S325P	1	Missense
171	22539			*TT*	AA	I326K	1	Missense
172	22542			*T*	C	V327A	1	Missense
173	22578			*G*	A	G339D	27	Missense
174	22599			*G*	A	R346K	5	Missense
175	22669			*T*	C	Y369Y	1	Synonymous
176	22673			*TC*	CT	S371L	22	Missense
177	22674			*C*	T	S371F	5	Missense
178	22679			*T*	C	S373P	27	Missense
179	22686			*C*	T	S375F	27	Missense
180	22688			*A*	G	T376A	5	Missense
181	22775			*G*	A	D405N	5	Missense
182	22786			*A*	C	R408S	5	Missense
183	22813			*G*	T	K417N	31	Missense
184	22882			*T*	G	N440K	7	Missense
185	22898			*G*	A	G446S	16	Missense
186	22917			*T*	G	L452R	4	Missense
187	22969			*AA*	GG	T470A	1	Missense
188	22992			*G*	A	S477N	21	Missense
189	22995			*C*	A	T478K	24	Missense
190	23012			*G*	A	E484K	6	Missense
191	23013			*A*	C	E484A	20	Missense
192	23040			*A*	G	Q493R	20	Missense

193	23048			*G*	A	G496S	15	Missense
194	23055			*A*	G	Q498R	20	Missense
195	23063			*A*	T	N501Y	29	Missense
196	23075			*T*	C	Y505H	19	Missense
197	23202			*C*	A	T547K	22	Missense
198	23271			*C*	A	A570D	3	Missense
199	23403			*A*	G	D614G	40	Missense
200	23525			*C*	T	H655Y	27	Missense
201	23599			*T*	G	N679K	27	Missense
202	23604			*C*	G	P681R	29	Missense
203	23604			*C*	A	P681H	5	Missense
204	23664			*C*	T	A701V	6	Missense
205	23709			*C*	T	T716I	3	Missense
206	23854			*C*	A	N764K	27	Missense
207	23948			*G*	T	D796Y	27	Missense
208	24130			*C*	A	N856K	22	Missense
209	24183			*C*	T	T874I	1	Missense
210	24410			*G*	A	D950N	4	Missense
211	24424			*A*	T	Q954H	27	Missense
212	24469			*T*	A	N969K	26	Missense
213	24503			*C*	T	L981F	21	Missense
214	24506			*T*	G	S982A	3	Missense
215	24821			*G*	T	A1087S	1	Missense
216	24872			*G*	T	V1104L	2	Missense
217	24914			*G*	C	D1118H	3	Missense
218	24942			*A*	G	D1127G	1	Missense
219	25000			*C*	T	D1146D	27	Synonymous
220	25352			*G*	T	V1264L	3	Missense
221	25469	ORF3a	9	*C*	T	S26L	4	Missense

222	25546			*C*	T	L52F	4	Missense
223	25563			*G*	T	Q57H	6	Missense
224	25584			*C*	T	T64T	27	Synonymous
225	25587			*C*	T	L65L	5	Synonymous
226	25665			*C*	T	Y91Y	1	Synonymous
227	25904			*C*	T	S171L	6	Missense
228	26058			*C*	T	D222D	1	Synonymous
229	26060			*C*	T	T223I	5	Missense
230	26270	Envelop e	5	*C*	T	T9I	27	Missense
231	26296			*C*	A	L18I	1	Missense
232	26392			*A*	T	S50C	1	Missense
233	26438			*T*	C	L65P	2	Missense
234	26456			*C*	T	P71L	6	Missense
235	26530	Membra ne	7	*A*	G	D3G	19	Missense
236	26577			*C*	G	Q19E	25	Missense
237	26709			*G*	A	A63T	24	Missense
238	26767			*T*	C	I82T	4	Missense
239	26858			*C*	T	F112F	4	Synonymous
240	26997			*T*	G	C159G	1	Missense
241	27059			*C*	T	Y179Y	4	Synonymous
242	27193	3’UTR	1	*T*	C	27193	1	extragenic
243	27209	ORF6	4	*A*	G	H3R	1	Missense
244	27259			*A*	C	M19M	27	Synonymous
245	27382			*GAT*	CTC	D61L	5	Missense
246	27384			*T*	C	D61D	1	Synonymous
247	27638	ORF7a	2	*T*	C	V82A	4	Missense
248	27752			*C*	T	T120I	4	Missense

249	27807	ORF7b	9	*C*	T	L17L	20	Synonymous
250	27873			.	G	E39	1	Insertion
251	27874			*C*	A	T40N	1	Missense
252	27874			*C*	T	T40I	4	Missense
253	27875			.	C	T40	1	Insertion
254	27879			*C*	.	C41	1	Deletion
255	27880			*A*	T	H42L	1	Missense
256	27882			*G*	.	H42	1	Deletion
257	27883			*C*	T	A43V	1	Missense
258	27921	ORF8	9	*A*	T	I10F	1	Missense
259	27972			*C*	T	Q27*	3	SNP_stop
260	27998			*C*	T	D35D	1	Synonymous
261	28048			*G*	T	R52I	3	Missense
262	28079			*G*	C	V62V	1	Synonymous
263	28111			*A*	G	Y73C	3	Missense
264	28248			*GATTTC*	.	D119	4	Deletion
265	28253			*C*	T	F120F	1	Synonymous
266	28253			*CA*	TC	I121L	5	Missense
267	28271	3’UTR	2	*A*	T	28271	27	extragenic
268	28273			*A*	.	28273	7	extragenic
269	28280	Nucleocapsid	20	*GAT*	CTA	D3L	3	Missense
270	28310			*C*	T	P13S	1	Missense
271	28311			*C*	T	P13L	26	Missense
272	28362			*GAGAACGCA*	.	E31	26	Deletion
273	28461			*A*	G	D63G	4	Missense
274	28881			*G*	T	R203M	2	Missense
275	28881			*GG*	TT	R203I	2	Missense

276	28881			*GGG*	AAC	RG203 KR	30	Missense
277	28887			*C*	T	T205I	6	Missense
278	28916			*G*	T	G215C	4	Missense
279	28977			*C*	T	S235F	3	Missense
280	28985			*G*	T	G238C	1	Missense
281	29014			*T*	C	T247T	1	Synonymous
282	29095			*C*	T	F274F	1	Synonymous
283	29144			*C*	T	E290E	1	Synonymous
284	29192			*T*	C	F307L	2	Missense
285	29304			*C*	T	P344L	5	Missense
286	29402			*G*	T	D377Y	4	Missense
287	29433			*A*	G	K387R	2	Missense
288	29510			*A*	C	S413R	5	Missense
289	29545	3’UTR	4	*C*	T	29545	2	extragenic
290	29734			*GAGGCCA CGCGGA GTACGAT CGAGTG*	.	29734	4	extragenic
291	29736			*G*	T	29736	1	extragenic
292	29742			*G*	T	29742	4	extragenic

Appendix 2: Alpha Varian SNP Mutation table

Ref genome position	Region/ Gene	# 0f Mutations	reference nucleotide	sample Nucleotide	AA change	Frequency	Mutation Type
	5'UTR	1					
241			C	T	241	3	extragenic
	NSP2	1					
913			C	T	S36S	3	Synonymous
	NSP3	8					
3037			C	T	F106F	3	Synonymous
3177			C	T	P153L	3	Missense
3267			C	T	T183I	3	Missense
4255			G	T	P512P	2	Synonymous
5388			C	A	A890D	3	Missense
5986			C	T	F1089F	3	Synonymous
6954			T	C	I1412T	3	Missense
8499			C	T	A1927V	2	Missense
	NSP6	1					
11288			TCTGGTTTT	.	S106	3	deletion
	NSP12b	4					
14408			C	T	P314L	3	Missense

14676			C	T	P403P	3	Synonymous
15279			C	T	H604H	3	Synonymous
16176			T	C	T903T	3	Synonymous
	NSP15	1					
19684			G	T	V22L	1	Missense
	NSP16	19					
21082			GA	CT	E142L	2	Missense
21512			A	C	N285T	3	Missense
21517			AG	CA	R287Q	3	Missense
21520			G	C	V288L	3	Missense
21523			GTT	TCG	V289S	3	Missense
21528			T	C	I290I	3	Synonymous
21532			A	T	S292C	3	Missense
21534			.	A	S292	2	Insertion
21535			G	.	S292	1	Deletion
21536			.	TC	D293	1	Insertion
21537			T	.	D293	1	Deletion
21539			T	A	V294D	1	Missense
21542			T	.	L295	1	Deletion
21543			TGTT	AACG	V296T	1	Missense
21543			TGTTA	.	L295	1	Deletion
21547			AA	.	V296	1	Deletion

21549			C	G	N297K	1	Missense
21550			A	.	N297	1	Deletion
21551			A	T	N298I	2	Missense
	Spike	10					
21765			TACATG	.	H69	3	deletion
21993			ATT	.	Y145	3	deletion
23063			A	T	N501Y	3	Missense
23271			C	A	A570D	3	Missense
23403			A	G	D614G	3	Missense
23604			C	A	P681H	3	Missense
23709			C	T	T716I	3	Missense
24183			C	T	T874I	3	Missense
24506			T	G	S982A	3	Missense
24914			G	C	D1118H	3	Missense
	ORF8	3					
27972			C	T	Q27*	3	Stop
28048			G	T	R52I	3	Missense
28111			A	G	Y73C	3	Missense
	3'UTR	1					
28273			A	.	28273	3	extragenic
	Nucleocapsid	4					
28280			GAT	CTA	D3L	3	Missense

28881			GGG	AAC	RG203KR	3	Missense
28977			C	T	S235F	3	Missense
28985			G	T	G238C	1	Missense

Appendix 3: Beta Variant SNP Mutation table

Ref genome position	Region/ Gene	# 0f Mutations	reference nucleotide	sample Nucleotide	AA change	Frequency	Mutation Type
	5'UTR	2					
174			G	T	174	6	extragenic
241			C	T	241	6	extragenic
	NSP2	3					
940			G	A	K45K	1	Synonymous
1059			C	T	T85I	6	Missense
2692			A	T	T629T	2	Synonymous
	NSP3	6					
3037			C	T	F106F	6	Synonymous
4150			T	C	V477V	2	Synonymous
4276			C	T	Y519Y	2	Synonymous
4540			C	T	Y607Y	2	Synonymous
5230			G	T	K837N	6	Missense
6106			T	C	Y1129Y	1	Synonymous
	NSP4	1					
9344			C	T	L264F	1	Missense
	NSP5	2					
10181			A	G	I43V	1	Missense
10323			A	G	K90R	6	Missense
	NSP6	1					
11288			TCTGGTTTT	.	S106	6	deletion

	NSP8	1					
12473			C	T	K127K	1	Synonymous
	NSP10	1					
13339			T	A	N105K	2	Missense
	NSP12b	2					
13592			C	T	T42I	2	Missense
14408			C	T	P314L	6	Missense
	NSP13	1					
17999			C	T	T588I	4	Missense
	NSP15	1					
19803			C	T	R61R	4	Synonymous
	NSP16	10					
21110			C	T	T151I	4	Missense
21512			A	C	N285T	1	Missense
21517			AG	CA	R287Q	1	Missense
21520			G	C	V288L	1	Missense
21523			GTT	TCG	V289S	1	Missense
21528			T	C	I290I	1	Synonymous
21532			A	T	S292C	1	Missense
21537			T	A	D293E	1	Missense
21539			T	A	V294D	1	Missense
21547			AA	CT	N297L	1	Missense
	Spike	15					
21614			C	T	L18F	1	Missense
21801			A	C	D80A	6	Missense
22206			A	G	D215G	6	Missense
22286			CTTGCTTTA	.	L242	4	deletion
22481			A	C	T307P	1	Missense

22511			AA	TT	N317F	1	Missense
22535			T	C	S325P	1	Missense
22539			TT	AA	I326K	1	Missense
22542			T	C	V327A	1	Missense
22813			G	T	K417N	4	Missense
23012			G	A	E484K	6	Missense
23063			A	T	N501Y	6	Missense
23403			A	G	D614G	6	Missense
23664			C	T	A701V	6	Missense
24821			G	T	A1087S	1	Missense
	ORF3a	4					
25546			C	T	L52F	4	Missense
25563			G	T	Q57H	6	Missense
25904			C	T	S171L	6	Missense
26058			C	T	D222D	1	Synonymous
	Envelope	1					
26456			C	T	P71L	6	Missense
	Membrane	1					
27059			C	T	Y179Y	4	Synonymous
	3'UTR	1					
27193			T	C	27193	1	extragenic
	ORF7b	7					
27873			.	G	E39	1	Insertion
27874			C	A	T40N	1	Missense
27875			.	C	T40	1	Insertion
27879			C	.	C41	1	Deletion
27880			A	T	H42L	1	Missense
27882			G	.	H42	1	deletion

27883			C	T	A43V	1	Missense
	ORF8	3					
27921			A	T	I10F	1	Missense
28253			CA	TC	I121L	5	Missense
28253			C	T	F120F	1	Synonymous
	Nucleocapsid	5					
28310			C	T	P13S	1	Missense
28887			C	T	T205I	6	Missense
29095			C	T	F274F	1	Synonymous
29144			C	T	E290E	1	Synonymous
29192			T	C	F307L	2	Missense

Appendix 4: Delta Variant SNP Mutation table

Ref genome position	Region/ Gene	# 0f Mutations	reference nucleotide	sample Nucleotide	AA change	Frequency	Mutation Type
	5'UTR	2					
210			G	T	210	4	extragenic
241			C	T	241	4	extragenic
	NSP2	1					
926			T	C	F41L	2	Missense
	NSP3	8					
3037			C	T	F106F	4	Synonymous
4181			G	T	A488S	4	Missense
5953			G	A	L1078L	2	Synonymous
6402			C	T	P1228L	4	Missense
6568			C	T	D1283D	1	Synonymous
7124			C	T	P1469S	4	Missense
7923			C	T	S1735F	1	Missense
8035			G	T	M1772I	2	Missense
	NSP4	4					
8986			C	T	D144D	4	Synonymous
9053			G	T	V167L	4	Missense
9474			C	T	A307V	1	Missense
10029			C	T	T492I	4	Missense
	NSP5	1					
10691			A	G	I213V	4	Missense

	NSP6	3					
11201			A	G	T77A	4	Missense
11332			A	G	V120V	4	Synonymous
11653			C	T	L227L	2	Synonymous
	NSP8	1					
12452			C	T	P121S	2	Missense
	NSP12b	2					
14408			C	T	P314L	4	Missense
15451			G	A	G662S	4	Missense
	NSP13	1					
16466			C	T	P77L	4	Missense
	NSP14	1					
19220			C	T	A394V	4	Missense
	NSP15	1					
20132			C	T	A171V	2	Missense
	Spike	9					
21618			C	G	T19R	4	Missense
21846			C	T	T95I	4	Missense
21987			G	A	G142D	4	Missense
22029			AGTTCA	.	E156	4	deletion
22917			T	G	L452R	4	Missense
22995			C	A	T478K	4	Missense
23403			A	G	D614G	4	Missense
23604			C	G	P681R	4	Missense
24410			G	A	D950N	4	Missense
	ORF3a	2					

25469			C	T	S26L	4	Missense
25665			C	T	Y91Y	1	Synonymous
	Envelope	1					
26438			T	C	L65P	2	Missense
	Membrane	1					
26767			T	C	I82T	4	Missense
	ORF7a	2					
27638			T	C	V82A	4	Missense
27752			C	T	T120I	4	Missense
	ORF7b	1					
27874			C	T	T40I	4	Missense
	ORF8	1					
28248			GATTTC	.	D119	4	deletion
	3'UTR	1					
28273			A	.	28273	4	extragenic
	Nucleocapsid	6					
28461			A	G	D63G	4	Missense
28881			G	T	R203M	2	Missense
28881			GG	TT	R203I	2	Missense
28916			G	T	G215C	4	Missense
29402			G	T	D377Y	4	Missense
29433			A	G	K387R	2	Missense
	3'UTR	1					
29742			G	T	29742	4	extragenic

Appendix 5: Omicron Variant SNP Mutation table

Ref genome position	Region/ Gene	# 0f Mutations	reference nucleotide	sample Nucleotide	AA change	Frequency	Mutation Type
	5'UTR	2					
222			C	T	222	1	extragenic
241			C	T	241	15	extragenic
	NSP1	2					
478			C	T	I71I	1	Synonymous
670			T	G	S135R	4	Missense
	NSP3	22					
2790			C	T	T24I	4	Missense
2832			A	G	K38R	17	Missense
3037			C	T	F106F	23	Synonymous
4184			G	A	G489S	4	Missense
4321			C	T	A534A	4	Synonymous
4576			A	T	T619T	1	Synonymous
4579			T	A	L620L	2	Synonymous
4655			C	T	R646W	1	Missense
5161			T	A	F814L	1	Missense
5260			T	A	T847T	1	Synonymous
5386			T	G	A889A	19	Synonymous
5866			C	T	F1049F	1	Synonymous
6513			GTT	.	S1265	18	Deletion
6968			C	T	L1416L	1	Synonymous
7069			A	G	E1450E	1	Synonymous
7071			G	T	G1451V	1	Missense
7086			C	T	T1456I	2	Missense
7252			G	C	L1511F	1	Missense

7292			T	C	G1524G	1	Synonymous
7787			G	C	A1690P	1	Missense
8290			C	T	L1857L	1	Synonymous
8393			G	A	A1892T	19	Missense
	NSP4	6					
8723			A	C	I57L	1	Missense
9344			C	T	L264F	4	Missense
9424			A	G	V290V	2	Synonymous
9534			C	T	T327I	4	Missense
9866			C	T	L438F	4	Missense
10029			C	T	T492I	23	Missense
	NSP5	4					
10198			C	T	D48D	4	Synonymous
10447			G	A	R131R	4	Synonymous
10449			C	A	P132H	23	Missense
10892			A	G	T280A	1	Missense
	NSP6	6					
11083			G	T	L37F	1	Missense
11109			C	T	A46V	1	Missense
11286			TGTCTGGTT	.	L105	19	Deletion
11288			TCTGGTTTT	.	S106	4	Deletion
11537			A	G	I189V	19	Missense
11713			T	C	S247S	1	Synonymous
	NSP7	1					
11941			C	T	V33V	1	Synonymous
	NSP8	2					
12406			C	T	N105N	1	Synonymous

12557			A	C	I156L	1	Missense
	NSP9	2					
12786			C	T	T34I	1	Missense
12880			C	T	I65I	4	Synonymous
	NSP10	1					
13195			T	C	V57V	19	Synonymous
	NSP12b	6					
13756			A	G	I97V	1	Missense
14408			C	T	P314L	23	Missense
15240			C	T	N591N	19	Synonymous
15714			C	T	L749L	4	Synonymous
15939			T	C	D824D	1	Synonymous
15960			C	T	A831A	1	Synonymous
	NSP13	4					
16406			T	G	V57G	2	Missense
16636			G	T	A134S	1	Missense
17280			G	T	V348V	1	Synonymous
17410			C	T	R392C	4	Missense
	NSP14	1					
18163			A	G	I42V	23	Missense
	NSP15	5					
19884			C	T	Y88Y	1	Synonymous
19955			C	T	T112I	4	Missense
19999			G	T	V127F	1	Missense
20055			A	G	E145E	4	Synonymous
20094			A	G	K158K	1	Synonymous
	NSP16	1					
20741			A	G	Q28R	1	Missense
	Spike	58					

21618			C	T	T19I	4	Missense
21633			TACCCCCTG	.	L24	4	Deletion
21762			C	.	A67	19	Deletion
21764			A	.	A67	19	Deletion
21767			CATG	.	I68	19	Deletion
21846			C	T	T95I	19	Missense
21987			GTGTTTATT	.	G142	19	Deletion
21987			G	A	G142D	4	Missense
22139			G	C	V193L	1	Missense
22161			A	G	Y200C	1	Missense
22193			.	T	I210	15	Insertion
22194			ATT	.	N211	1	Deletion
22195			T	G	N211K	15	Missense
22197			TA	GC	L212C	15	Missense
22200			T	G	V213G	4	Missense
22201			.	AGC	S214	15	insertion
22202			.	A	V213	15	Insertion
22203			.	A	R214	15	Insertion
22204			T	A	R214R	15	Synonymous
22289			G	T	A243S	1	Missense
22293			T	C	L244S	1	Missense
22578			G	A	G339D	23	Missense
22599			G	A	R346K	5	Missense
22669			T	C	Y369Y	1	Synonymous
22673			TC	CT	S371L	19	Missense
22674			C	T	S371F	4	Missense
22679			T	C	S373P	23	Missense

22686			C	T	S375F	23	Missense
22688			A	G	T376A	4	Missense
22775			G	A	D405N	4	Missense
22786			A	C	R408S	4	Missense
22813			G	T	K417N	23	Missense
22882			T	G	N440K	7	Missense
22898			G	A	G446S	13	Missense
22969			AA	GG	T470A	1	Missense
22992			G	A	S477N	16	Missense
22995			C	A	T478K	17	Missense
23013			A	C	E484A	17	Missense
23040			A	G	Q493R	17	Missense
23048			G	A	G496S	13	Missense
23055			A	G	Q498R	17	Missense
23063			A	T	N501Y	17	Missense
23075			T	C	Y505H	16	Missense
23202			C	A	T547K	19	Missense
23403			A	G	D614G	23	Missense
23525			C	T	H655Y	23	Missense
23599			T	G	N679K	23	Missense
23604			C	A	P681H	23	Missense
23854			C	A	N764K	23	Missense
23948			G	T	D796Y	23	Missense
24130			C	A	N856K	19	Missense
24424			A	T	Q954H	23	Missense
24469			T	A	N969K	22	Missense
24503			C	T	L981F	18	Missense
24872			G	T	V1104L	1	Missense
24942			A	G	D1127G	1	Missense

25000			C	T	D1146D	23	Synonymous
25352			G	T	V1264L	2	Missense
	ORF3a	3					
25584			C	T	T64T	23	Synonymous
25587			C	T	L65L	4	Synonymous
26060			C	T	T223I	4	Missense
	Envelope	2					
26270			C	T	T9I	23	Missense
26392			A	T	S50C	1	Missense
	Membrane	5					
26530			A	G	D3G	16	Missense
26577			C	G	Q19E	23	Missense
26709			G	A	A63T	20	Missense
26858			C	T	F112F	3	Synonymous
26997			T	G	C159G	1	Missense
	ORF6	4					
27209			A	G	H3R	1	Missense
27259			A	C	M19M	23	Synonymous
27382			GAT	CTC	D61L	4	Missense
27384			T	C	D61D	1	Synonymous
	ORF7b	1					
27807			C	T	L17L	16	Synonymous
	ORF8	2					
27998			C	T	D35D	1	Synonymous
28079			G	C	V62V	1	Synonymous
	3'UTR	1					
28271			A	T	28271	23	extragenic

	Nucleocapsid	5					
28311			C	T	P13L	22	Missense
28362			GAGAACGCA	.	E31	22	Deletion
28881			GGG	AAC	RG203KR	23	Missense
29304			C	T	P344L	4	Missense
29510			A	C	S413R	4	Missense
	3'UTR	3					
29545			C	T	29545	1	extragenic
29734			GAGGCCACGCGGAGTACGATCGAGTG	.	29734	4	extragenic
29736			G	T	29736	1	extragenic

Appendix 6: Ethical approval

Plot No. 1, Cnr Joseph Mwilwa & Great East Road
Rhodes Park, Lusaka - Zambia
Tel: +260 955 155 633
+260 955 155 634
Cell: +260 977 493220
Email: eresconvergeltd@gmail.com

I.R.B. No. 00005948
F.W.A. No. 00011697

28th August, 2020

Ref. No. 2020-Aug-008

The Principal Investigator
Prof. Victor Mukonka,
Zambia National Public Health Institute,
13 Reedbuck Road, Kabulonga
LUSAKA.

Dear Prof. Mukonka,

RE: GENOMIC CHARACTERIZATION OF SARS-COV-2 AND EVALUATION OF NOVEL DIAGNOSTIC TESTS FOR COVID-19 IN ZAMBIA.

Reference is made to your protocol resubmission dated 25th August, 2020. The IRB resolved to approve this study and your participation as Principal Investigator for a period of one year.

Review Type	**Fast Track**	Approval No. **2020-Aug-008**
Approval and Expiry Date	Approval Date: 28th August, 2020	Expiry Date: 27th August, 2021
Protocol Version and Date	Version - Nil.	27th August, 2021
Information Sheet, Consent Forms and Dates	• English.	27th August, 2021
Consent form ID and Date	Version - Nil	27th August, 2021
Recruitment Materials	Nil	27th August, 2021
Other Study Documents	Data Collection Sheet.	27th August, 2021
Number of participants approved for study		27th August, 2021

Specific conditions will apply to this approval. As Principal Investigator it is your responsibility to ensure that the contents of this letter are adhered to. If these are not adhered to, the approval may be suspended. Should the study be suspended, study sponsors and other regulatory authorities will be informed.

Conditions of Approval

- No participant may be involved in any study procedure prior to the study approval or after the expiration date.
- All unanticipated or Serious Adverse Events (SAEs) must be reported to the IRB within 5 days.
- All protocol modifications must be IRB approved prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address.
- All protocol deviations must be reported to the IRB within 5 working days.
- All recruitment materials must be approved by the IRB prior to being used.
- Principal investigators are responsible for initiating Continuing Review proceedings. Documents must be received by the IRB at least 30 days before the expiry date. This is for the purpose of facilitating the review process. Any documents received less than 30 days before expiry will be labelled "late submissions" and will incur a penalty.
- Every 6 (six) months a progress report form supplied by ERES IRB must be filled in and submitted to us.
- A reprint of this letter shall be done at a fee.

Should you have any questions regarding anything indicated in this letter, please do not hesitate to get in touch with us at the above indicated address.

On behalf of ERES Converge IRB, we would like to wish you all the success as you carry out your study.

Yours faithfully,
ERES CONVERGE IRB

Dr. Jason Mwanza
Dip. Clin. Med. Sc., BA., M.Soc., PhD
CHAIRPERSON

Appendix 7: Published article associated with this thesis

viruses

MDPI

Article

Genomic Surveillance of SARS-CoV-2 in the Southern Province of Zambia: Detection and Characterization of Alpha, Beta, Delta, and Omicron Variants of Concern

Ben Katowa [1,2,†], Annie Kalonda [3,4,5,†], Benjamin Mubemba [6,7], Japhet Matoba [1], Doreen Mainza Shempela [8], Jay Sikalima [8], Boniface Kabungo [9], Katendi Changula [10], Simbarashe Chitanga [3,11,12], Mpanga Kasonde [13], Otridah Kapona [13], Nathan Kapata [13], Kunda Musonda [13], Mwaka Monze [14], John Tembo [15], Matthew Bates [15,16], Alimuddin Zumla [17], Catherine G. Sutcliffe [18,19], Masahiro Kajihara [20], Junya Yamagishi [21,22], Ayato Takada [4,5,20,22,23], Hirofumi Sawa [4,5,21,23,24,25,26], Roma Chilengi [13,27], Victor Mukonka [13], Walter Muleya [2] and Edgar Simulundu [1,4,*]

1 Macha Research Trust, Choma 20100, Zambia
2 Department of Biomedical Sciences, School of Veterinary Medicine, University of Zambia, Lusaka 10101, Zambia
3 Department of Biomedical Sciences, School of Health Sciences, University of Zambia, Lusaka 10101, Zambia
4 Department of Disease Control, School of Veterinary Medicine, University of Zambia, Lusaka 10101, Zambia
5 Africa Centre of Excellence for Infectious Diseases of Humans and Animals, School of Veterinary Medicine, University of Zambia, Lusaka 10101, Zambia
6 Department of Wildlife Sciences, School of Natural Resources, Copperbelt University, Kitwe 50100, Zambia
7 Department of Biomedical Sciences, School of Medicine, Copperbelt University, Ndola 50100, Zambia
8 Churches Health Association of Zambia, Lusaka 10101, Zambia
9 Southern Provincial Health Office, Ministry of Health, Choma 20100, Zambia
10 Department of Paraclinical Studies, School of Veterinary Medicine, University of Zambia, Lusaka 10101, Zambia
11 Department of Preclinical Studies, School of Veterinary Medicine, University of Namibia, Windhoek Private Bag 13301, Namibia
12 School of Life Sciences, College of Agriculture, Engineering and Sciences, University of KwaZulu-Natal, Private Bag X54001, Durban 4000, South Africa
13 Zambia National Public Health Institute, Ministry of Health, Lusaka 10101, Zambia
14 Virology Laboratory, University Teaching Hospital, Lusaka 10101, Zambia
15 HerpeZ Infection Research and Training, University Teaching Hospital, Lusaka 10101, Zambia
16 School of Life and Environmental Sciences, University of Lincoln, Lincoln, Lincolnshire LN6 7TS, UK
17 Division of Infection and Immunity, Centre for Clinical Microbiology, University College London, NIHR Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London NW3 2PF, UK
18 Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
19 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
20 Division of Global Epidemiology, International Institute for Zoonosis Control, Hokkaido University, N20 W10, Kita-ku, Sapporo 001-0020, Japan
21 Division of Collaboration and Education, International Institute for Zoonosis Control, Hokkaido University, N20 W10, Kita-ku, Sapporo 001-0020, Japan
22 International Collaboration Unit, International Institute for Zoonosis Control, Hokkaido University, N20 W10, Kita-ku, Sapporo 001-0020, Japan
23 One Health Research Center, Hokkaido University, N18 W9, Kita-ku, Sapporo 001-0020, Japan
24 Division of Molecular Pathobiology, International Institute for Zoonosis Control, Hokkaido University, N20 W10, Kita-ku, Sapporo 001-0020, Japan
25 Division of International Research Promotion, Hokkaido University International Institute for Zoonosis Control, N20 W10, Kita-ku, Sapporo 001-0020, Japan
26 Global Virus Network, 725 W Lombard Street, Baltimore, MD 21201, USA
27 Republic of Zambia State House, Lusaka 10101, Zambia
* Correspondence: esikabala@yahoo.com
† These authors contributed equally to this work.

check for updates

Citation: 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. https://doi.org/10.3390/v14091865

Academic Editors: Marta Giovanetti and Luiz Carlos Junior Alcantara

Received: 12 July 2022
Accepted: 19 August 2022
Published: 24 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern (VOCs) have significantly impacted the global epidemiology of the pandemic. From December 2020 to April 2022, we conducted genomic surveillance of SARS-CoV-2 in the Southern Province of Zambia, a region that shares international borders with Botswana, Namibia, and Zimbabwe and is a major tourist destination. Genetic analysis of 40 SARS-CoV-2 whole genomes revealed the circulation of Alpha (B.1.1.7), Beta (B.1.351), Delta (AY.116), and multiple Omicron subvariants with the BA.1 subvariant being predominant. Whereas Beta, Delta, and Omicron variants were associated with the second, third, and fourth pandemic waves, respectively, the Alpha variant was not associated with any wave in the country. Phylogenetic analysis showed evidence of local transmission and possible multiple introductions of SARS-CoV-2 VOCs in Zambia from different European and African countries. Across the 40 genomes analysed, a total of 292 mutations were observed, including 182 missense mutations, 66 synonymous mutations, 23 deletions, 9 insertions, 1 stop codon, and 11 mutations in the non-coding region. This study stresses the need for the continued monitoring of SARS-CoV-2 circulation in Zambia, particularly in strategically positioned regions such as the Southern Province which could be at increased risk of introduction of novel VOCs.

Keywords: SARS-CoV-2; COVID-19; variants of concern; spike mutations; whole-genome sequencing; Zambia

1. Introduction

As of 6 July 2022, the ongoing coronavirus disease 2019 (COVID-19) pandemic has caused over 548,990,094 confirmed cases including 6,341,637 deaths [1]. In Africa, despite having a total population of about 1.3 billion, the official reports show a low burden of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections when compared with other continents. The total number of confirmed cases and fatalities reported in Africa were 9,138,803 and 173,674, respectively, representing a global burden of 1.7% [1]. However, post-mortem and serological studies in some African countries suggest that the true burden of SARS-CoV-2 infections and deaths may be higher than what is officially reported [2–5]. Further, a recent systematic review by the World Health Organisation (WHO) on the seroprevalence of SARS-CoV-2 in Africa revealed that over two-thirds of the African population had been infected by SARS-CoV-2 [6]. The analysis further revealed that the true number of SARS-CoV-2 infections on the African continent was 97 times higher than the reported confirmed cases and the sharp rise in incidence was attributed to the introduction of the highly transmissible Alpha and Delta variants [6,7].

The first COVID-19 case in Africa was reported in Egypt on 14 February 2020 [8,9] followed by Algeria, with its first case being reported on 25 February 2020 [10] and Nigeria on 27 February 2020 [11]. Most African countries including Cameroon, Morocco, Senegal, South Africa, Togo, and Tunisia reported their first cases by mid-March 2020 [8,12] and most of the index cases were imported cases from Europe which by then had become the epicentre of the pandemic [8,12]. Within three months of Africa's COVID-19 index case, 54 of 55 African Union (AU) Member States (except Western Sahara) had reported over 100,000 cases which included imported and community transmissions [8]. The early phase of the pandemic in Africa was characterized by the predominance of lineage B.1 which was introduced multiple times in African countries [13]. However, due to a ban on international air travel in most African countries and the world at large in March/April 2020, the number of SARS-CoV-2 importations into Africa decreased and the pandemic entered a phase that was characterized by sustained low levels of within-country spread and occasional international viral dissemination between neighbouring countries, presumably via road and rail links between these countries [13].

As the pandemic progressed, several SARS-CoV-2 variants carrying mutations with concerning phenotypic implications on current pandemic management strategies emerged [14]. Of particular significance to the ongoing pandemic are SARS-CoV-2 variants designated

variants of concern (VOCs). Several VOCs have been described including Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529). VOCs are associated with enhanced transmissibility or virulence, reduction in neutralization by antibodies obtained through natural infection or vaccination, the ability to evade detection, or a decrease in therapeutic or vaccination effectiveness [15,16]. Further, all the five reported VOCs have mutations in the receptor-binding domain (RBD) and the N-terminal domain (NTD), of which N501Y mutation located on the RBD is common to all variants except the Delta variant [16]. The N501Y mutation results in increased binding affinity of the spike (S) protein to angiotensin-converting enzyme (ACE) 2 receptors thereby enhancing the viral attachment and its subsequent entry into the host cells [17,18]. Other genomic changes have been reported, including the extensive deletion in the open reading frame (ORF) 7a, ORF8 [19–21], and a deletion in the nsp2 genes [22], but these deletions have been associated with mild to moderate clinical symptoms compared to the infection caused by the wildtype SARS-CoV-2 [21,23].

To date, four VOCs, namely Alpha, Beta, Delta, and Omicron have been detected on the African continent. The first VOC, designated Alpha (B.1.1.7), was detected in September 2020 in the United Kingdom (UK) and was introduced into Africa between November 2020 and February 2021 with evidence of local transmission in Nigeria and Ghana [13]. This variant is characterised by nine mutations in the S protein, increased transmissibility, and increased risk of hospitalisation [24,25]. The second VOC was the Beta (B.1.351 lineage) variant which was first detected in South Africa in October 2020 and became the most common variant in many African countries [26]. This VOC is characterized by mutations in the S protein, including in the RBD—K417N, E484K, and N501Y [14,26]. In addition, the Beta variant is known to cause severe disease in young and healthy individuals [26]. Whereas the Beta variant was associated with the second wave of SARS-CoV-2 in Africa, the Alpha variant did not predominate in many African countries possibly due to a lack of selective advantage over the other VOCs [27]. These variants were replaced by the highly transmissible Delta (B.1.617.2 lineage) variant which was initially detected in India in December 2020 and spread worldwide among vaccinated as well as unvaccinated individuals [28]. This variant seeded the third wave of the pandemic in 2021 and was introduced in Africa in June 2021. The Omicron variant, characterised by several mutations in the S protein, including a set of mutations previously observed in other VOCs and novel mutations, was first reported in South Africa on 24 November 2021 and became the dominant driver of the fourth global wave of SARS-CoV-2 [29].

In Zambia, the first known COVID-19 cases were reported on 18 March 2020 from travellers returning from Europe [30]. Within days, the government implemented restrictions on international travel, school closures, halting of non-essential business, and confinement of people to their homes. Despite these measures, the virus spread to all parts of the country with over 300,000 cases and over 4000 deaths as of 6 July 2022 [31]. The course of the pandemic in Zambia can be divided into four major waves: the first wave occurred from July to September 2020 and was mainly driven by B.1.1 and its sub-lineages; the second wave occurred from December 2020 to April 2021 and was dominated by the Beta variant, while the Delta variant dominated the third wave from May to September 2021 [32,33]. The Omicron variant has dominated the fourth pandemic wave in Zambia, with cases peaking in early January 2022 and then rapidly decreasing to low levels. In the Southern Province, which shares international borders with Botswana, Namibia, and Zimbabwe and is a major tourist destination, SARS-CoV-2 was first detected in May 2020 [31]. As the course of the pandemic continues to evolve, it remains crucial to monitor and understand the virus evolution and outbreak dynamics, particularly in strategically positioned regions such as the Southern Province which is a trade entry point of Zambia for all imports and exports from Southern Africa. However, there is limited data regarding the molecular epidemiology of SARS-CoV-2 in Zambia, with only two genomic studies reporting the detection of SARS-CoV-2 belonging to lineage B.1.1. [34] and the B.1.351 variant [35]. Moreover, to our best knowledge, no reports have described the genetic characteristics of SARS-CoV-2 VOCs

circulating in the Southern Province. Therefore, this study used whole-genome sequencing (WGS) and phylogenetic analyses to describe the genetic characteristics of SARS-CoV-2 in the Southern Province of Zambia.

2. Materials and Methods

2.1. Study Site and Sample Collection

The study samples were collected between December 2020 and April 2022 from eight districts in the Southern Province (Figure 1). Sample collection was conducted through the Zambia National Public Health Institute under the coordination of the Zambia Genomic Sequencing Consortium. The samples were collected through routine surveillance (i.e., point of entry screening and routine screening for influenza-like illnesses) and targeted surveillance of cluster outbreaks. A total of 198 samples were collected from different parts of the Southern Province and were brought to Macha Research Trust (MRT) for WGS. WGS was conducted in collaboration with the Churches Health Association of Zambia (CHAZ) with 161 samples collected between December 2020 and November 2021 being transported to MRT, while 37 samples collected from December 2021 to April 2022 were transported to CHAZ Complex laboratory for sequencing. Upon receipt, all samples were retested to determine the cycle threshold (Ct) value of each sample. Samples that had a Ct value of $\leq$30 and were submitted with the relevant metadata were included to undergo WGS. Samples that did not meet the inclusion criteria and those that could not be amplified or had poor genomic coverage were excluded from further analysis.

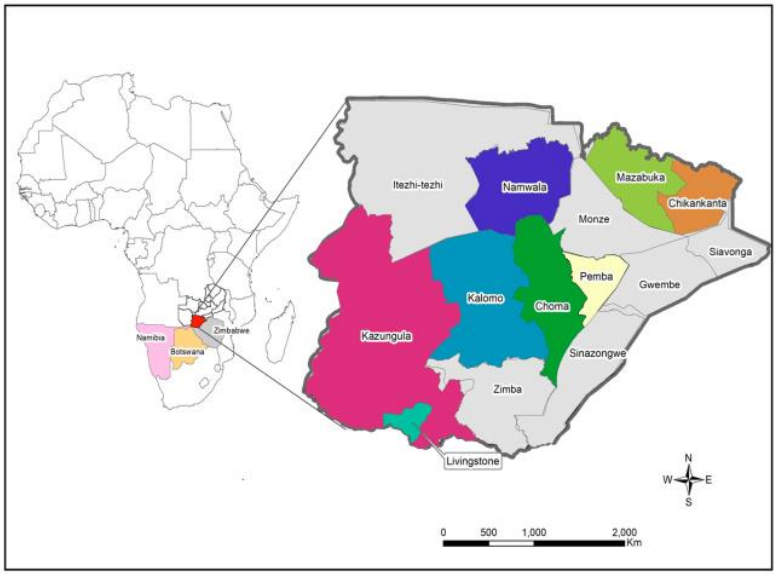

Figure 1. Map showing the location of study sites in Southern Province. The locator map depicts Zambia with neighbouring countries that share the border with Southern Province. The insert map shows the Southern Province of Zambia with the study sites namely Chikankata, Choma, Kalomo, Kazungula, Livingstone, Mazabuka, Namwala and Pemba districts. The maps were generated using Quantum Geographic Information System (QGIS) version 3.10 (http://www.qgis.org (accessed on 8 August 2022).

2.2. RNA Extraction and Virus Genome Amplification

Viral RNA was extracted from nasopharyngeal swabs using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) or the MagMax kit (Thermo Fisher Scientific, Waltham,

MA, USA) on an automated KingFisher Flex platform (Thermo Fisher Scientific, USA) according to manufacturer specifications and protocols. Amplification of the SARS-CoV-2 genome in preparation for WGS was conducted using the Centre for Disease Control and Prevention (CDC) SARS-CoV-2 qRT-PCR assay [36].

2.3. Next-Generation Sequencing

Whole-genome sequencing was performed using the Oxford Nanopore technologies and Illumina NextSeq platforms. For Oxford Nanopore, a cDNA synthesis reaction was performed on 36 samples (based on cycle threshold values < 30) using SuperScript IV Reverse Transcriptase kit (Invitrogen, Waltham, MA, USA), following the manufacturer's instructions. Library preparation was conducted using the ARTIC protocol version 3 [37,38]. Whole-genome sequencing was conducted using custom-designed primers (Tables S3 and S4) [34]. The PCR products were cleaned using AMPure XP beads (Beckman Coulter, Brea, CA, USA) and DNA quantification was conducted using a Qubit fluorometer (Thermo Fisher Scientific). End-repair on the amplified samples was conducted using NEBNext Ultra II End Repair Module (New England BioLabs, Ipswich, MA, USA). Native barcode expansion kits 1–12 and 13–24 was used in combination with Ligation Sequencing Kit (SQK-LSK109) (Oxford Nanopore Technologies). Subsequently, genomic sequencing was conducted using the MinION 1MkB (Oxford Nanopore Technologies, Oxford, UK). The RAMPART (v1.0.6) software package was used to monitor sequencing performance in real-time, with runs proceeding until a minimum of approximately 200-fold coverage was achieved across all amplicons. At this point, the run was terminated and the resulting reads were basecalled using Guppy (4.0.14). Consensus sequence generation was conducted using the ARTIC bioinformatics pipeline (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html (accessed on 7 October 2021).

For Illumina NextSeq, V.3 primers pools designed by ARTIC Network were used (https://github.com/joshquick/artic-ncov2019/blob/master/primer_schemes/nCoV-2019/V3/nCoV-2019.tsv (accessed on 7 October 2021)). Sequencing libraries for 37 samples were prepared using the Illumina COVIDSeq kit on the automated Hamilton robotic instrument, ABI 7500 fast, and the Quant Studio thermo-cyclers. After successful library clean-up and pooling, pooled samples were quantified and normalized using a Qubit dsDNA HS Assay kit by diluting from a starting concentration of 4 nM to a final loading concentration of 1 nM. Thereafter, 25 µL was loaded on the Illumina NextSeq 2000 instrument through a cartridge loaded with a flow cell for SARS-CoV-2 genomic sequencing. A customized version of the DRAGEN DNA pipeline was used to perform Kmer-based detection of SARS-CoV-2. The Nextseq 2000 then aligned the reads to a reference genome, calls variants, and generates consensus genome sequences. The NextSeq 2000 optionally performs lineage/clade analysis using Pangolin and NextClade.

2.4. Genome Annotation and Phylogenetic Analysis

Whole-genome sequences were annotated using the reference genome of hCoV-19/Wuhan/Hu-1/2019 | EPI_ISL_402125 [33]. A dataset of 180 whole genomes was created, which included 40 generated from this study and 140 retrieved from the GISAID database. Audacity *Instant* was used to retrieve SARS-CoV-2 whole-genome sequences from GISAID that were most similar to the sequences generated in this study. We also included reference sequences of VOCs detected in Southern Africa and other parts of the world, targeting those isolated within the same period and belonging to the same lineage as those characterized in this study. Reference sequences with stretches of more than 10% 'NNNN' were excluded from the analysis. Multiple sequence alignment of the sequences was performed using the FFT-NS-2 algorithm available in the multiple sequence alignment programme (MAFFT), but otherwise using default settings (https://mafft.cbrc.jp/alignment/server/index.html (accessed on 9 August 2022) [39]. The alignment was inspected in Geneious Prime v2022.0.1 (https://www.geneious.com (accessed on 9 August 2022) and gaps were trimmed. Follow-

ing alignment, a maximum likelihood (ML) phylogenetic tree was constructed using the PhyML Online server (www.atgc-montpellier.fr/phyml/ (accessed on 9 August 2022) [40] using the smart model selection (SMS) [41] and the Bayesian Information Criterion. Branch support was estimated through the SH-like approximate likelihood ratio test (SH-aLRT). The ML tree was then rooted using TempEst v1.5.3 [42], which estimated the best-fitting root of this phylogeny using the heuristic residual mean squared function, aimed at minimizing the variance of root-to-tip distances. The resultant ML tree file was edited using Interactive Tree of Life (iTOL) v5, an online tool for phylogenetic tree display and annotation [43].

PANGO lineage identification was performed using Pangolin v3.1.16 (https://pangolin.cog-uk.io/ (accessed on 8 August 2022)). Identification of single nucleotide polymorphisms (SNPs) was performed using the coronapp web application http://giorgilab.unibo.it/coronannotator/ (accessed on 8 August 2022). SNPs were identified based on the number of high confidence base calls (consensus sequence variations of the assembly) that do not agree with the reference bases for the genome position of interest. These variations were then exported to a vcf file and visualized in Microsoft Excel. The GISAID accession IDs of the genomes generated in this study can be found in Table S1.

3. Results

3.1. Characteristics of Patients with COVID-19 from the Southern Province of Zambia

A total of 198 samples were received for WGS from districts in the Southern Province, 74 were negative for SARS-CoV-2, 51 had Ct values > 30, and 33 had a low genome coverage. Only 40 samples were successfully sequenced, 13 samples at MRT and 27 at CHAZ Complex laboratory.

Demographic data were analyzed for all the 198 samples and the majority of the samples (104/198; 52.5%) were from females as shown in Table 1. The mean age of the participants was 28 (range: 0–82). The data set for gender and age were not available for one and five samples, respectively (Table 1).

Table 1. Characteristics of the genotyped samples infected with SARS-CoV-2.

Parameters	Sample Distribution *n* (%), Overall, *n* = 198
Age Group	
0–14 Years	7 (3.5)
15–50 Years	171 (86.4)
>50 Years	15 (7.6)
Unknown	5 (2.5)
Gender	
Female	104 (52.5)
Male	93 (47.0)
Unknown	1 (0.5)

3.2. SARS-CoV-2 Lineage Assignment and Distribution in Southern Province

SARS-CoV-2 lineage assignment using the PANGOLIN application (https://pangolin.cog-uk.io/ (accessed on 8 August 2022), showed that the 40 genomes detected in this study were distributed into seven lineages, namely AY.116 (Delta), B.1.1.7 (Alpha), B.1.351 (Beta), and Omicron (BA.1, BA.1.1, BA.1.14, and BA.2) (Figure 2A). The largest number of the sequences (*n* = 17, 42.5%) belonged to lineage BA.1/GRA (Figure 2A). All lineage AY.116 sequences came from Choma District, whereas the six B.1.351 lineage was detected in Choma (*n* = 2), Namwala (*n* = 2), Kalomo (*n* = 1), and Mazabuka (*n* = 1) districts (Figure 2B; Table S1). The Alpha variant (B.1.1.7) viruses were found in Namwala, Pemba, and Chikankata districts (Figure 2B; Table S1). Of the 27 Omicron variants, 11 (40.7%) were from Livingstone, 9 (33.3%) from Chikankata, 4 (14.8%) from Kazungula, 2 (7.4%) from Choma and 1 (3.7%) from Namwala. Most lineage BA.1 viruses were detected in Chikankata and Livingstone districts where 7/27 (25.9%) viruses of this lineage were found in each district. Lineage BA.1.1 was detected in Chikankata and Kazungula districts

whereas B.1.14 was only detected in Livingstone (Table S1). Three of the five BA.2 lineage viruses were detected in Livingstone whereas the other two were detected in Choma and Kazungula districts as shown in Figure 2B and Table S1.

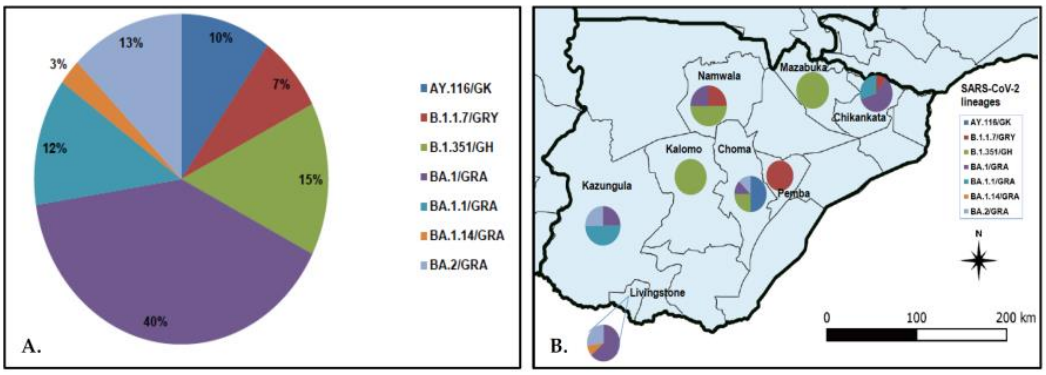

Figure 2. Distribution of SARS-CoV-2 lineage in the Southern Province of Zambia. Panel (**A**) pie chart showing SARS-CoV-2 lineages detected in the Southern Province; panel (**B**) proportionate distribution of SARS-CoV-2 lineages in the eight districts of the Southern Province. The map was generated using Quantum Geographic Information System (QGIS) version 3.10 (http://www.qgis.org (accessed on 8 August 2022)).

3.3. Phylogenetic Analysis

Phylogenetic analysis revealed that the sequences separated into four clades namely Delta, Beta, Alpha, and Omicron (Figure 3). In the Delta clade four Southern Province sequences (Zambia/SP250/2021 | EPI ISL 6761088, Zambia/SP253/2021 | EPI ISL 6762977, Zambia/SP251/2021 | EPI ISL 6761106, and Zambia/SP252/2021 | EPI ISL 6761100), separated into two groups of which two formed a distinct cluster with a Zambian isolate whereas the other two clustered with sequences from Angola, Eswatini, and Zambia (Figure 3). Six sequences analysed in this study belonged to the Beta clade and they separated into four distinct clusters. Two of the sequences (Zambia/SP30/2021 | EPI ISL 6760973 and Zambia/SP87/2021 | EPI ISL 6764745) analysed in this study clustered with isolates from Zambia, Zimbabwe, England, and the Democratic Republic of Congo (DRC), and another set of two formed a distinct cluster with sequences from Zambia. The last two sequences (Zambia/SP11/2021 | EPI_ISL_6760905 and Zambia/SP10/2021 | EPI_ISL_6760707) belonged to separate clusters with the former sequence grouping with Zambian sequences whereas the latter was closely related to sequences obtained in Malawi, Eswatini, and Botswana (Figure 3). In the Alpha clade, three Southern Province sequences, namely Zambia/SP32/2021 | EPI_ISL_6761015, Zambia/SP37/2020 | EPI_ISL_6761027, and Zambia/SP172/2021 | EPI_ISL_6761052 formed a distinguishable cluster with sequences from England and Zambia (Figure 3). The Omicron clade was separated into two clusters (Figure 3). The majority (22/27; 81.5%) of the Zambian sequences in this clade belonged to the BA.1 sub-lineage cluster whereas the rest (5/27; 18.5%) were of the BA.2 lineage. Phylogenetic analysis further showed that the Omicron sequences from this study were mainly closely related to sequences from European and African countries (Figure 3).

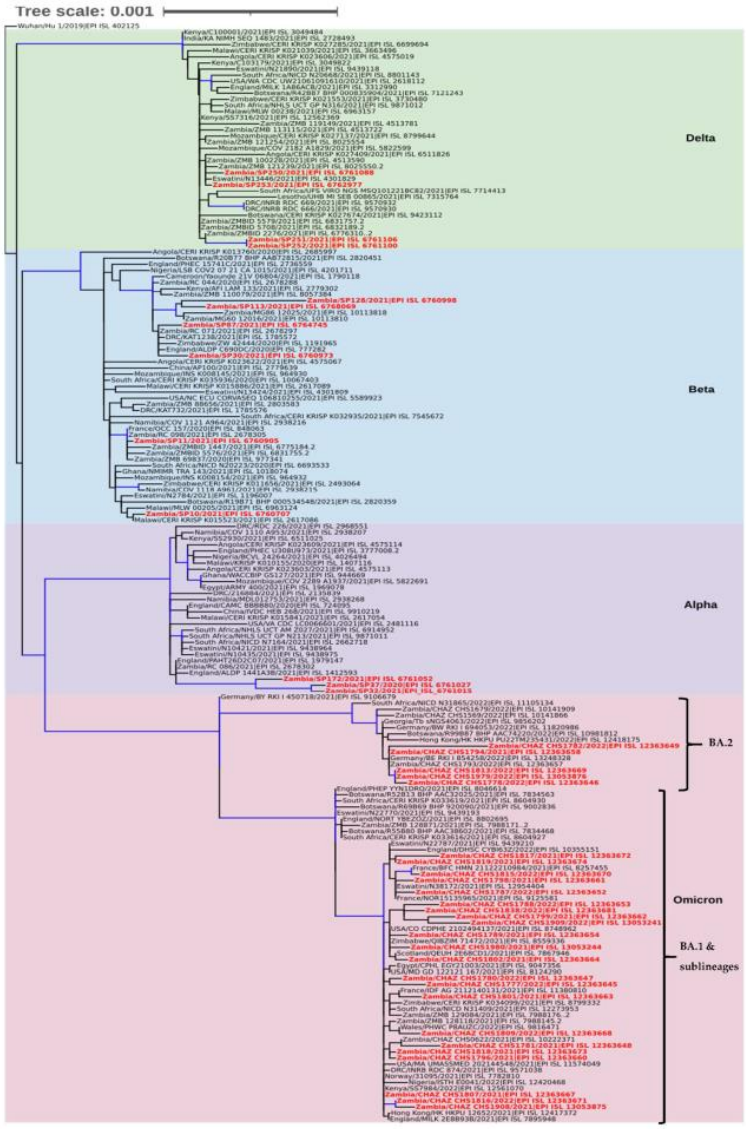

Figure 3. Phylogenetic analysis of SARS-CoV-2 genomes from Zambia and other countries. The genomes generated in this study are indicated in red whereas the shaded areas indicate the clades of variants of concern. Each sequence was named with the country name first followed by the isolate name and then the GISAID accession number. The tree branches highlighted in blue indicate tree branches that had a strong maximum likelihood ratio greater than 0.9, whereas the tree scale represents the nucleotide substitutions per site.

3.4. Molecular Analysis

A total of 292 different mutations were detected from the 40 genomes studied when compared to the Wuhan/Hu 1/2019 | EPI ISL 402125 reference sequence (Table 2). Most (96.2%) mutations were detected in the coding regions of the genomes. Of the mutations detected in the coding region, 64.8% (182/281) were missense mutations, 23.5% (66/281) were synonymous mutations, 8.2% (23/281) were deletions, 3.2% (9/281) insertions, and one was a stop codon (0.4%), gained with a single nucleotide polymorphism (SNP) on the ORF8 (Tables 2 and S2). Deletions and insertion included in-frame and out-of-frame mutations. When gene mutations were stratified according to the VOCs, the Alpha variant had a total of 53 different mutations of which 31 (58.5%) were missense mutations and 8 (15.1%) synonymous mutations. The number of mutations in the Alpha variant genomes ranged between 41 and 45 with (EPI_ISL_6761027) having the most mutations. The Beta variant had a total of 68 different mutations with 44 (64.7%) missense mutations and 15 (22.1%) synonymous mutations. The mutations in the Beta variant genomes ranged between 26 and 45 with one sequence (EPI_ISL_6760998) having the most mutations. Further, sequences of the Delta variant had a total of 50 mutations with 37 (74%) missense mutations and 7 (14%) synonymous mutations. The Delta variant mutations ranged between 39 and 44 with two sequences (EPI_ISL_6761106; EPI_ISL_6761100) having the most mutations. Sequences of the Omicron variant had the highest number of mutations; 149 different mutations with 90 (60.4%) missense and 39 (26.2%) synonymous mutations, with the genomes having a mutation range between 48 and 67 with three sequences (EPI_ISL_12363648; EPI_ISL_12363649; EPI_ISL_12363661) having the most mutations. Deletions, insertions, stop-codons, and upstream/downstream gene variants had a frequency below 18% in all the VOCs.

Table 2. Distribution of mutations along different genomic regions of SARS-CoV-2 sequences detected in Southern Province.

Genome Segment	Missense Mutation	Synonymous Mutation	Deletion	Insertion	Others	Total Mutation
Coding Region						
ORF1ab	74	48	9	3	0	134
Spike	65	3	10	4	0	82
ORF3a	5	4	0	0	0	9
Envelope	5	0	0	0	0	5
Membrane	5	2	0	0	0	7
ORF6	2	2	0	0	0	4
ORF7a	2	0	0	0	0	2
ORF7b	4	1	2	2	0	9
ORF8	4	3	1	0	1 [1]	9
Nucleocapsid	16	3	1	0	0	20
Non-coding Region [2]						
5′UTR	0	0	0	0	4	4
3′UTR	0	0	0	0	7	7
Total	182	66	23	9	12	292

[1] Stop codon in the ORF8; [2] all the mutations in the non-coding region are extragenic.

When the number of mutations per gene was counted only once, the S protein was the most mutated gene with 82 mutations whereas the second mutated gene was the NSP3 protein with 42 mutations (Table S2). Of the 82 mutations in the S protein, 65/82 were missense mutations, 3/82 synonymous mutations, 10/82 deletions, and 4 insertions as shown in Table 2. Among all the SNPs, the most common change was C > T followed by A > G and G > A. Further, a large deletion of 26 nucleotides was observed on position 29734 of the 3′UTR of the four sequences (EPI_ISL_12363646, EPI_ISL_12363658, EPI_ISL_12363649 and EPI_ISL_12363669).

The most common mutation was the D614G substitution on the S protein and P314L substitution on the NSP12b (RdRp) protein which occurred in all the sequences studied and

67.5% (27/40) showed other amino acid substitution in the S protein including T95I, G339D, S373P, S375F, H655Y, N679K, N764K, D796K, Q954H, and D1146D (Table S2). The second most common amino acid change (39/40; 97.5%) was the F106F substitution on the NSP3 followed by the K417N (31/40; 77.5%) substitution on the S protein T492I substitution on the NSP4, followed by P681H (30/40; 75%), and (29/40; 72.5%) N501Y substitutions on the S protein. In addition to these mutations, several substitutions, deletions, and insertions in other genomic areas were also present (Table S2).

Comparison of mutations on the S protein of the SARS-CoV-2 variants in this study with the wildtype (Wuhan-Hu-1) SARS-CoV-2 revealed that the Omicron variant had the highest number of mutations in this protein compared to the other VOCs in this study. The Omicron variant had 58 amino acid (AA) mutations which included six deletions and four insertions (Table 3). Of the 60 AA mutations in the Omicron variant, 22 were found to be in the RBD of the S protein including G339D, R346K, Y369Y, S371L, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446S, T470A, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, and Y505H (Tables 2 and 3). The other AA variations in the RBD included N501Y in Alpha variants, S325P, I326K, V327A, K417N, E484K, and N501Y in the Beta variant and L452R and T478K (Table 3).

Table 3. Spike protein mutations in different SARS-CoV-2 variants compared to the wild-type (Wuhan-Hu-1).

SARS-CoV-2 Variants	Spike Mutations [1]
Wuhan-Hu-1 (wild-type)	-
Alpha (B.1.1.7)	ΔH69, ΔY145, **N501Y**, A570D, D614G, P681H, T716I, T874I, S982A, D1118H
Beta (B.1.351)	L18F, D80A, D215G, ΔL242, T307P, N317F, **S325P, I326K, V327A, K417N, E484K, N501Y**, D614G, A701V, A1087S
Delta (AY.116)	T19R, T95I, G142D, ΔE156, **L452R, T478K,** D614G, P681R, D950N
Omicron (BA.1, BA.1.1, BA.1.14, BA.2)	T19I, ΔL24, ΔA67, ΔA67, ΔI68, T95I, ΔG142, G142D, V193L, Y200C, insI210, ΔN211, N211K, L212C, V213G, insS214, insV213, insR214, insV213, insR214, R214R, A243S, L244S, **G339D, R346K, Y369Y, S371L, S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446S, T470A, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H,** T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F, V1104L, D1127G, D1146D, V1264L

[1] Receptor-binding domain (residues 319–541) is marked as bold in all the variants. Δ Represents deletion, ins represent insertion.

4. Discussion

In this study, from the 198 samples that were obtained for genomic sequencing in eight districts of the Southern Province of Zambia, 40 SARS-CoV-2 whole genomes were successfully sequenced and analysed. Our dataset revealed that there were more cases of COVID-19 observed in females compared to males. However, other studies have recorded a higher disease burden in males compared to females [44–46]. The mean age of patients was 28 with a minimum and maximum age of 0 and 82 years, respectively. However, it cannot be ruled out that the small number of samples analysed in this study may have had an impact on the observed gender distribution and the mean age of COVID-19 patients. Furthermore, lineage assignment revealed that BA.1 was the most prevalent lineage among our sequences followed by B.1.351. This could be explained by the fact that most of the successfully sequenced samples were collected during the Omicron wave. The B.1.351 predominated in the second wave, AY.116 in the third wave, and BA.1 in the fourth wave. The findings corroborate those of other authors who reported the predominance of Beta (B.1.351), Delta (B.1.617.2), and Omicron BA.1 variants in the second, third, and fourth waves of the pandemic in Africa, respectively [13,26,47,48]. Moreover, the detection of AY.116 and B.1.351 coincided with a rapid increase in the number of confirmed cases and deaths in Zambia [35,49]. Despite the small sample size of this study, SARS-CoV-2 lineages were detected in different districts of the Southern Province. The majority of the Omicron

variants were detected in Chikankata and Livingstone districts, with the latter having more subvariants. It is plausible that Livingstone, being a border town, a tourist capital, and a major transportation link to Zambia's neighbouring countries, the area could be at increased risk of the introduction of novel VOCs. Except for the Alpha variant, all the other VOCs detected in this study were found in Choma District. This could be explained by the fact that Macha Research Trust where sequencing was conducted is located in the Choma District and thus the institution was more likely to receive samples throughout the different phases of the COVID-19 waves.

Phylogenetic analysis revealed that the 40 SARS-CoV-2 genomes generated in this study belonged to four SARS-CoV-2 VOCs namely Alpha, Beta, Delta, and Omicron variants. These VOCs have presented a formidable public health challenge during the COVID-19 pandemic because of their increased viral transmissibility and disease severity [50]. Additionally, the early detection of some of the VOCs in Africa highlights the importance of coordinated molecular surveillance systems in all parts of the world and the role Africa has played in enabling the early detection and characterization of new lineages and informing the global pandemic response. The close phylogenetic relatedness of sequences generated in this study with those from European and African countries supports the idea of possible multiple introductions of the virus from different regions. Phylogenetic analysis further revealed that some sequences from this study clustered together and among other Zambian sequences which may signify the local circulation of these viruses. Notably, sequences obtained in this study that grouped within the Alpha variant clade were phylogenetically distinguishable and were detected in three different districts, which may suggest independent introductions, particularly from Europe, as these sequences were closely related to isolates from England. This introduction could be attributed to the relaxation of flight restrictions at the time these samples were collected. The Zambian Alpha cluster also displayed a longer branch length compared to the other sequences in this clade indicating the continued evolution as the virus circulated. Interestingly, the Alpha variant has not been associated with any COVID-19 wave in Zambia. This observation may suggest that the Alpha variant has no selective advantage over the other VOCs such as the Beta and Delta variants [27]. Although some Beta and Delta variants were closely related to isolates from Europe and Zambia, others showed a close relationship to isolates from Eswatini, DRC, Malawi, and Zimbabwe, suggesting that public health measures implemented by the authorities may have been compromised by porous borders and thus permitting the variants to spread within the region. Phylogenetic analysis also revealed that Omicron variants separated into two major clusters, BA.1 and BA.2, signifying the continued evolution of this VOC. The BA.4 and BA.5 subvariants which have been associated with driving current waves of infection in South Africa [51,52] were not detected in this study.

The S glycoprotein of SARS-CoV-2 plays a pivotal role in viral infection and pathogenesis because of its role in host cell receptor recognition, viral attachment, and entry [53–57]. The present study demonstrated the presence of the D614G mutation in the S protein in all 40 genomes. Similar findings have been reported in many countries including Turkey [58], Oman [59], Egypt [60], and the Comoros Island [61]. In addition to the D614G mutation in the S glycoprotein (23403A > G), a P314L mutation (14408C > T) in the NSP12/RdRp was detected in all the sequences analysed. This finding agrees with previous research which reported a high co-occurrence of these mutations around the globe [62–64]. The D614G mutation is associated with a high viral load, infectivity, and transmissibility [50] whereas mutations in the RdRp protein results in a dysfunctional enzyme that generates errors during RNA synthesis, increasing the chances of mutations occurring [62,64,65]. It is also suggested that the co-occurrence of the D614G and NSP12_P314L mutations may enhance viral entry and replication, respectively [66]. Therefore, the S protein mutations and their effects on virulence should be closely monitored and evaluated, as this protein is the main target for vaccine development [67].

Alpha, Beta, and Omicron variants share the N501Y mutation, located in the receptor-binding domain (RBD) of the S protein. It is known to confer an increased binding affinity of the RBD for the ACE2 receptor, raising the viral transmission rate [68]. This mutation was detected in all Alpha, Beta, and 20 Omicron variants of our sequences. Furthermore, the K417N and E484K mutations in the S protein, common to all Beta variants [26] were also detected in our sequences. Other mutations in this present study included the Q27 stop in the ORF8 in all three Alpha variants. This mutation has been observed in the Alpha (B.1.1.7) variant and is known to truncate the ORF8 protein or make it inactive, allowing the accumulation of additional mutations in other regions [68]. Further, eight mutations namely D614G, D950N, F157Δ, L452R, P681R, R158Δ, T19R, and T478K were detected in the S protein in the four sequences of the Delta variant identified in this study. These mutations are identical to those detected in the Indian Delta variants (B.1.617.2) [69]. Deletions, insertions, frameshift variants, and up/downstream variants were much rarer. This observation is also in line with the finding of Malune et al., whose study reported less than 10% of these mutations [70].

Sequences of the Omicron variant obtained in this study were highly mutated, having 149 mutations across the 27 sequences examined. The findings are consistent with the findings of Saxena et al., who detected more mutations in Omicron variants than the Delta variant [71]. When the S protein mutations of the VOCs in this study were compared to the hCoV-19/Wuhan/Hu-1/2019|EPI_ISL_402125, Omicron was highly mutated with 58 mutations and 22 amino acid mutations in the RBD. These mutations are crucial as they are thought to increase the overall risk of reinfection and partial resistance to existing vaccines [72]. In addition to mutations in the S protein, several substitutions and deletions in other genomic regions are also present in all the SARS-CoV-2 variants in this study. Moreover, mutations have an adverse impact on the pathogenicity of SARS-CoV-2 and the development of diagnostic assays, antivirals, and vaccines. Therefore, monitoring of mutations and characterization of their roles in virulence-related conditions in SARS-CoV-2 is very vital in the control and prevention of the spread of the virus.

The limitations of the study are that most of the samples could not be successfully sequenced because they had a Ct > 30, whereas others had poor genomic coverage. We believe poor sample quality was the main reason for the considerably low number of sequences obtained in this study which may have been due to poor storage and transportation conditions (i.e., failure to maintain a good cold chain), as some of the samples came from far-lying rural districts. For improved SARS-CoV-2 genomic surveillance, strengthening the capacity for sample storage and courier in rural areas should be prioritized by the Zambian Ministry of Health.

5. Conclusions

The findings highlighted the circulation of four VOCs in the Southern Province of Zambia namely Alpha (B.1.1.7), Beta (B.1.351), Delta (AY.116), and Omicron (BA.1, BA.1.1, BA.1.14 and BA.2). Phylogenetic analysis revealed that our genomes were closely related to genomes from Europe and Southern Africa indicating intra- and intercontinental introductions of the virus to the country. Additionally, some sequences that clustered with Zambian sequences may signify local transmission of the virus. The Omicron variant exhibited the highest number of amino acid substitutions in the S glycoprotein as compared to the other three variants in this study. Moreover, SARS-CoV-2 with the D614G and P314L mutation was the major circulating virus in Southern Province, Zambia. Our findings stress the need for continued monitoring of SARS-CoV-2 circulation in Zambia, especially in strategically positioned regions such as the Southern Province which could be at increased risk of introduction of novel VOCs. This analysis further represents the first genomic study in the Southern Province of Zambia and highlights the importance of the Zambia Genomic Sequencing Consortium in the expansion of SARS-CoV-2 genomic surveillance in understanding the spread of the virus at national and community levels. It has further contributed to the decentralization of sequencing facilities encompassing

among them public, private, and academic public health laboratories which have led to the rapid dissemination of sequences into the public domain.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/v14091865/s1, Table S1: Characteristics of 40 SARS-CoV-2 whole-genomes from the Southern Province of Zambia; Table S2: Nucleic acid and amino acid mutations observed in the 40 SARS-CoV-2 genomes obtained from the Southern Province of Zambia; Table S3: SARS-CoV-2 primer sequences; Table S4: SARS-CoV-2 genome map.

Author Contributions: Conceptualization, B.K. (Ben Katowa), A.K., B.M., W.M. and E.S.; methodology, B.K. (Ben Katowa), A.K., B.M., J.M., W.M. and E.S.; software, B.M.; formal analysis, B.K. (Ben Katowa), A.K. and B.M.; investigation, B.K. (Ben Katowa), A.K., B.M., W.M., and E.S.; resources, M.K. (Masahiro Kajihara), J.Y., A.T., H.S., R.C., V.M. and E.S.; writing—original draft preparation, B.K. (Ben Katowa) and A.K.; writing—review and editing, B.K. (Ben Katowa), A.K., B.M., J.M., D.M.S., J.S., B.K. (Boniface Kabungo), K.C., S.C., M.K. (Mpanga Kasonde), O.K., N.K., K.M., M.M., J.T., M.B., A.Z., C.G.S., M.K. (Masahiro Kajihara), J.Y., A.T., H.S., R.C., W.M. and E.S; supervision, W.M. and E.S.; project administration, M.K. (Masahiro Kajihara), J.Y., A.T., H.S., R.C., V.M. and E.S.; funding acquisition, A.T., H.S., R.C., V.M., M.B., A.Z. and E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Africa Centre of Excellence for Infectious Disease of Humans and Animals (ACEIDHA) project (grant number P151847) funded by the World Bank. Financial support was also provided in part by the Science and Technology Research Partnership for Sustainable Development (SATREPS) (grant number JP22jm0110019), grants for the Japan Program for Infectious Diseases Research and Infrastructure (grant number JP21wm0125008, JP20wm0225003) from Japan Agency for Medical Research and Development (AMED), and by the European and Developing Countries Clinical Trials Partnership (EDCTP2) programme under the PANDORA-ID-NET Consortium (EDCTP Reg/Grant RIA2016E-1609). The funders had no role in study design, collection, analysis and interpretation of data, manuscript writing, and decision to submit the article for publication.

Institutional Review Board Statement: Ethical approval to conduct this study was obtained from ERES Converge (Ref. No. 2020-Aug-008).

Informed Consent Statement: Patient consent was waived as the study utilized residual samples and patient information was de-identified.

Data Availability Statement: The sequences have been deposited in the GISAID EpiCoV (https://www.gisaid.org/ (accessed on 8 August 2022) under accession numbers: EPI_ISL_6768069; EPI_ISL_6764745; EPI_ISL_6760707; EPI_ISL_6760905; EPI_ISL_6760973; EPI_ISL_6760998; EPI_ISL_6761015; EPI_ISL_6761027; EPI_ISL_6761052; EPI_ISL_6761088; EPI_ISL_6761100; EPI_ISL_6762977; EPI_ISL_6761106; EPI_ISL_12363645; EPI_ISL_12363646; EPI_ISL_12363647; EPI_ISL_12363648; EPI_ISL_12363649; EPI_ISL_12363652; EPI_ISL_12363653; EPI_ISL_12363654; EPI_ISL_12363658; EPI_ISL_12363660; EPI_ISL_12363661; EPI_ISL_12363662; EPI_ISL_12363663; EPI_ISL_12363664; EPI_ISL_12363667; EPI_ISL_12363668; EPI_ISL_12363669; EPI_ISL_12363670; EPI_ISL_12363671; EPI_ISL_12363672; EPI_ISL_12363673; EPI_ISL_12363674; EPI_ISL_12363681; EPI_ISL_13053241; EPI_ISL_13053244; EPI_ISL_13053875 and EPI_ISL_13053876.

Acknowledgments: We thank Mukuma Lubinda from Macha Research Trust, Choma, Zambia for generating the map used in this study. Authors E.S., A.Z., J.T., and M.B. are members of the Pan-African Network for Rapid Research, Response, Relief and Preparedness for Infectious Diseases Epidemics funded by the European and Developing Countries Clinical Trials Partnership, which is supported by Horizon 2020, the EU's Framework Programme for Research and Innovation. A.Z. holds a UK National Institute of Health Research Senior Investigator Award and is a Mahathir Science Prize and Pascoal Mocumbi Prize laureate. All authors have an interest in epidemic infections.

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

References

1. WHO Coronavirus (COVID-19) Dashboard. 2022. Available online: https://covid19.who.int/ (accessed on 7 July 2022).
2. Uyoga, S.; Adetifa, I.M.O.; Karanja, H.K.; Nyagwange, J.; Tuju, J.; Wanjiku, P.; Aman, R.; Mwangangi, M.; Amoth, P.; Kasera, K.; et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* **2021**, *371*, 79–82. [CrossRef]
3. Mwananyanda, L.; Gill, C.J.; MacLeod, W.; Kwenda, G.; Pieciak, R.; Mupila, Z.; Lapidot, R.; Mupeta, F.; Forman, L.; Ziko, L.; et al. COVID-19 deaths in Africa: Prospective systematic postmortem surveillance study. *BMJ* **2021**, *372*, n334. [CrossRef]
4. Chisale, M.R.O.; Ramazanu, S.; Mwale, S.E.; Kumwenda, P.; Chipeta, M.; Kaminga, A.C.; Nkhata, O.; Nyambalo, B.; Chavura, E.; Mbakaya, B.C. Seroprevalence of anti-SARS-CoV-2 antibodies in Africa: A systematic review and meta-analysis. *Rev. Med. Virol.* **2022**, *32*, e2271. [CrossRef]
5. Gelanew, T.; Seyoum, B.; Mulu, A.; Mihret, A.; Abebe, M.; Wassie, L.; Gelaw, B.; Sorsa, A.; Merid, Y.; Muchie, Y.; et al. High seroprevalence of anti-SARS-CoV-2 antibodies among Ethiopian healthcare workers. *BMC Infect. Dis.* **2022**, *22*, 261. [CrossRef]
6. Lewis, H.C.; Ware, H.; Whelan, M.G.; Subissi, L.; Li, Z.; Ma, X.; Nardone, A.; Valenciano, M.; Cheng, B.; Noel, K.C.; et al. SARS-CoV-2 infection in Africa: A systematic review and meta-analysis of standardised seroprevalence studies, from January 2020 to December 2021. *medRxiv* **2022**, 2022.02.14.22270934. [CrossRef]
7. World Health Organisation. Over Two-Thirds of Africans Exposed to Virus Which Causes COVID-19: WHO Study. 2022. Available online: https://www.afro.who.int/news/over-two-thirds-africans-exposed-virus-which-causes-covid-19-who-study (accessed on 1 June 2022).
8. Massinga Loembé, M.; Tshangela, A.; Salyer, S.J.; Varma, J.K.; Ouma, A.E.O.; Nkengasong, J.N. COVID-19 in Africa: The spread and response. *Nat. Med.* **2020**, *26*, 999–1003. [CrossRef]
9. Oluniyi, P. First African SARS-CoV-2 Genome Sequence from Nigerian COVID-19 Case: Virological.org; 2020. Available online: https://virological.org/t/first-african-sars-cov-2-genome-sequence-from-nigerian-covid-19-case/421 (accessed on 31 May 2022).
10. World Health Organisation. A Second COVID-19 Case is Confirmed in Africa. 2020. Available online: https://www.afro.who.int/news/second-covid-19-case-confirmed-africa (accessed on 31 May 2022).
11. Elimian, K.O.; Ochu, C.L.; Ilori, E.; Oladejo, J.; Igumbor, E.; Steinhardt, L.; Wagai, J.; Arinze, C.; Ukponu, W.; Obiekea, C.; et al. Descriptive epidemiology of coronavirus disease 2019 in Nigeria, 27 February–6 June 2020. *Epidemiol. Infect.* **2020**, *148*, e208. [CrossRef]
12. Dufailu, O.A.; Afriyie-Asante, A.; Gyan, B.; Kwabena, D.A.; Yeboah, H.; Ntiakoh, F.; Asare-Werehene, M. COVID-19 in Africa: An ovarian victory? *J. Ovarian Res.* **2021**, *14*, 70. [CrossRef]
13. Wilkinson, E.; Giovanetti, M.; Tegally, H.; San, J.E.; Lessells, R.; Cuadros, D.; Martin, D.P.; Rasmussen, D.A.; Zekri, A.R.N.; Sangare, A.K.; et al. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* **2021**, *374*, 423–431. [CrossRef]
14. Sanyaolu, A.; Okorie, C.; Marinkovic, A.; Haider, N.; Abbasi, A.F.; Jaferi, U.; Prakash, S.; Balendra, V. The emerging SARS-CoV-2 variants of concern. *Ther. Adv. Infect. Dis.* **2021**, *8*, 20499361211024372. [CrossRef]
15. Dinnon, K.H.; Leist, S.R.; Schäfer, A.; Edwards, C.E.; Martinez, D.R.; Montgomery, S.A.; West, A.; Yount, B.L., Jr.; Hou, Y.J.; Adams, L.E.; et al. A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures. *Nature* **2020**, *586*, 560–566. [CrossRef]
16. Aleem, A.; Akbar Samad, A.B.; Slenker, A.K. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2022.
17. 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]
18. 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. [CrossRef]
19. Holland, L.A.; Kaelin, E.A.; Maqsood, R.; Estifanos, B.; Wu, L.I.; Varsani, A.; Halden, R.U.; Hogue, B.G.; Scotch, M.; Lim, E.S. An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *J. Virol.* **2020**, *94*, e00711-20. [CrossRef]
20. Addetia, A.; Xie, H.; Roychoudhury, P.; Shrestha, L.; Loprieno, M.; Huang, M.-L.; Jerome, K.R.; Greninger, A.L. Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J. Clin. Virol.* **2020**, *129*, 104523. [CrossRef]
21. Mazur-Panasiuk, N.; Rabalski, L.; Gromowski, T.; Nowicki, G.; Kowalski, M.; Wydmanski, W.; Szulc, P.; Kosinski, M.; Gackowska, K.; Drweska-Matelska, N.; et al. Expansion of a SARS-CoV-2 Delta variant with an 872 nt deletion encompassing ORF7a, ORF7b and ORF8, Poland, July to August 2021. *Eurosurveillance* **2021**, *26*, 2100902. [CrossRef]
22. Bal, A.; Destras, G.; Gaymard, A.; Bouscambert-Duchamp, M.; Valette, M.; Escuret, V.; Frobert, E.; Billaud, G.; Trouillet-Assant, S.; Cheynet, V.; et al. Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del). *Clin. Microbiol. Infect.* **2020**, *26*, 960–962. [CrossRef]
23. Young, B.E.; Fong, S.-W.; Chan, Y.-H.; Mak, T.-M.; Ang, L.W.; Anderson, D.E.; Lee, C.Y.-P.; Amrun, S.N.; Lee, B.; Goh, Y.S.; et al. Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: An observational cohort study. *Lancet* **2020**, *396*, 603–611. [CrossRef]

24. 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]
25. Karadag, E. Increase in COVID-19 cases and case-fatality and case-recovery rates in Europe: A cross-temporal meta-analysis. *J. Med. Virol.* **2020**, *92*, 1511–1517. [CrossRef]
26. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **2021**, *592*, 438–443. [CrossRef]
27. Hirabara, S.M.; Serdan, T.D.A.; Gorjao, R.; Masi, L.N.; Pithon-Curi, T.C.; Covas, D.T.; Curi, R.; Durigon, E.L. SARS-CoV-2 Variants: Differences and potential of immune evasion. *Front. Cell. Infect. Microbiol.* **2022**, *11*, 781429. [CrossRef]
28. Singanayagam, A.; Hakki, S.; Dunning, J.; Madon, K.J.; Crone, M.A.; Koycheva, A.; Derqui-Fernandez, N.; Barnett, J.L.; Whitfield, M.G.; Varro, R.; et al. Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: A prospective, longitudinal, cohort study. *Lancet Infect. Dis.* **2022**, *22*, 183–195. [CrossRef]
29. Telenti, A.; Hodcroft, E.B.; Robertson, D.L. The evolution and biology of SARS-CoV-2 variants. *Cold Spring Harb. Perspect. Med.* **2022**, *12*, a041390. [CrossRef]
30. Chipimo, P.J.; Barradas, D.T.; Kayeyi, N.; Zulu, P.M.; Muzala, K.; Mazaba, M.L.; Hamoonga, R.; Musonda, K.; Monze, M.; Kapata, N.; et al. First 100 persons with COVID-19-Zambia, March 18–April 28, 2020. *MMWR. Morb. Mortal. Wkly. Rep.* **2020**, *69*, 1547–1548. [CrossRef]
31. Zambia National Public Health Institute. Zambia COVID-19 Dashboard. 2022. Available online: https://www.arcgis.com/apps/dashboards/3b3a01c1d8444932ba075fb44b119b63 (accessed on 21 May 2022).
32. Worldometer. Coronavirus. Available online: https://www.worldometers.info/coronavirus/country/zambia/ (accessed on 13 June 2022).
33. Pandemic Coronavirus Causing COVID-19. 2020. Available online: https://www.epicov.org/epi3/frontend#1739f3 (accessed on 21 May 2022).
34. 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]
35. Mwenda, M.; Saasa, N.; Sinyange, N.; Busby, G.; Chipimo, P.J.; Hendry, J.; Kapona, O.; Yingst, S.; Hines, J.Z.; Minchella, P.; et al. Detection of B.1.351 SARS-CoV-2 Variant Strain-Zambia, December 2020. *MMWR. Morb. Mortal. Wkly. Rep.* **2021**, *70*, 280–282. [CrossRef]
36. Lu, X.; Wang, L.; Sakthivel, S.K.; Whitaker, B.; Murray, J.; Kamili, S.; Lynch, B.; Malapati, L.; Burke, S.A.; Harcourt, J.; et al. US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **2020**, *26*, 1654. [CrossRef]
37. Quick, J. nCoV-2019 Sequencing Protocol v3 (LoCost). Protocols.io. 2020. Available online: https://protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye (accessed on 5 April 2021).
38. Tyson, J.R.; James, P.; Stoddart, D.; Sparks, N.; Wickenhagen, A.; Hall, G.; Choi, J.H.; Lapointe, H.; Kamelian, K.; Smith, A.D.; et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* **2020**, 2020.09.04.283077. [CrossRef]
39. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinform.* **2017**, *20*, 1160–1166. [CrossRef]
40. Guindon, S.; Lethiec, F.; Duroux, P.; Gascuel, O. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **2005**, *33*, W557–W559. [CrossRef] [PubMed]
41. Lefort, V.; Longueville, J.-E.; Gascuel, O. SMS: Smart model selection in PhyML. *Mol. Biol. Evol.* **2017**, *34*, 2422–2424. [CrossRef] [PubMed]
42. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2016**, *2*, vew007. [CrossRef] [PubMed]
43. 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] [PubMed]
44. Taboada, B.; Zárate, S.; Iša, P.; Boukadida, C.; Vazquez-Perez, J.A.; Muñoz-Medina, J.E.; Ramírez-González, J.E.; Comas-García, A.; Grajales-Muñiz, 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]
45. Jin, J.M.; Bai, P.; He, W.; Wu, F.; Liu, X.F.; Han, D.M.; Liu, S.; Yang, J.K. Gender differences in patients with COVID-19: Focus on severity and mortality. *Front. Public Health* **2020**, *8*, 152. [CrossRef]
46. Yadav, P.D.; Nyayanit, D.A.; Majumdar, T.; Patil, S.; Kaur, H.; Gupta, N.; Shete, A.M.; Pandit, P.; Kumar, A.; Aggarwal, N.; et al. An epidemiological analysis of SARS-CoV-2 genomic sequences from different regions of India. *Viruses* **2021**, *13*, 925. [CrossRef]
47. Tegally, H.; Wilkinson, E.; Lessells, R.J.; Giandhari, J.; Pillay, S.; Msomi, N.; Mlisana, K.; Bhiman, J.N.; von Gottberg, A.; Walaza, S.; et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.* **2021**, *27*, 440–446. [CrossRef]
48. 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]

49. Salyer, S.J.; Maeda, J.; Sembuche, S.; Kebede, Y.; Tshangela, A.; Moussif, M.; Ihekweazu, C.; Mayet, N.; Abate, E.; Ouma, A.O.; et al. The first and second waves of the COVID-19 pandemic in Africa: A cross-sectional study. *Lancet* **2021**, *397*, 1265–1275. [CrossRef]
50. Plante, J.A.; Liu, Y.; Liu, J.; Xia, H.; Johnson, B.A.; Lokugamage, K.G.; Zhang, X.; Muruato, A.E.; Zou, J.; Fontes-Garfias, C.R.; et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **2021**, *592*, 116–121. [CrossRef]
51. Khan, K.; Karim, F.; Ganga, Y.; Bernstein, M.; Jule, Z.; Reedoy, K.; Cele, S.; Lustig, G.; Amoako, D.; Wolter, N.; et al. Omicron BA.4/BA.5 escape neutralizing immunity elicited by BA.1 infection. *Nat. Commun.* **2022**, *13*, 4686. [CrossRef]
52. Tegally, H.; Moir, M.; Everatt, J.; Giovanetti, M.; Scheepers, C.; Wilkinson, E.; Subramoney, K.; Moyo, S.; Amoako, D.G.; Althaus, C.L.; et al. Continued emergence and evolution of omicron in South Africa: New BA.4 and BA.5 lineages. *medRxiv* **2022**, 2022.05.01.22274406. [CrossRef]
53. Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: Implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* **2020**, *17*, 613–620. [CrossRef]
54. Du, L.; He, Y.; Zhou, Y.; Liu, S.; Zheng, B.-J.; Jiang, S. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **2009**, *7*, 226–236. [CrossRef]
55. Lu, L.; Liu, Q.; Zhu, Y.; Chan, K.-H.; Qin, L.; Li, Y.; Wang, Q.; Chan, J.F.-W.; Du, L.; Yu, F.; et al. Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. *Nat. Commun.* **2014**, *5*, 3067. [CrossRef]
56. Du, L.; Tai, W.; Yang, Y.; Zhao, G.; Zhu, Q.; Sun, S.; Liu, C.; Tao, X.; Tseng, X.T.C.-T.K.; Perlman, S.; et al. Introduction of neutralizing immunogenicity index to the rational design of MERS coronavirus subunit vaccines. *Nat. Commun.* **2016**, *7*, 13473. [CrossRef]
57. Duan, L.; Zheng, Q.; Zhang, H.; Niu, Y.; Lou, Y.; Wang, H. The SARS-CoV-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: Implications for the design of spike-based vaccine immunogens. *Front. Immunol.* **2020**, *11*, 576622. [CrossRef]
58. Sahin, E.; Bozdayi, G.; Yigit, S.; Muftah, H.; Dizbay, M.; Tunccan, O.G.; Fidan, I.; Caglar, K. Genomic characterization of SARS-CoV-2 isolates from patients in Turkey reveals the presence of novel mutations in spike and nsp12 proteins. *J. Med. Virol.* **2021**, *93*, 6016–6026. [CrossRef]
59. Al-Mahruqi, S.; Al-Wahaibi, A.; Khan, A.L.; Al-Jardani, A.; Asaf, S.; Alkindi, H.; Al-Kharusi, S.; Al-Rawahi, A.N.; Al-Salmani, M.; Al-Shukri, I.; et al. Molecular epidemiology of COVID-19 in Oman: A molecular and surveillance study for the early transmission of COVID-19 in the country. *Int. J. Infect. Dis.* **2021**, *104*, 139–149. [CrossRef]
60. Zekri, A.-R.N.; Amer, K.E.; Hafez, M.M.; Hassan, Z.K.; Ahmed, O.S.; Soliman, H.K.; Bahnasy, A.A.; Hamid, W.A.; Gad, A.; Ali, M.; et al. Genomic characterization of SARS-CoV-2 in Egypt. *J. Adv. Res.* **2021**, *30*, 123–132. [CrossRef] [PubMed]
61. Agoti, C.N.; Githinji, G.; Mohammed, K.S.; Lambisia, A.W.; de Laurent, Z.R.; Mburu, M.W.; Ong'Era, E.M.; Morobe, J.M.; Otieno, E.; Azali, H.A.; et al. Detection of SARS-CoV-2 variant 501Y.V2 in comoros islands in January 2021. *Wellcome Open Res.* **2021**, *6*, 192. [CrossRef]
62. Obeid, D.A.; Alsanea, M.S.; Alnemari, R.T.; Al-Qahtani, A.A.; Althawadi, S.I.; Mutabagani, M.S.; Almaghrabi, R.S.; Alhadheq, F.M.; Alahideb, B.M.; Alhamlan, F.S. SARS-CoV-2 genetic diversity and variants of concern in Saudi Arabia. *J. Infect. Dev. Ctries.* **2021**, *15*, 1782–1791. [CrossRef]
63. Flores-Alanis, A.; Cruz-Rangel, A.; Rodríguez-Gómez, F.; González, J.; Torres-Guerrero, C.A.; Delgado, G.; Cravioto, A.; Morales-Espinosa, R. Molecular epidemiology surveillance of SARS-CoV-2: Mutations and genetic diversity one year after emerging. *Pathogens* **2021**, *10*, 184. [CrossRef] [PubMed]
64. Rahman, M.M.; Kader, S.B.; Rizvi, S.M.S. Molecular characterization of SARS-CoV-2 from Bangladesh: Implications in genetic diversity, possible origin of the virus, and functional significance of the mutations. *Heliyon* **2021**, *7*, e07866. [CrossRef] [PubMed]
65. Eskier, D.; Karakülah, G.; Suner, A.; Oktay, Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. *PeerJ* **2020**, *8*, e9587. [CrossRef]
66. Haddad, D.; John, S.E.; Mohammad, A.; Hammad, M.M.; Hebbar, P.; Channanath, A.; Nizam, R.; Al-Qabandi, S.; Al Madhoun, A.; Alshukry, A.; et al. SARS-CoV-2: Possible recombination and emergence of potentially more virulent strains. *PLoS ONE* **2021**, *16*, e0251368. [CrossRef]
67. Korber, B.; Fischer, W.M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E.E.; Bhattacharya, T.; Foley, B.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **2020**, *182*, 812–827. [CrossRef]
68. Gómez, C.E.; Perdiguero, B.; Esteban, M. Emerging SARS-CoV-2 Variants and Impact in Global Vaccination Programs against SARS-CoV-2/COVID-19. *Vaccines* **2021**, *9*, 243. [CrossRef]
69. Kang, N.; Prevention, N.G.C.F.D.C.A.; Liu, Y.; Bi, F.; Wang, J.; Chen, H.; Liang, Z.; Chen, M.; Wei, G.; Li, A.; et al. Detection of variants of B.1.617 lineage from five returning Chinese nationals at a Guangxi-vietnam border port—Guangxi Zhuang Autonomous Region, China. *China CDC Wkly.* **2021**, *30*, 653–655. [CrossRef]
70. Malune, P.; Piras, G.; Monne, M.; Fiamma, M.; Asproni, R.; Fancello, T.; Manai, A.; Carta, F.; Fancello, P.; Rosu, V.; et al. Molecular characterization of severe acute respiratory syndrome coronavirus 2 isolates from central inner sardinia. *Front. Microbiol.* **2022**, *12*, 827799. [CrossRef] [PubMed]

71. Saxena, S.K.; Kumar, S.; Ansari, S.; Paweska, J.T.; Maurya, V.K.; Tripathi, A.K.; Abdel-Moneim, A.S. Characterization of the novel SARS-CoV-2 Omicron (B.1.1.529) variant of concern and its global perspective. *J. Med. Virol.* **2022**, *94*, 1738–1744. [CrossRef] [PubMed]
72. Geddes, L. From Alpha to Omicron: Everything You Need to Know about Coronavirus Variants of Concern: Gavi the Vaccine Alliance. 2021. Available online: https://www.gavi.org/vaccineswork/alpha-omicron-%20everything-you-need-know-about-coronavirus-variants-concern (accessed on 24 May 2022).