Molecular epidemiology of M and E protein coding genes from South African SARS-CoV-2 strains, 2020 to 2021.

by

Fabian Marsden

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

A Dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of

Master of Sciences (MSc) in Medicine (Medical Virology)

2023

Declaration

I, Fabian Marsden declare that this thesis is my own, unaided work. It is being submitted for the Degree of Masters of Virology at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

___________Fabian Marsden______________

(Signature of candidate)

____________________day of _____________22/03/2023____________at_______13:37______

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the current pathogen causing the COVID-19 pandemic across the world. While vaccines that elicit anti-SARS-CoV-2 antibodies have been developed and licenced, there is a reduced protection against variants of concern (VOCs) such as Beta, Delta and Omicron. This is due to the mutations within the spike (S) protein which is the antigen targeted by most vaccines. Other potential vaccine targets include the structural proteins namely the membrane (M) and envelope (E) proteins of SARS-CoV-2 which are more conserved. In this study we aimed to determine the extent of genetic diversity in the M and E protein genes from South African SARS-CoV-2 strains and its impact on predicted B and T cell epitopes. M and E gene sequences were extracted from South African SARS-CoV-2 genomes obtained from the Global Initiative on Sharing All Influenza Data (GISAID) database for the period 01March 2020 to 31 December 2021. Maximum-likelihood phylogenetic tree analysis shows that among South African E gene sequences only the Omicron VOC sequences form a distinct cluster. Similarly, the Omicron M gene sequences also form a distinct cluster compared to the Wuhan reference strain, Beta and Delta sequences. The predicted T cell and B cell epitopes of M and E proteins were identified with specific regions that have shown to have identical regions in both the variants and the reference strain, this shows the conserved nature of the M and E genes. SARS-CoV-2 are shown to have varying antigenic probabilities for M and E proteins from each of the variants considered as probable antigens. The allergenicity and toxicity of the M and E proteins was assessed in the context of potential vaccine development with certain peptides of each shown to have toxic properties. The predicted B and T cell epitopes show that despite the presence of mutations in the VOCs' derived protein sequences, there is a common epitopic region that is shared between the reference and the variants. There is a strong 9-mer coverage by the natural sequences despite some non-coverage due to non-silent mutations. The results from the epitope predication and HLA typing shows the conserved nature of the M and E proteins which highlights the potential use for the development of vaccines.

Acknowledgements

I would like to thank the support and guidance of my supervisor, Dr Florette K. Treurnicht, the support of several colleague Kathleen Subramoney, Avani Bharuthram and Houcemeddine Othman as well as the support of my family.

List of Abbreviations

+ssRNA:	Single-stranded positive-sense RNA
A:	Adenine
ACE2:	Angiotensin-converting enzyme 2
ARDS:	Acute respiratory distress syndrome
BEAST:	Bayesian Evolutionary Analysis Sampling Trees
C:	Cytosine
COVID-19:	Coronavirus disease 2019
CTD:	C-terminal domain
CTL:	Cytotoxic T lymphocytes
DlgA:	Drosophila disc large tumour suppressor
DNA:	Deoxyribonucleic acid
E:	Envelope
ERGIC:	ER-Golgi intermediate compartment
ESS:	Effective Sampling Size
G:	Guanine
GISAID:	Global Initiative on Sharing All Influenza Data
GTR:	General Time Reversible
hACE2:	human Angiotensin-converting enzyme 2
HLA:	Human leukocyte antigen
J&J:	Johnson & Johnson
kbp:	Kilobase pairs
kDa:	kiloDalton
LINK:	Predicted disordered central linker
M:	Membrane
MEGA:	Molecular Evolutionary Genetics Analysis
MERS-CoV:	Middle East respiratory syndrome coronavirus
MHC:	Major histocompatibility complex
mRNA:	Messenger RNA
N:	Nucleocapsid
NSP:	Non-structural protein
NTD:	N-terminal domain
ORFs:	Open reading frame

PDZ:	Postsynaptic density protein of 95 kDa Drosophila disc large tumour suppressor and zomula occludens-1 protein
PSD95:	Postsynaptic density protein of 95 kDa
RBD:	Receptor-binding domain
RNA:	Ribonucleic acid
RT-PCR	Real-Time Polymerase Chain Reaction
S:	Spike
SARS-CoV:	Severe acute respiratory syndrome coronavirus
SARS-CoV-2:	Severe acute respiratory syndrome coronavirus 2
SVM:	Support vector machine
T:	Thymine
TLR:	Toll-like receptor
TMD:	Transmembrane domain
TMPRSS2:	Transmembrane protease serine 2
VOC:	Variant of concern
VOI:	Variant of interest
Zo-1:	Zomula occludens-1 protein

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the global health crisis causing worldwide deaths, rippling effects on the global economy and overwhelmed many healthcare system (Cascella *et al.*, 2022). SARS-CoV-2 causes the disease Coronavirus disease 2019 (COVID-19) which primarily affects the respiratory and cardiovascular systems causing pneumonia and hypercoagulability, respectively (Cascella *et al.*, 2022). Hypercoagulability is a recognised severe complication of COVID-19. SARS-CoV-2 is prone to genetic mutations that are associated with increased transmissibility and reduction in neutralization by the vaccine stimulated immunity (Cascella *et al.*, 2022).

There are a few strains of SARS-CoV-2 that are considered as variants of concern (VOCs) (Cascella *et al.*, 2022). These VOCs include Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B1.617.2) and the most recent Omicron (B.1.1.529) (Cascella *et al.*, 2022). The mutations in the spike protein of the VOCs are responsible for increases in transmissibility and reduction in vaccines' efficacy (Sanyaolu *et al.*, 2021). These mutations on the spike protein are found most notably on the S1 subunit, resulting in the variations in immune responses and transmission (Khateeb *et al.*, 2021).

The two VOCs first identified in South Africa are Beta and Omicron (Tegally *et al.*, 2021b, Wei *et al.*, 2021). The Alpha variant was briefly detected in South Africa with no detections since August 2021 (Wolter *et al.*, 2021). The Beta variant posses' three substitutions namely K417N, E484K and N501Y in the receptor-binding domain of the spike protein and the Omicron VOC has over 37 mutations in the spike protein with 26 mutations that are completely distinct to Omicron like E484A and S477N (Venkatakrishnan *et al.*, 2021, Tegally *et al.*, 2021c, Zhao *et al.*, 2022). The Omicron variant is the leading cause of the third wave of infections globally which started in December of 2021 (Yang and Shaman, 2021). There are several lineages of Omicron that became prominent in South Africa and include the BA.1 and BA.2 lineages which dominated the fourth wave as well as the BA.4 and BA.5 lineages that dominated the fifth wave (Mohapatra *et al.*, 2022, Wolter *et al.*, 2022).

1.2 Taxonomy of SARS-CoV-2

Coronaviruses are one of the largest groups viruses with the family *Coronaviridae* is separated into two subfamilies, *Letovirinae and Orthocoronvirinae* (Helmy *et al.*, 2020). *Letovirinae* includes the genus *Alphaletovirus* and *Orthocoronvirinae* includes the four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* (Helmy *et al.*, 2020). The taxonomy of SARS-CoV-2 (Table 1.1) classify it in the genus *Betacoronavirus* and the subgenus *Sarbecovirus* (Gorbalenya *et al.*, 2020). The genome of SARS-CoV-2 is composed of a single-stranded positive-sense RNA (+ssRNA), with a length of between 27-32 kb (Alanagreh *et al.*, 2020). The genome has a 38% of GC content, 11 protein-coding genes for12 expressed proteins (Figure 1.1) (Naqvi *et al.*, 2020, Alanagreh *et al.*, 2020). The genetic arrangement of the ORFs of SAR-CoV-2 closely resemble the ORFs rearrangement of other coronaviruses (Naqvi *et al.*, 2020).

Table 1.1 Classification of SARS-CoV-2 (Gorbalenya *et al.*, 2020)

Realm	Riboviria
Order	*Nidovirales*
Suborder	*Cornidovirineae*
Family	*Coronaviridae*
Subfamily	*Orthocoronavirinae*
Genus	*Betacoronavirus*
Subgenus	*Sarbecovirus*
Species	Severe acute respiratory syndrome-related coronavirus
Individuum	SARS-CoV-2

1.3 Phenotypic characteristics of SARS-CoV-2

SARS-CoV-2 is a spherical enveloped, non-segmented, positive sense RNA virus, it has a diameter of around 91nm and contains a single strand of RNA and is surrounded by crown-like spikes on the outer shell (Astuti and Ysrafil, 2020, Ke *et al.*, 2020). SARS-CoV-2 has four main structural proteins which include a small envelope (E) glycoprotein, membrane (M) glycoprotein, nucleocapsid (N) protein and spike (S) glycoprotein and well as several accessory proteins (Astuti and Ysrafil, 2020). The S glycoprotein is a heavily glycosylated trimer with each protomer composed of 1260 amino acids and is composed of two subunits (Duan *et al.*, 2020). The S glycoprotein which facilitates viral adhesion and fusion consists of the two subunits S1 and S2 (Kumar *et al.*, 2020). The surface subunit S1 is composed of 672 amino acids and is organized in four domains: N-terminal domain (NTD), C-terminal domain (CTD) and two subdomains (Duan *et al.*, 2020). The transmembrane S2 subunit is composed of 588 amino acids and contains two heptad repeats, a cytoplasmic tail, N-terminal hydrophobic fusion peptide and a transmembrane domain (Duan *et al.*, 2020).

The E protein is an integral membrane proteins of 75-109 amino acids with a single helical transmembrane domain (TMD) and a long amphiphilic region comprising of one or two α-helices at the C-terminus (Kuzmin *et al.*, 2021). The M glycoprotein is composed of 222 amino acids spans the membrane bilayer with a NH2-terminal domain outside and a long COOH-terminus inside (Thomas, 2020, Gorkhali *et al.*, 2021). The M and E proteins are required for the morphogenesis and budding of the virus (Kumar *et al.*, 2020). The N protein is composed of 419 amino acids and is divided into five domains: a dimerization domain, a predicted disorder C-terminal domain, an intrinsically disordered N-terminal domain (NTD), a RNA-binding domain (RBD) and a predicted disordered central linker (LINK) (Cubuk *et al.*, 2021, Gao *et al.*, 2021). The N protein is associated with the viral genome and are released into the host cell during fusion to facilitates replication, particle assembly and virion release (Zeng *et al.*, 2020).

1.4 Epidemiology of SARS-CoV-2

The wet market in Wuhan, China is considered to be the early and major epicentre of SARS-CoV-2 due to the epidemiological data (Holmes *et al.*, 2021). SARS-CoV-2 mode of transmission from human-to-human occurs via droplets, direct or indirect contact with contaminated surfaces (Guo *et al.*, 2020). The population groups with a higher vulnerability to SARS-CoV-2 are individuals over the age of 50 and those who have underlining conditions such as diabetes, cardiovascular disease and hypertension (Guo *et al.*, 2020). SARS-CoV-2 has a high transmission rate due to asymptomatic persons unknowingly transmitting the virus and a lax in the practice of self-isolation and social distancing (Rahman *et al.*, 2020a).

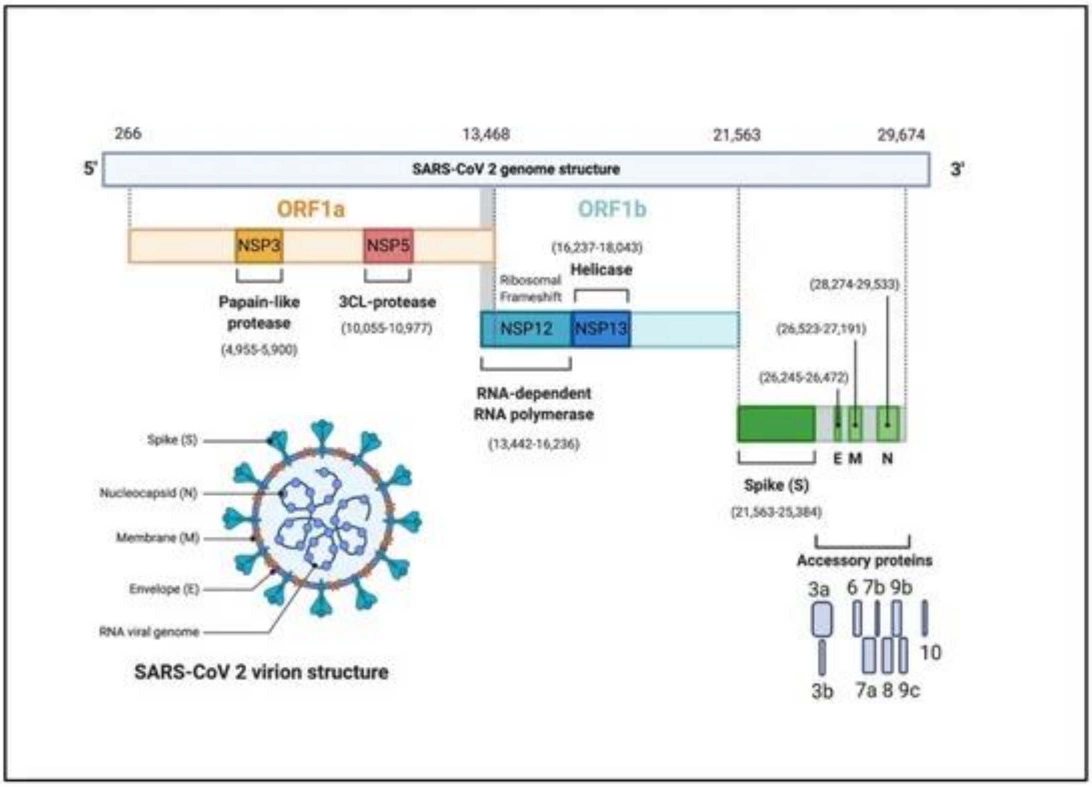


Figure 1.1 Graphical display of the genomic organization of SARS-CoV-2. Depiction of the open-reading frames (ORFs), proteins and the virion structure (Alanagreh *et al.*, 2020). NSP: non-structural protein.

1.5 Pathogenesis of SARS-CoV-2

The entry of SARS-CoV-2 into target cells is mediated by the S glycoprotein (Harrison *et al.*, 2020). The receptor-binding domain (RBD) of S glycoprotein binds to the cellular receptor, angiotensin-converting enzyme 2 (ACE2) and the S1/S2 polybasic cleavage site is proteolytically cleaved by the cellular cathepsin L and transmembrane protease serine 2 (TMPRSS2) (Harrison *et al.*, 2020). TMPRSS3 functions to allow viral entry into the plasma membrane, whereas cathepsin L activates the Spike in the endosomes and can allow for viral entry in the absence of TMPRSS2 (Figure 1.2) (Harrison *et al.*, 2020). The virus enters the host cell via an endosome(Hernandez Acosta *et al.*, 2022). The viral genome is released from the endosome when it is cleaved by cell protease, cathepsin or a decrease in pH (Hernandez Acosta *et al.*, 2022). The viral genome is then taken to the ER-Golgi intermediate compartment (ERGIC) for virion assembly (Klein *et al.*, 2020). The virion is then released from the host cell by exocytosis (V'kovski *et al.*, 2021).

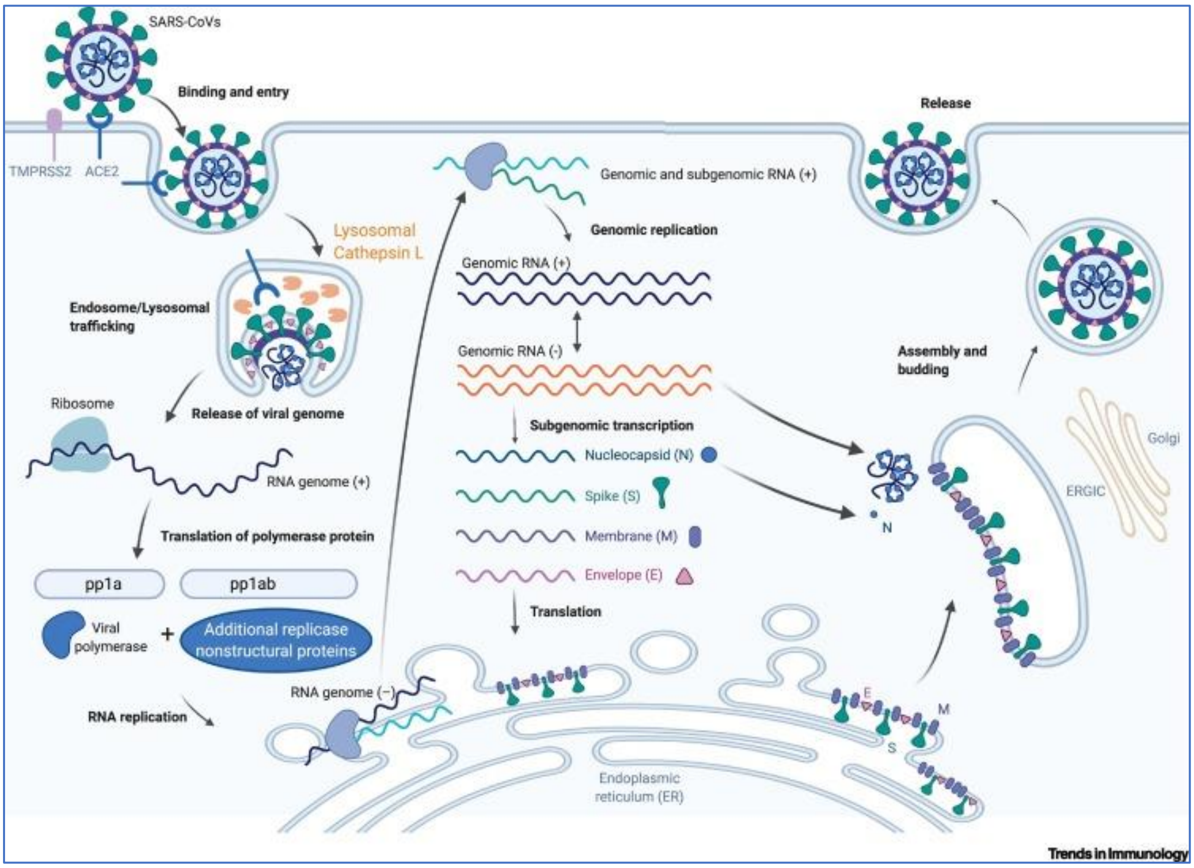


Figure 1.2 Graphical representation of the SARS-CoV-2 lifecycle. Indication of the stages of binding and entry, viral replication, assembly, virion budding and release (Harrison *et al.*, 2020) (Permission obtained from Elsevier by Rightslink).

1.6 Clinical manifestation of SARS-CoV-2

The typical symptoms of COVID-19 occur 5 days after infection and express symptoms such as a fever, dry cough and fatigue (Hu *et al.*, 2021). The onset of symptoms usually occurs after the 5 day incubation period and symptom duration can vary with mild and severe symptoms persisting for 8 days and critical symptoms lasting for 16 days (Hu *et al.*, 2021). Other potential symptoms include a sore throat, chest pains, sneezing, nasal congestion, nausea and sputum production (Lotfi *et al.*, 2020). Severe complications have been reported such as hypoxaemia, acute cardiac injury and acute kidney injury (Harapan *et al.*, 2020). Severe COVID-19 is characterized by a hyper-inflammatory response with acute respiratory distress syndrome (ARDS), multiorgan failure, and coagulation disorders leading to death (Lariccia *et al.*, 2020).

1.7 Diagnosis of SARS-CoV-2 and impact of variants

The diagnosis of SARS-CoV-2 infection is important aspect of managing the treatment and prevention as the correct treatment strategies and required containments need to coincide with an accurate detection (Kevadiya *et al.*, 2021). There are diagnostic tests that detect viral nucleic acids, antibodies and viral proteins but the use of Real-Time Polymerase Chain Reaction (RT-PCR) is the stranded for detection of viral nucleic acid (Kevadiya *et al.*, 2021). The preferred choice of specimen for RT-PCR is nasopharyngeal specimen (Mathuria *et al.*, 2020).The targets of RT-PCR are the helicase (ORF 1ab), nucleocapsid (N), transmembrane, S glycoproteins and the envelope (E) genes within the RNA of the virus (Mathuria *et al.*, 2020). Most RT-PCR have an accurate result of 100% but false negative can occur as a result of sampling error or inappropriate timing of sampling (Mathuria *et al.*, 2020). However, genetic changes associated with variants can be present in common RT-PCR targets such as the N-gene, S-gene and ORF1ab gene which may impact molecular diagnostic tests negatively. This impact has been noticeable while using TaqPath assay where inability to detect the S gene (S gene fallout or S gene target failure =SGTF) were linked to the emergence of the Alpha variant with deletions at S gene codon positions 69 and 70 (Firkins *et al.*, 2021). Subsequently, SGTF was also associated with the Omicron BA.1 lineage (Subramoney *et al.*, 2022). The point of care Xpert SARS-CoV-2 assays (Cepheid, Sunnyvale, CA) has been associated with failure to detect the N gene target (N gene target failure = NGTF) due to point mutations (C29200T, C29200A, G29140T and C29197T) in the targeted N fragment (Lopez *et al.*, 2022). The C29197T associated NGTF were associated with lineage B.1.1.519 and B.1.427/B.1.429 viruses. However, this mutation was also sporadically found in VOCs like Alpha, Beta and Gamma (Lopez *et al.*, 2022). NGTF was also associated point mutations in Delta variants which affected detection of the N target with the TaqPath assay (Holland *et al.*, 2022).

1.8 Treatment and prevention

Most preventive strategies implemented in Turkey were aimed at isolation of infected patients and infection control (Güner *et al.*, 2020). Prevention and control measures similar to what was implemented in South Africa included mandatory quarantine and other practices such as avoiding crowds, hand hygiene, use of personal protective equipment, and social distancing (Güner *et al.*, 2020). South Africa used a risk-adjusted strategy for restrictions during the global outbreak (Majam *et al.*, 2021). Another crucial factor in prevention was increasing testing capacity in order to pinpoint where infections occur (Güner *et al.*, 2020).

The most effective way to slow down the spread of SARS-CoV-2 is the use of herd immunity (Liu *et al.*, 2020). Herd immunity comes from the effects of acquired immunity in individual through either natural infection or immunization with a vaccine at the population level (Randolph and Barreiro, 2020). It has been estimated that the COVID-19 pandemic will decrease due to high levels of herd immunity (Liu *et al.*, 2020). It was estimated that to achieve herd immunity levels of 75% to 80% the COVID-19 vaccines should have an efficacy of 80% which most vaccines failed to achieve with a single dose (Britton *et al.*, 2020). However, another concern to achieve and maintain levels of herd immunity was the observation of rapid waning of immunity following natural infection. Figure 1.3 summarises the types of vaccines that been developed for the prevention of COVID-19. The live attenuated vaccines use a weakened or inactivated antigen to trigger an immune response while an inactivated uses only dead cells of the virus (Uttarilli *et al.*, 2021). Viral vector vaccines use attenuated viral vectors that retain the ability to replicate thus enhancing the immune response (Uttarilli *et al.*, 2021). The protein and synthetic peptide vaccines use a protein use a protein conjugated with a non-infectious recombinant proteins or synthetic peptides to trigger the immune system (Uttarilli *et al.*, 2021). Nucleic acid-based vaccines contain the specific genetic constructs of the virus which express the specific components to activate the immune responses (Uttarilli *et al.*, 2021). Synthetic peptides can target viral structures essential for viral replication (Schütz *et al.*, 2020).

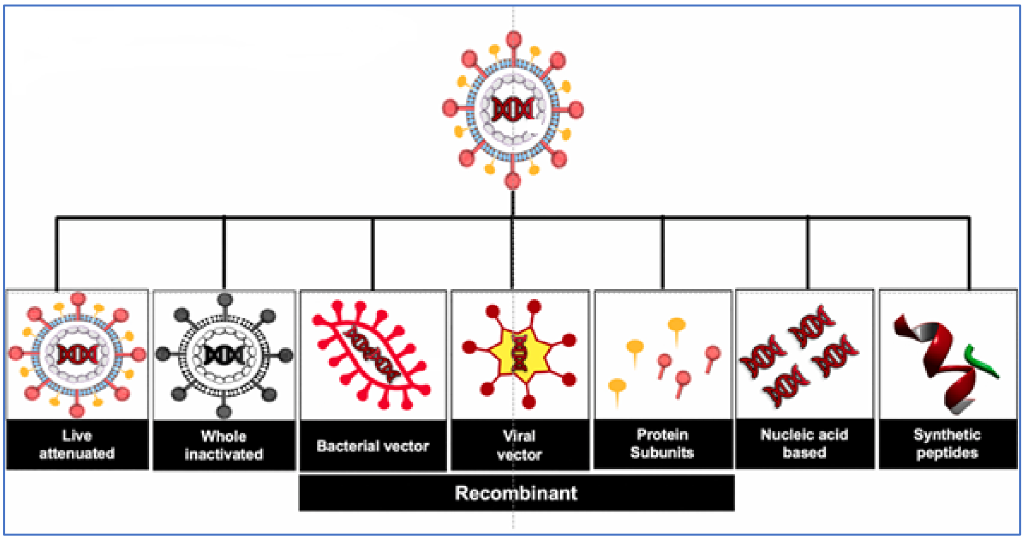


Figure 1.3 Graphical overview of the different types of vaccines that can produced against SARS-CoV-2 (Uttarilli *et al.*, 2021).

The S glycoprotein is recognized by the immune system triggering protective cellular and humoral immune responses which is why vaccines against SARS-CoV-2 are based on the Spike glycoprotein (Martínez-Flores *et al.*, 2021). There are a variety of vaccines platforms that are used: viral vector vaccine, whole virus vaccines, nucleic acid vaccine and protein-based vaccines (Tavilani *et al.*, 2021). There are four main vaccines that are mainly used globally: BNT162b2/ Comirnaty vaccine (Pfizer-BioNTech/ Fosun Pharma), mRNA-1273 vaccine (Moderna/ NIAID), JNJ-78436735/ Ad26.COV2.S vaccine (Janssen Pharmaceutical companies of Johnson & Johnson) and the AZD1222 vaccine (AstraZeneca/ Oxford University) (Tavilani *et al.*, 2021). One of the main vaccines approved for use in South Africa is the Johnson & Johnson (J&J) vaccine which uses a adenovirus vector (Livingston *et al.*, 2021). The other vaccine approved for use is the Pfizer vaccine which is a mRNA-based vaccine (Shahcheraghi *et al.*, 2021). Both vaccine types work to induce production of multiple pro-inflammatory cytokines and chemokines as well as the stimulation of S protein-specific T and B cells responses (Teijaro and Farber, 2021). The two vaccine types differ in that the mRNA vaccine triggers the RNA sensor called Toll-like receptor (TLR) 7 and melanoma differentiation-associated protein 5 whereas the adenovirus vector stimulates the major double-stranded DNA sensor TLR 9 (Teijaro and Farber, 2021). Toll-like receptor stimulates the innate immune responses through the recognition of specific molecular patterns on the virus. The Beta variant which was the predominant strain in the 2nd wave of infection in South Africa has resulted in lower efficacy of various vaccines such as the AstraZeneca vaccine which exhibited an efficacy as low as 10% which are no longer used as vaccines after being suspended due concerns about the actual benefit to vaccine recipients (Venter *et al.*, 2021, Wise, 2021). However, from 11 May 2021 South Africa entered a 3rd wave of infection driven predominantly by the Delta variant that emerged in India (Tegally *et al.*, 2021a). The Delta VOC displayed a 10-20% reduction in vaccine effectiveness (Harder *et al.*, 2021).

Various antivirals, anti-inflammatory, anticoagulant and other drugs are used in the treatment of COVID-19 (Niknam *et al.*, 2022). Remdesivir is a nucleotide prodrug inhibitor of SARS-CoV-2 RNA-dependent RNA polymerase and has shown a shortened recovery time for patients in hospitals (Gottlieb *et al.*, 2021). Baricitinib is the first immunomodulatory treatment for SARS-CoV-2 to receive FDA approval (Rubin, 2022). Baricitinib is an oral inhibitor of Janus Kinase (JAK) and has demonstrated multiple mechanisms of anti-viral activity (Marconi *et al.*, 2021).

1.9 Variation in SARS-CoV-2

The emergence of VOCs which contain sets of mutation that impact the transmissibility and antigenicity of SARS-CoV-2 have played a role in changing the immune profile of the virus (Harvey *et al.*, 2021). VOCs of SARS-CoV-2 are classified as being associated with one or more of the following changes that have public health significance such as an increase in transmissibility, virulence or a deceases in effectiveness in vaccines and therapeutics (Choi and Smith, 2021). The mutations in the S protein that allow for an escape from the immune system are the E484K, R346K, N501Y of the Beta variant which reduces the neutralizing ability of antibodies. (Harvey *et al.*, 2021, Koyama *et al.*, 2022). The Delta variant has three key amino acid mutations in the S protein, the L452R, T478K and P681R associated with increase transmissibility and immune escape (Tian *et al.*, 2021). The Omicron variant BA.4 and BA.5 lineage have two major mutations in the L452R and F486V which contribute to immune escape (Tuekprakhon *et al.*, 2022). Both the M and E proteins are highly conserved with low mutations rates when compared to S (Thakur *et al.*, 2022).

All viruses accumulate mutation in their genomes over time (Otto *et al.*, 2021). The genomes of coronaviruses such as SARS-CoV-2 encode a proofreading domain that reduce its mutation rate compared to other RNA viruses (Otto *et al.*, 2021). The estimated mutation rate of 1.87 x 10^{-6} nucleotide substitution per site per day, thus approximately 20 genetic changes per year within a lineage (Otto *et al.*, 2021). For ease of communication to the general public the World Health Organization (WHO) has provided a new nomenclature system SARS-CoV-2 variants of concern (VOCs) in late 2020: Alpha , Beta , Delta and Gamma (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/). Figure 1.4 Show the various SARS-CoV-2 VOCs that have appeared over time. The Alpha and Beta variants was first identified in September 2020 with the Alpha variant was identified in the United Kingdom and the Beta variant was identified in South Africa (Flores-Vega *et al.*, 2022). The Delta variant was identified in October 2020 in India and the Gamma variant was identified in November 2020 in Japan and Brazil (Flores-Vega *et al.*, 2022). The Omicron variant was identified on the 11 November 2021 in Botswana and South Africa (Flores-Vega *et al.*, 2022).

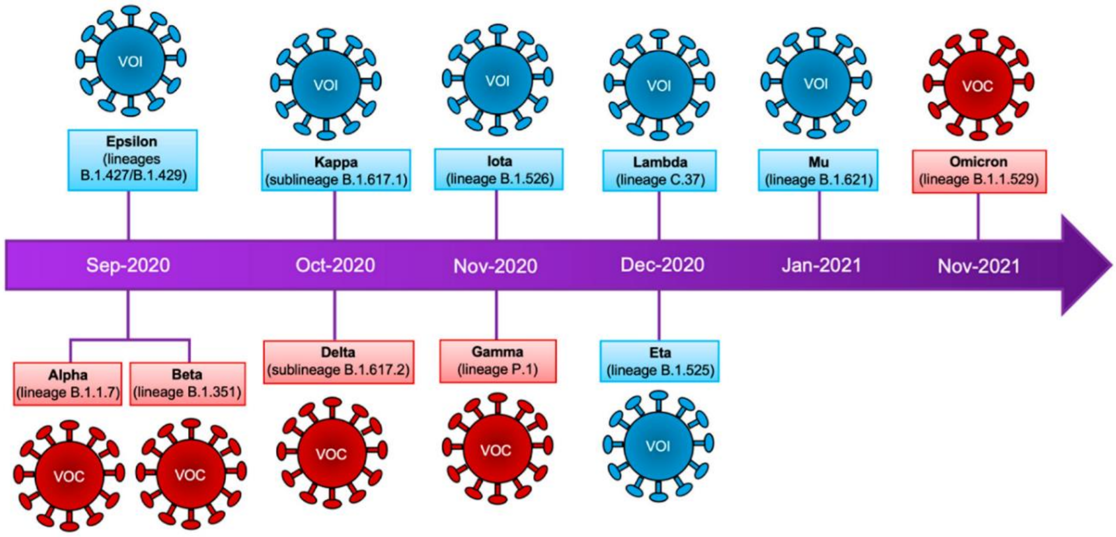


Figure 1.4 Overview of the different SARS-CoV-2 variants of concern or interest (VOC [red] or VOI [blue]) and the dates that the variants were detected (Flores-Vega *et al.*, 2022).

In December 2020, the SARS-CoV-2 Beta variant was identified in South Africa (Tegally *et al.*, 2021c). This variant has been associated with increased transmissibility and has three mutations (K417N, E484K and N501Y) at the key sites in the receptor binding domain (RBD) of the S protein (Tegally *et al.*, 2020). These mutation in the RBD allow the virus to escape neutralization by neutralizing antibodies (Wibmer *et al.*, 2021). There have been several cases in which this has caused a reduction in current vaccines efficiency due to the mutations in the S proteins (Zhou *et al.*, 2021). Despite the reduction in vaccine efficacy caused by the , some vaccines are still able to fully neutralize the virus (Madhi *et al.*, 2021). The Alpha variant originating in the United Kingdom (UK) and Gamma variant, originating in Brazil, are also of great concern due to greater transmission and immune escape of the virus (Fontanet *et al.*, 2021). The Alpha variant has a mutation (N501Y) in the RBD domain of the S protein that cause a higher transmission rate (Fontanet *et al.*, 2021). The Gamma variant has a set of characteristic mutations (N501Y, E484K and K417T) in the S protein (Fontanet *et al.*, 2021). The Delta variant first described in India in 2020 and drove the 3rd wave of infections in South Africa in May 2021 (Mwangi *et al.*, 2022). SARS-CoV-2 Delta variant has shown a higher transmission rate than previously identified variant (Kumar *et al.*, 2021). The Delta variant has a L452R and E484 mutation that increases the binding energy between the S protein and human Angiotensin-converting enzyme 2 (hACE2) (Kumar *et al.*, 2021).

A VOI of SARS-CoV-2 is Epsilon, identified in the United States of America (USA), which has displayed immune evasion properties (McCallum *et al.*, 2021). This strain has three mutations, a S131, W152C and L452 which cause a reduction of neutralization potency from vaccine-elicited plasma (McCallum *et al.*, 2021). The emergence of these variants has highlighted the need for the development of a vaccine that targets the more conserved regions of SARS-CoV-2 in order to prevent the virus from escaping the immune system by mutating (Lawton, 2021). The two lineages responsible for hospitalization in South Africa were the Omicron lineages of BA.1 and BA.2 (Islam *et al.*, 2022). The BA.3 lineage of Omicron had a low number of infections and low spread compared to BA.1 and BA.2 (Desingu *et al.*, 2022).Two new lineages of the Omicron variant called BA.4 and BA.5 have led to a resurgence of COVD-19 cases in South Africa rapidly replacing the BA.2 lineage (Mohapatra *et al.*, 2022). The BA.4 and BA.5 lineages have changes in the L452R and F486V of the S-protein RBD and has led to increased hospitalizations (Mohapatra *et al.*, 2022).

1.10 M and E proteins as vaccines candidates

The alternative surface proteins such as M and E proteins were considered as potential targets for SARS-CoV-2 vaccines (Dai and Gao, 2021). However, M and E proteins were initially described as poorly immunogenic for induction of humoral responses; presumably due to their small ectodomains (Dai and Gao, 2021). Higher sequence identity between these proteins when compared to the S protein suggests that M and E proteins can potentially be used as vaccine targets that stimulate cross-reactive T cell immune responses (Dai and Gao, 2021). Ayyagari *et al.* (2020) explored the use of the SARS-CoV-2 M protein on its own to develop vaccines that induce T-cell responses. For this purpose they used an immunoinformatic approach to design a epitope vaccine which included both B and T-cell epitopes (Ayyagari *et al.*, 2020). New evidence suggest that the E protein is reportedly highly antigenic due to the fact that the E protein is a more conserved outer surface protein which is also easier for produce using molecular recombinant technology due to its small size (Bhattacharya *et al.*, 2021).

Rahaman *et al.* (2020b) explored M and E proteins in combination for developing peptide vaccines. They showed that epitopes in the receptor-binding domain (RBD) and the N-terminal domains (NTD) of S protein in combination with M and E proteins has potential antigenicity and are non-allergenic warranting further assessment for use as vaccines (Rahman *et al.*, 2020b). The epitopes of E, M and N proteins are well characterised (Lu *et al.*, 2021). However, Lu *et al.* (2020) successfully predicted SARS-CoV-2 epitopes based on the 3D structure of the proteins. Both B and T cell epitopes have been identified on the M protein of the SARS-CoV (type 1) strain from the early 2000's (Liu *et al.*, 2010). The M and E proteins are also target for antivirals which aim to block virus assembly and siRNA which interfere with the expression of various SARS-CoV-2 proteins are being evaluated in preclinical trails (Saxena, 2020). The emergence of new SARS-CoV-2 strains with increased fitness in various countries around the world, due to mutations in variable regions of the S protein contribute to the varying efficacy of vaccines. Therefore, evidence that M and E proteins may potentially be good vaccine candidates due to their conserved and immunogenic nature warrants further investigation. In addition, besides S, M and E proteins' transmembrane domains may also be exposed to the immune system during infection therefore M/E specific antibodies may also contribute to the blocking S binding the host cell receptor human angiotensin-converting enzyme 2 (hACE2). Describing the molecular epidemiology of SARS-CoV-2 M and E genes will help us understand the evolution and impact of mutations on antigen and immune escape. To date the immunogenicity of candidate SARS-CoV-2 M/E gene vaccine constructs have only been simulated *in silico* using the C-ImmSim tool (Rapin *et al.*, 2010, Rahman *et al.*, 2020b). There is no information regarding if the mutations in the M and E protein play any role in immune evasion by SARS-CoV-2.

Study rationale

The SARS-CoV-2 pandemic has had a global effect on public health and the global economy. Despite the development of vaccines against SARS-CoV-2, the rise of mutations in the vaccine target, the spike protein, have impacted the overall effectiveness of the vaccines developed against SARS-CoV-2 (Choi and Smith, 2021). The resulting mutations have led to the emergence of VOCs such as the Beta, Delta and Omicron that are responsible for the resurgent waves (Flores-Vega *et al.*, 2022).

The other surface glycoproteins such as the membrane and envelope proteins are more conserved and immunogenic in nature make these proteins, alternative vaccine targets (Dai and Gao, 2021). The use of the M and E proteins as a vaccine candidates has only been simulated *in silico* and no information on the effects of mutations in these proteins on the SARS-CoV-2 ability to evade the immune system (Rahman *et al.*, 2020b).

This study focussed on South African SARS-CoV-2 M and E gene sequences to explore characteristics which will support the use of these proteins as vaccine candidates.

Aim of the study

This study aims to describe the molecular epidemiology of M and E genes and the impact of viral evolution over time on associated immune epitopes from South African SARS-CoV-2 strains, 2020 to 2021.

Specific Objectives

1. To describe the molecular epidemiology of SARS-CoV-2 M/E gene sequences from South Africa from 2020 to 2021.
2. To describe the impact of genetic mutations over time on the antigenic properties of translated SARS-CoV-2 M/E proteins from South Africa

CHAPTER 2

MATERIALS AND METHODS

2.1 M and E protein sequence data

The M and E proteins sequences were extracted from South African SARS-CoV-2 whole genome sequence available on the Global Initiative on Sharing All Influenza Data (GISAID, https://www.gisaid.org/, accessed 08/01/2021) for the period 01 March 2020 to 31 December 2021. We focussed on M and E genes from the Beta, Delta and Omicron variants that dominated the 2nd (epidemiological [epi] week 43 in 2020 to epi week 15 in 2021), 3rd (epi week 26 to 47 in 2021) and 4th (epi week 46 to 53 in 2021) waves of infections in South Africa.

The M and E gene sequences were extracted from South African SARS-CoV-2 genomes (n = 33,347) were aligned using Nextclade (https://clades.nextstrain.org/) online tool that embeds the NextAlign programme for multiple sequence alignments. Sequence dataset were then cleaned by removing duplicated genome data to obtain a dataset of 26,734 genomes.

The M and E genes were extracted from the South African SARS-CoV-2 sequences by selecting the known sequence range of the M gene at nucleotide positions 26523 to 27191 and the E gene at nucleotide positions 26245 to 26472 after further selection for the most complete sequences for total gene sizes of 669 bp and 228 bp respectively. The respective Beta, Delta and Omicron sequences were then combined into a single file for the M gene and a single file for the E gene.

2.2 Phylogenetics analysis of M and E genes

The Wuhan reference strain (Accession number NC_055512) was used as root or master sequence for all analyses. Due to the size of the South African dataset, representational data from Gauteng Province was used for phylogenetic tree construction using MEGA 6 for both the M and E genes. A reference dataset of non-South African SARS-CoV-2 genomes provided by Dr Cornelius Roemer was downloaded from GitHub (https://github.com/nextstrain/nextclade_data/blob/master/data/datasets/sars-cov-2/reference/MN908947/versions/2022-07-11T12:00:00Z/files/sequences.fasta; accessed 09092022). The identical sequences among all the South African strains were removed leaving a total number of 138 unique E gene sequences and 560 unique M gene sequences. The sequences were used to produced phylogenetic trees using the MEGA 6 software. Maximum-likelihood trees were constructed using the Hasegawa-Kishino-Yano model with 100 bootstraps to determine confidence level for clusters. Partial deletions were used for missing or gaps in data with a site coverage cut-off of 95%. The Wuhan reference strain (NC 045512.2:26523-27191) and EPI_ISL_402124-E were selected as root.

The HIV sequence database Highlighter tool was used to produce blots showing the mismatch and silent/non-silent mutations in the sequences (https://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/highlighter_top.html). The mismatches and silent/non-silent mutations blots were obtained with sequences sorted by similarity, sequence type set to nucleotides and the rest of the settings kept and default. The images produced were then cut and shrunk down to fit into the page and remerge into one image.

2.3 Identification of immune epitopes using Bioinformatics approaches

Allergenicity of the epitopes were determined with the online AllerTOPv.2.0 tool with proteins classified by the k- nearest neighbour algorithm (kNN, k=1) (https://www.ddg-pharmfac.net/AllerTOP/index.html) and toxicity were predicted using the ToxinPred tool (https://webs.iiitd.edu.in/raghava/toxinpred/algo.php). Predicted toxins were defined as having a support vector machine (SVM) threshold value of ≥ 0.00 and an expected (E) value cut-off of 10. The ToxinPred2 tool (https://webs.iiitd.edu.in/raghava/toxinpred2/) was used to assess toxin properties for the whole M and E proteins.

The antigenicity of the epitopes was determined using VaxiJen v.20 with default settings which was set to virus and sequence output (http://www.ddg-pharmfac.net/vaxijen/VaxiJen.html). The threshold of probable antigen was defined as a score ≥ 0.4.

Major Histocompatibility Complex (MHC) class-I epitopes were examined with NetCTL, a web based tool that predicts the (Cytotoxic T Lymphocytes) CTL epitopes on protein sequences (https://services.healthtech.dtu.dk/service.php?NetCTL-1.2) (Larsen *et al.*, 2007). The NetTepi-1.0 online tool was used to predict T-cell epitopes restricted by the most prevalent HLA-A and HLA-B receptors which included A-01:01, -02:01, -03:01, -11:01, -24:02, and -

26:01) and B-07:02, -15:01, -27:05, -35:01, -39:01, -40:01, and -58:01. A rank of 50 and above predicted an epitope (https://services.healthtech.dtu.dk/service.php?NetTepi-1.0)) (Trolle and Nielsen, 2014). The 138 unique E gene sequences and 560 unique M gene sequences was used for this analysis.

Identification of B-cell epitopes

The B-cell epitopes were identified using the BepiPred 2.0 tool (http://www.cbs.dtu/services/BepiPred/). The default settings were used to run the analysis and cut-off values for determining epitopes were defined as probability scores higher than 0.5. Major Histocompatibility Complex (MHC) class-I epitopes were examined with NetCTL, a web based tool that predicts the (Cytotoxic T Lymphocytes) CTL epitopes on protein sequences (https://services.healthtech.dtu.dk/service.php?NetCTL-1.2) (Larsen *et al.*, 2007). The NetTepi-1.0 online tool was used to predict T-cell epitopes restricted by the most prevalent HLA-A and HLA-B receptors which included A-01:01, -02:01, -03:01, -11:01, -24:02, and -26:01) and B-07:02, -15:01, -27:05, -35:01, -39:01, -40:01, and -58:01. A rank of 50 and above predicted an epitope (https://services.healthtech.dtu.dk/service.php?NetTepi-1.0) (Trolle and Nielsen, 2014). The 138 unique E gene sequences and 560 unique M gene sequences was used for this analysis.

The Mosaic vaccine tool suite was used to determine the positional CTL epitope coverage assessment for SARS-CoV-2 M and E sequences to show how much of the respective protein's length are covered by the derived mosaic peptide cocktails (https://www.hiv.lanl.gov/content/sequence/MOSAIC/). The images of the 9-mer coverage by position and unique 9-mer counts images were downloaded. For these tools identical sequences were removed, errors, gaps and non-specific nucleotides were removed, and sequence names shortened to acceptable lengths.

2.4 Ethics approval

Study protocol was submitted for ethical review and an ethics waiver (reference number W-CBP-210813-01) was obtained.

CHAPTER 3

RESULTS

3. Molecular epidemiology of SARS-CoV-2 M/E gene sequences from South Africa, 2020 to 2021.

3.1 Phylogenetic analysis of SARS-CoV-2 E

For the study period March 2020 to December 2021 SARS-CoV-2 31781 E gene sequences were extracted on GISAID from a total of 33 347 genomes from South Africa. The final 26,734 individual SARS-CoV-2 genomes after final selection of the most complete M and E gene sequences was grouped as a study dataset on GISAID with study identifier EPI_SET_221130zm which can be used to extract the dataset at any time (or by using the link doi: 10.55876/gis8.221130zm).

The maximum-likelihood phylogenetic tree for the E gene sequences (Figure 3.1) shows a distinct Omicron cluster that is separated from the Wuhan, Beta and Delta variants. Most Omicron sequences forms a distinct cluster with 50% bootstrap support. Similarly, most Beta sequences also forms a cluster with less than 50% bootstrap support. This shows the highly conserved nature of the E gene among SARS-CoV-2 strains. All the Omicron sequences have a cytosine (C) to thymine (T) mutation in gene position 26 which translates to a threonine to isoleucine amino acid change at position 9 of the E protein.

3.2. Mismatch, silent and non-silent mutation analysis of SARS-CoV-2 E genes

For the analysis of the mutations, the sequences were cleaned by removing all identical sequences leaving a total number of 138 unique E gene sequences. The mismatch bases for the E genes compared to the Wuhan reference strain's sequence are show in figure 3.2. The E gene has numerous mismatches in the Beta, Delta and Omicron variants compared to the reference sequence. The Omicron variant has a consistent mutation of ACA to ATA in the sequences which is a threonine to tyrosine mutation at position 26. The consistent change in the Beta variant is identified in the sequences to be a CCG to CTG at position 217 which is a proline to leucine mutation at amino acid position 72 (shown by the red circles).

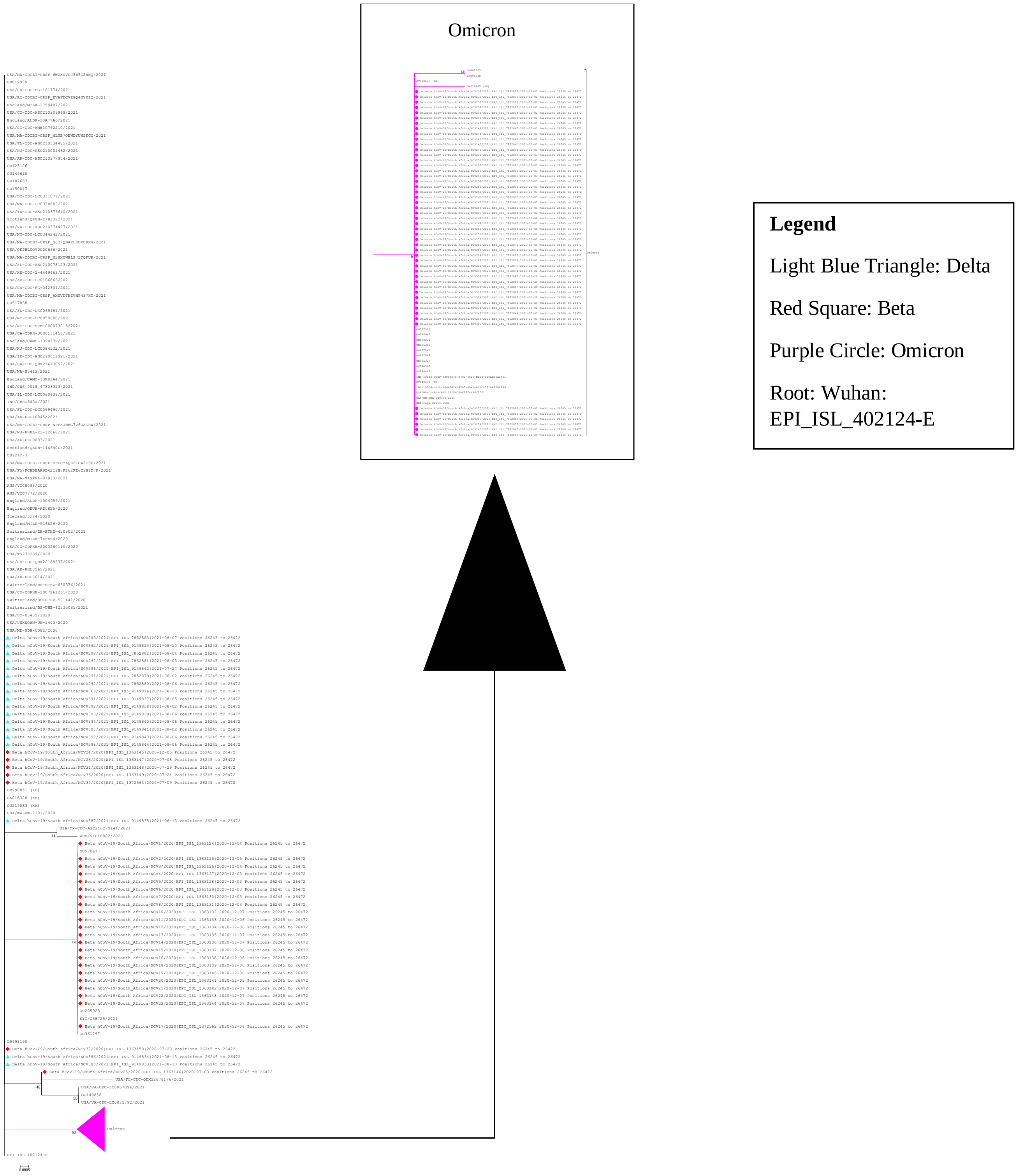


Figure 3.1. Phylogenetic tree of the E genes of SARS-CoV-2**.** The tree was constructed using a maximum likelihood tree with a bootstrap value of 100 and the Hasegawa-Kishino-Yano model. The E gene sequences were 228 bp in length and the Wuhan reference strain's sequence was used as the root of the tree.

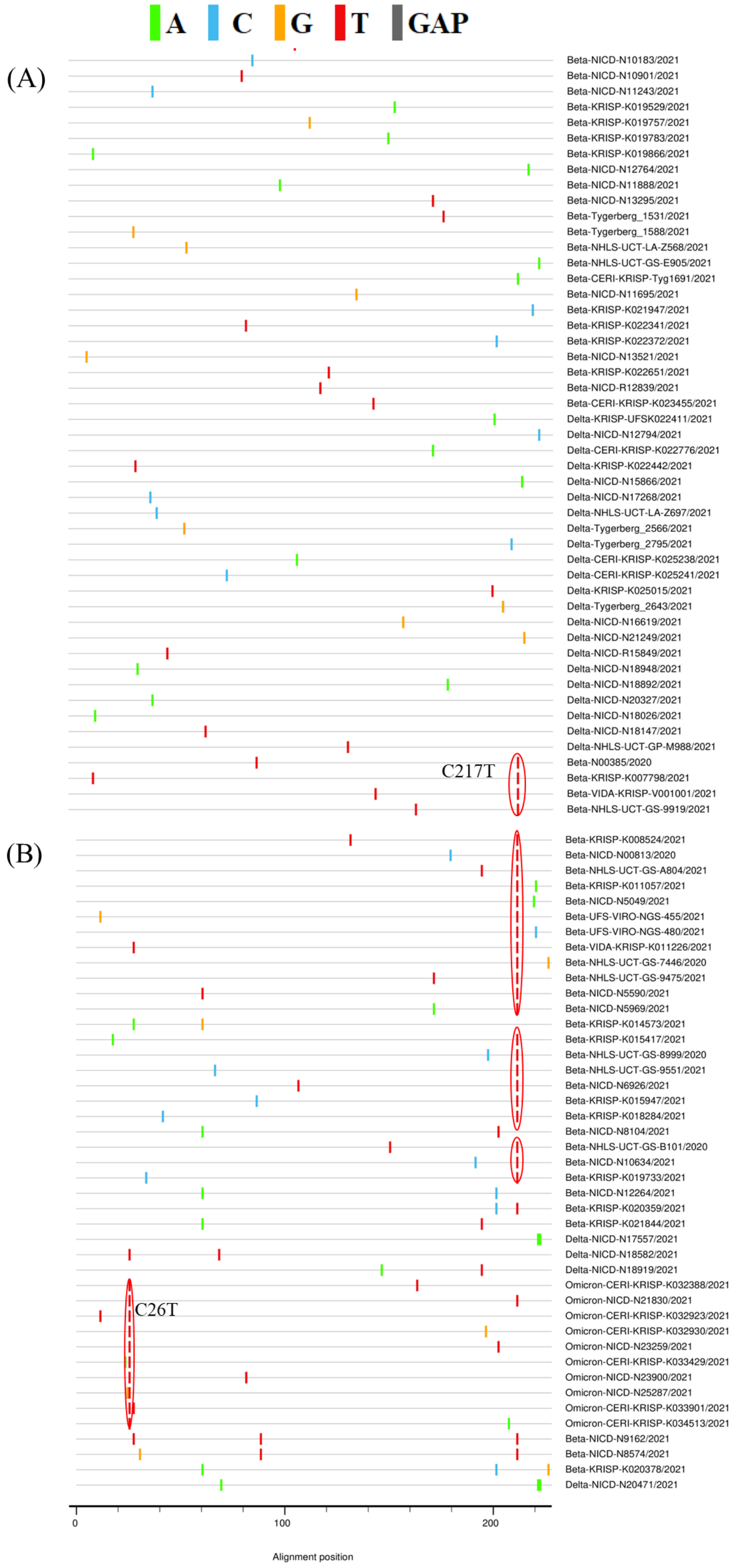


Figure 3.2 Omicron sequences compared to the Wuhan reference strain's sequence. Nucleotide changes in the sequences are shown by the colored ticks. C= cytosine and T= thymine.

The non-silent mutations mirror the mismatch mutations of the E gene with the consistent non-silent mutation circled in red (Figure 3.3). The Beta variant have a non-silent C217T mutation, and the Omicron variant have a C26T non-silent mutation. This shows that the consistent mutations have an effect on the overall structure of the protein due to being non-silent mutations.

3.3. Phylogenetic analysis of SARS-CoV-2 M genes

A total of 28983 M gene sequences were extracted for analysis from a total of 33 347 SARS-CoV-2 genomes from South Africa available on GISAID covering the period 01 March 2020 to 31 December 2021. The M gene maximum-likelihood phylogenetic tree (Figure 3.4) shows a distinct cluster for the Omicron sequences although with weak (19%) bootstrap support. The Delta cluster is characterised by an T to C mutation at gene position 245 which translates as an isoleucine to threonine change at position 82 of the M peptide sequence when compared to the Wuhan reference strain’s sequence. This mutation is also present in some Beta variant sequences.

The Omicron M gene sequences haves three mutations that differentiate them from the Wuhan reference strain: the first mutation is a adenine (A) to guanine (G) nucleotide change at positions 8 which translate to a aspartic acid to glycine amino acid mutation at position 3 of the protein, a second mutation of C to G at positions 55 translates to a glutamine to glutamic acid mutation at position 19 and thirdly a G to A mutation at positions 187 which results in a alanine to threonine amino acid change at position 63. There is one Omicron sequence (NCV262/2021) which is an outlier as it sits on the same branch as the Beta variant sequences. This indicates that this M gene sequence does not have the distinctive Omicron specific mutations.

3.4. Mismatch, silent and non-silent mutation analysis of SARS-CoV-2 M

The mismatch bases for the M genes compared to the Wuhan reference gene are show in the figure below. Figure 3.5 shows mismatches among 560 M gene sequences. Beta and Delta variants share a common C to T nucleotide mutation which results in an isoleucine to threonine mutation. The Omicron VOC have three consistent mutations found in the sequence of A8G,

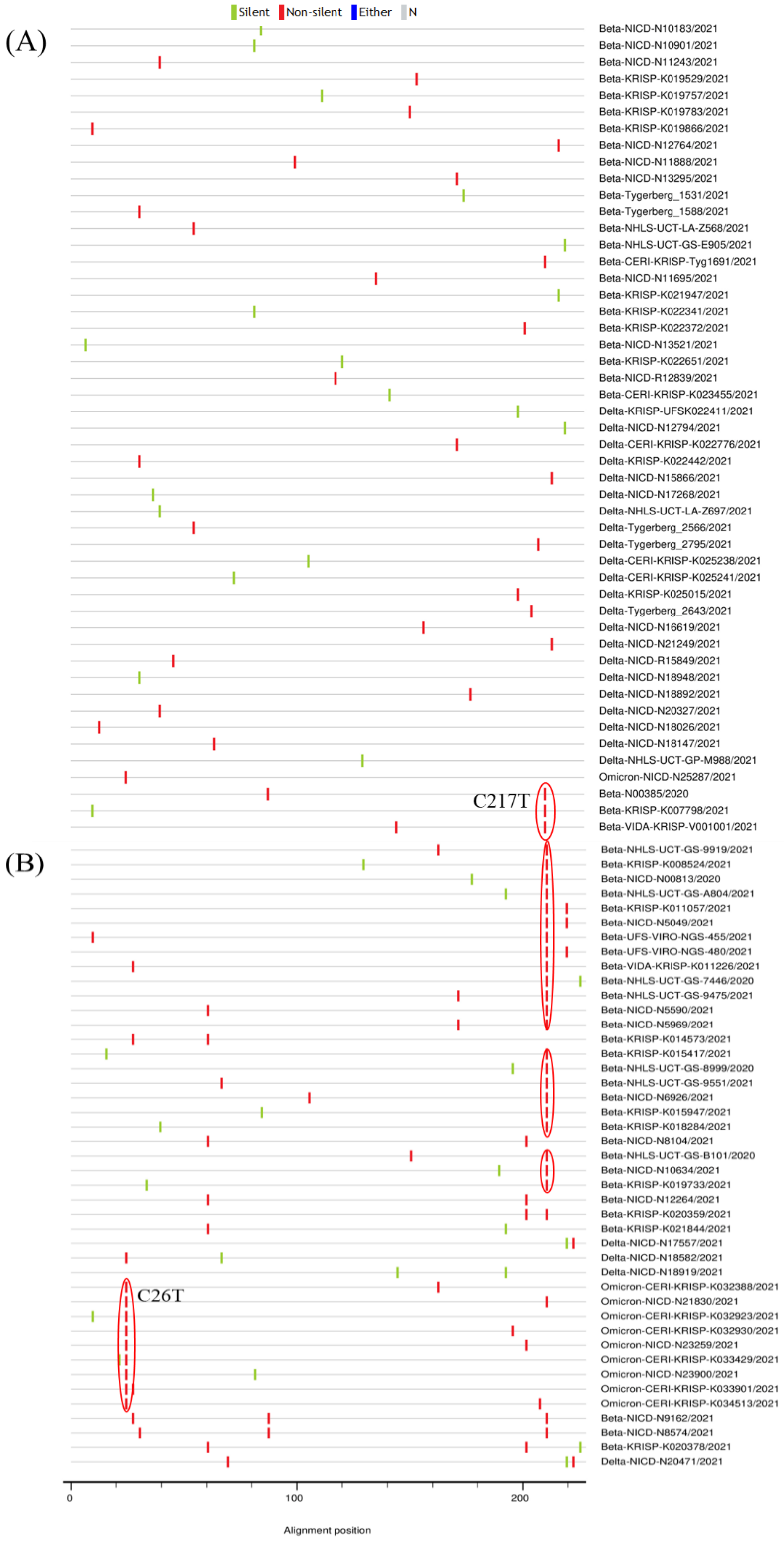


Figure 3.3 Figure 3.6. Highlighter plots providing a graphical representation of silent and non-silent mutations across the M genes of SARS-CoV-2 variants (A: Beta; B: Beta and Delta; C: Delta and Omicron) compared to the Wuhan reference strain's sequence. The silent and non-silent mutations in the sequences are shown by the colored ticks. N= unresolved amino acid position, T= thymine; C= cytosine, G= guanine, A= adenosine. The silent and non-silent mutations in the sequences are shown by the colored ticks. The exact change and position are shown in the images. N= unresolved amino acid position, C= cytosine and T= thymine.

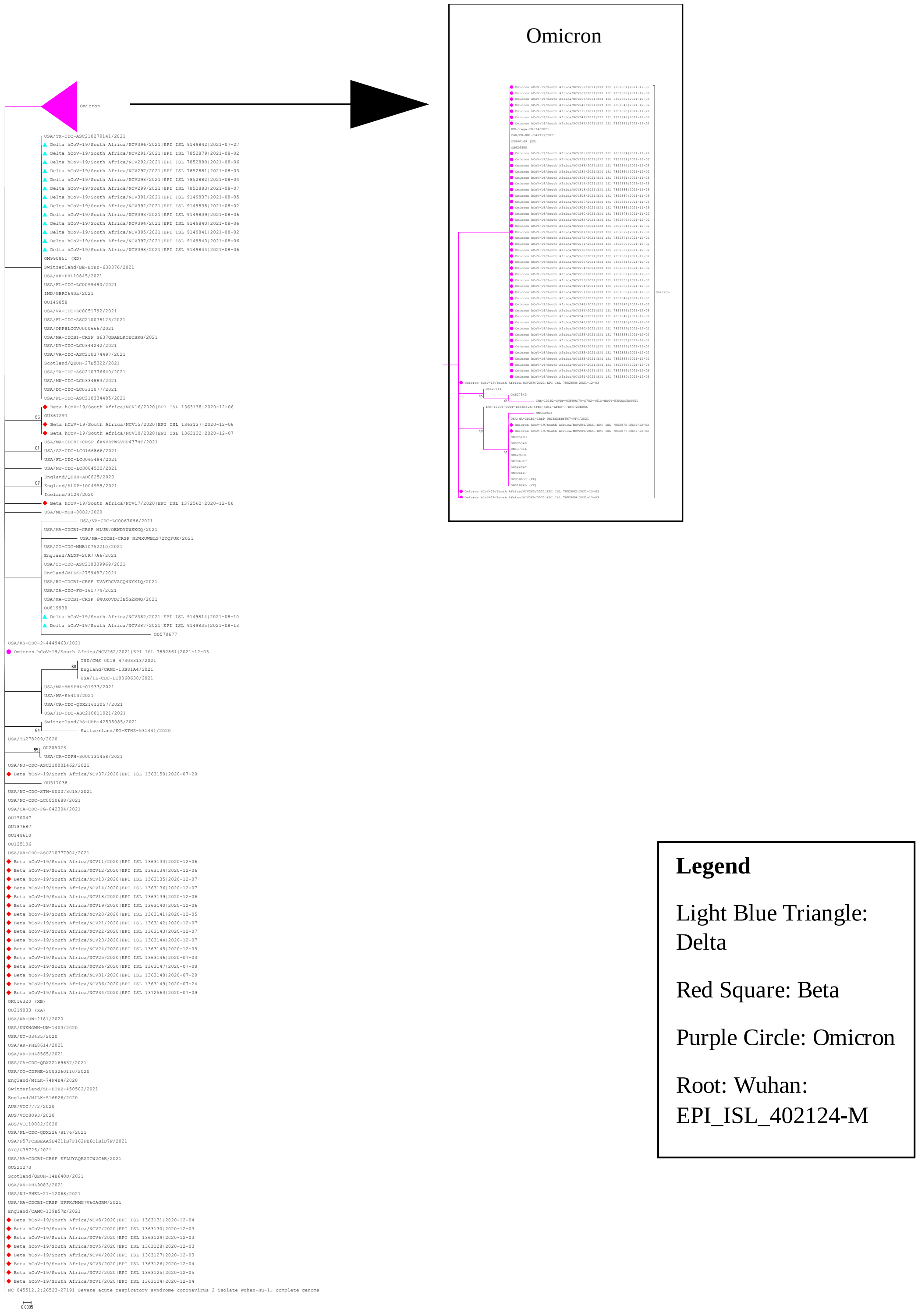


Figure 3.4 Phylogenetic tree of the M gene of SARS-CoV-2. The tree was constructed using a maximum likelihood tree with a bootstrap value of 100 and the Hasegawa-Kishino-Yano model. The M gene sequences were 669 bp in length and the Wuhan reference strain's sequence was used as the root of the tree.

C55G and A187G which led to respective mutations of aspartic acid to glycine, glutamine to glutamic acid and alanine to threonine.

The non-silent mutation in the M gene sequences which are circled red (Figure 3.6) mirrors the mismatch mutation of the M gene as shown in the figure 3.5. This highlights that the consistent nucleotide mutations actually result in amino acid changes. The Beta and Delta VOCs have a T245C non-silent mutation, and the Omicron VOC have A8G, C55G and G187A non-silent mutations.

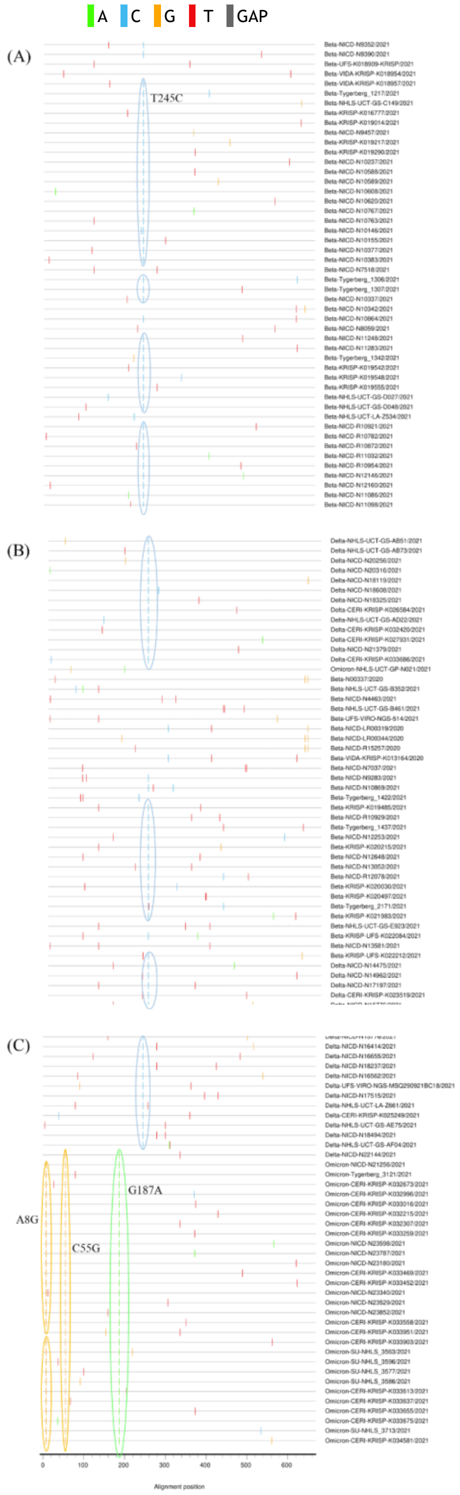


Figure 3.5 Highlighter plots providing a graphical representation of key nucleotide mismatch mutations across the M gene of South African SARS-CoV-2 variants (A: Beta, B: Beta and Delta; C = Delta and Omicron) compared to the Wuhan reference strain. Nucleotide changes in the sequences are shown by the colored ticks. All the consistent mismatch mutations are circled and the nucleotide position and changes are indicated for each (T= thymine; C= cytosine, G= guanine, A= adenosine).

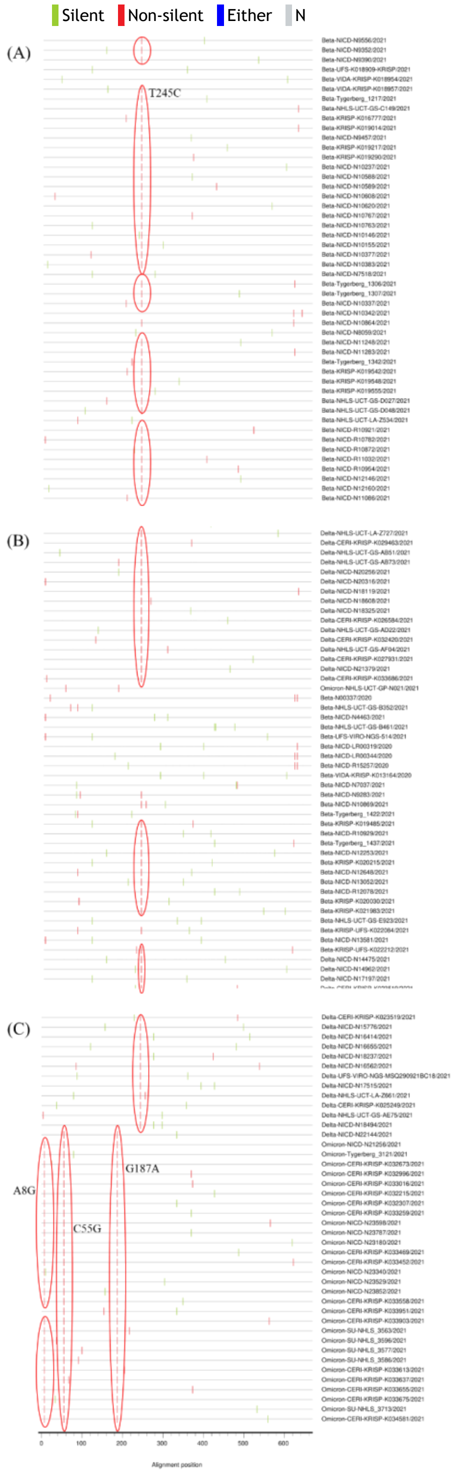


Figure 3.6. Highlighter plots providing a graphical representation of silent and non-silent mutations across the M genes of SARS-CoV-2 variants (A: Beta; B: Beta and Delta; C: Delta and Omicron) compared to the Wuhan reference strain's sequence. The silent and non-silent mutations in the sequences are shown by the colored ticks. N= unresolved amino acid position, T= thymine; C= cytosine, G= guanine, A= adenosine.

CHAPTER 4

RESULTS

4. Antigenicity of translated SARS-CoV-2 M/E proteins from South African strains, 2020 to 2021

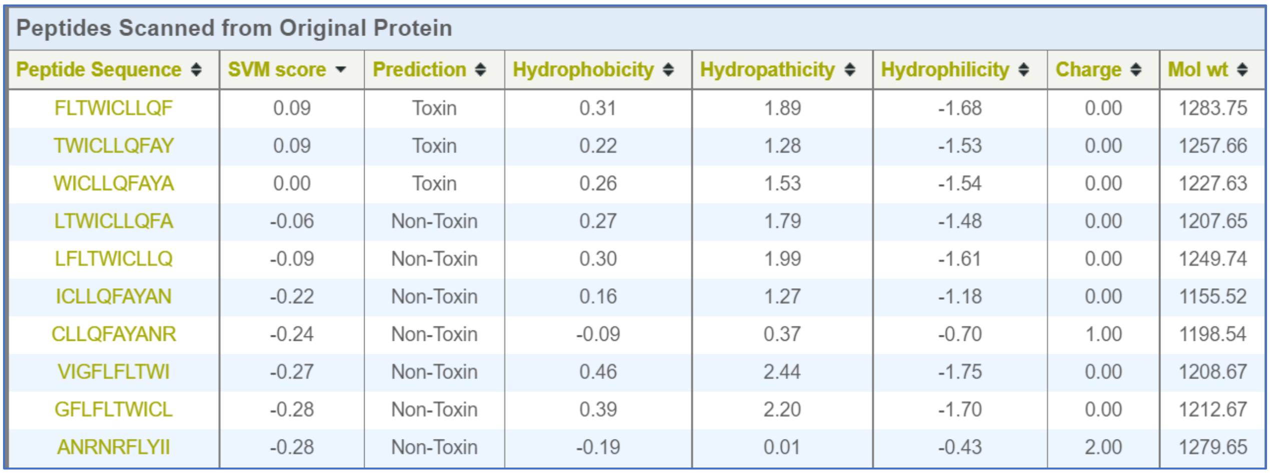
4.1 Allergenicity and toxicity of M and E proteins

The AllerTop result for the Wuhan reference strain and 560 South African SARS-CoV-2 variants' M protein sequences showed that the protein is defined as a non-allergen with the nearest protein being the voltage-dependent L-type calcium channel subunit alpha-1S in humans (https://www.uniprot.org/uniprotkb/Q13698/entry) and Cation transporter HKT2;2 (https://www.uniprot.org/uniprotkb/Q93XI5/entry) from Oryza sativa subsp. indica (Rice) specifically nearest for Omicron protein sequences. The AllerTop result for E gene sequences from the Wuhan reference strain and 138 South African SARS-CoV-2 variants showed that the proteins are also defined as a non-allergens with the nearest proteins being either the human C-C chemokine receptor type 1 (https://www.uniprot.org/uniprotkb/P32246/entry) or the human ATP synthase subunit a (P00846, UniProt Database, https://www.uniprot.org/uniprotkb/P00846/entry) which is a transmembrane protein specifically nearest for Omicron sequences.

The ToxinPred tool identified specific epitope regions in the M and E proteins of SARS-CoV-2 which have toxic properties. In the M protein at least 3 epitopes with toxin properties (positive Support vector machine (SVM) values) and 7 with non-toxin properties were identified (Table 4.1). The M epitopes with toxic properties are in the region FLTWICLLQFAYA which is located at amino acid positions 28 to 40. Using peptides that is longer or shorter than the peptides shown in table 4.1 may resolve some of the predicted toxic properties. The whole M and E proteins were all identified as non-toxins.

In the E protein 9 overlapping epitopes were identified as toxic within the ALRLCAYCCNIVNIVNVSLV epitopic region at positions 36 to 51 (Table 4.2).

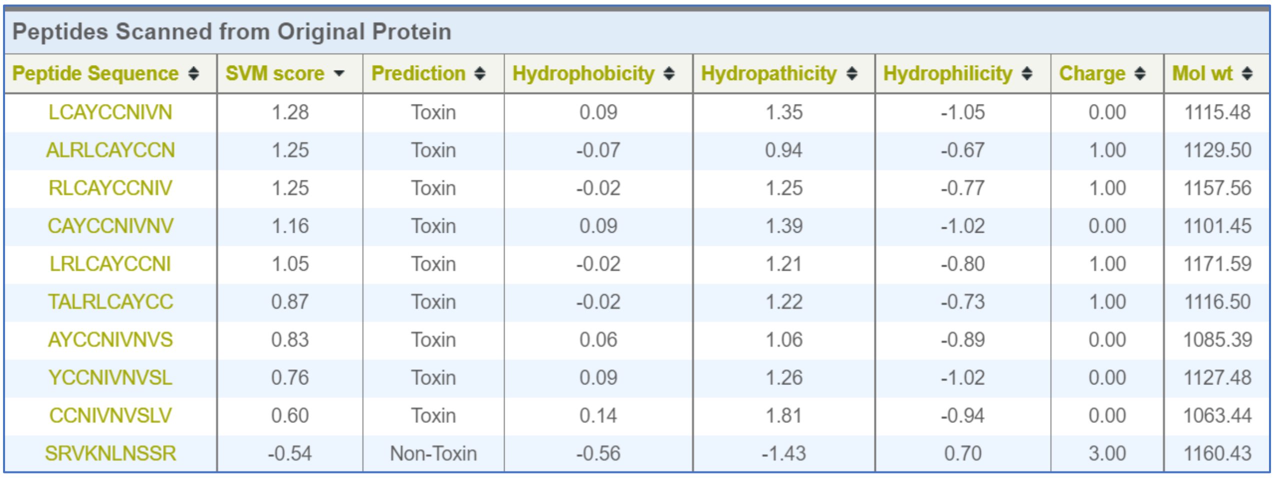
Table 4.1 The toxic regions in the SARS-CoV-2 M peptide from Toxinpred. The regions with a positive SVM score are defined as toxins.



Peptides Scanned from Original Protein

Peptide Sequence	SVM score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
FLTWICLLQF	0.09	Toxin	0.31	1.89	-1.68	0.00	1283.75
TWICLLQFAY	0.09	Toxin	0.22	1.28	-1.53	0.00	1257.66
WICLLQFAYA	0.00	Toxin	0.26	1.53	-1.54	0.00	1227.63
LTWICLLQFA	-0.06	Non-Toxin	0.27	1.79	-1.48	0.00	1207.65
LFLTWICLLQ	-0.09	Non-Toxin	0.30	1.99	-1.61	0.00	1249.74
ICLLQFAYAN	-0.22	Non-Toxin	0.16	1.27	-1.18	0.00	1155.52
CLLQFAYANR	-0.24	Non-Toxin	-0.09	0.37	-0.70	1.00	1198.54
VIGFLFLTWI	-0.27	Non-Toxin	0.46	2.44	-1.75	0.00	1208.67
GFLFLTWICL	-0.28	Non-Toxin	0.39	2.20	-1.70	0.00	1212.67
ANRNRFLYII	-0.28	Non-Toxin	-0.19	0.01	-0.43	2.00	1279.65

Table 4.2 The toxic regions in the SARS-CoV-2 E peptide from Toxinpred. The regions with a positive SVM score are defined as toxins.



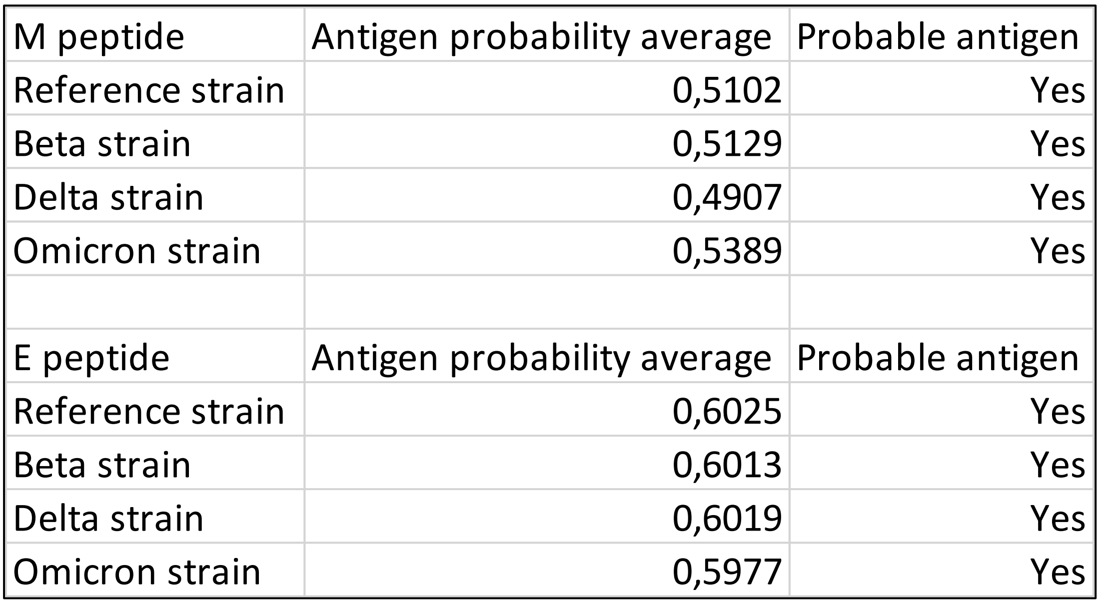
Peptides Scanned from Original Protein

Peptide Sequence	SVM score	Prediction	Hydrophobicity	Hydropathicity	Hydrophilicity	Charge	Mol wt
LCAYCCNIVN	1.28	Toxin	0.09	1.35	-1.05	0.00	1115.48
ALRLCAYCCN	1.25	Toxin	-0.07	0.94	-0.67	1.00	1129.50
RLCAYCCNIV	1.25	Toxin	-0.02	1.25	-0.77	1.00	1157.56
CAYCCNIVNV	1.16	Toxin	0.09	1.39	-1.02	0.00	1101.45
LRLCAYCCNI	1.05	Toxin	-0.02	1.21	-0.80	1.00	1171.59
TALRLCAYCC	0.87	Toxin	-0.02	1.22	-0.73	1.00	1116.50
AYCCNIVNVS	0.83	Toxin	0.06	1.06	-0.89	0.00	1085.39
YCCNIVNVSL	0.76	Toxin	0.09	1.26	-1.02	0.00	1127.48
CCNIVNVSLV	0.60	Toxin	0.14	1.81	-0.94	0.00	1063.44
SRVKNLNSSR	-0.54	Non-Toxin	-0.56	-1.43	0.70	3.00	1160.43

4.2 Probability of antigens

Table 4.3 shows the likelihood of M and E proteins being recognised as antigens as determined using the VaxiJen tool which does not consider the alignment of proteins but their physicochemical properties which is dependent on principal amino acids. The Wuhan reference strain and 560 South African SARS-CoV-2 variants' M protein sequences were identified as probable protective antigens as indicated by probability scores above 0.4. The average probability for the M peptides of Beta variants are slightly higher than the M peptide of the reference strain. The M peptides of the Delta variants had the lowest average probability score whereas the average probability score for Omicron variants was the highest. Similarly, the Wuhan reference strain and all 138 South African SARS-CoV-2 E protein sequences were

Table 4.3 The probability of SARS-CoV-2 M and E peptide being Antigens obtain from the VaxiJen database. Probability scores above 0.4 indicate potential (probable) antigens.



M peptide	Antigen probability average	Probable antigen
Reference strain	0,5102	Yes
Beta strain	0,5129	Yes
Delta strain	0,4907	Yes
Omicron strain	0,5389	Yes
E peptide	Antigen probability average	Probable antigen
Reference strain	0,6025	Yes
Beta strain	0,6013	Yes
Delta strain	0,6019	Yes
Omicron strain	0,5977	Yes

identified as probable protective antigens. The average probability scores of the E peptide were higher than for M peptides with Omicron variants displaying the lowest score and the reference strain displaying the highest score.

4.3.1 Mutations associated with immune escape in M protein

For the analysis of mutations in predicted CTL epitopes a total of 560 unique M gene sequences were used. The T cell epitopes of the M protein sequences were summarised into the tables 4.4A to 4.4E and a total of 26 epitopes were identified across the M protein sequence. The resulting HLA-A and HLA-B epitopes identified from Wuhan, Beta, Delta and Omicron sequences were combined in tables. The identified epitopes restricted by HLA-A (01:01, 02:01, 03:01, 11:01, 24:02, 26:01) and HLA-B (07:02, 15:01, 27:05, 35:01, 39:01, 40:01, 58:01) are summarised tables 4.4A to 4.4E.

In table 4.4A some Beta, Delta and Omicron variants retained the wildtype sequence in M protein peptide epitopes $_{1}$MADSNGTIT/VEELKK, $_{13}$K/KLLEQWNL/VIGF, $_{21}$LVIGFLFLTW/ICLLQF and $_{30}$WICLLQFAY. CTL escape mutations was observed at several positions across the overlapping peptides. Six epitopes with escape mutations shown in table 4.4B are: $_{34}$LQFAYANRN/FLY, $_{40}$NRNRFLYII/KLIFLW, $_{49}$KLIFLWLLW/PVTL, $_{54}$WLLWPVTLA/CFV, $_{60}$TLACFVLAA/V and $_{64}$FVLAAVYRI/NW. Table 4.4C summarises the following five CTL epitopes with escape mutations: $_{70}$YRINWITGG/I, $_{75}$ITGG/IAIAMA/CLV, $_{81}$IAMACLVGL/MW, $_{85}$CLVGLMWLS/YFIA, $_{91}$WLSY/FIASF/RLFAR. One mutation resulted in a potential N-linked glycosylation gain for the peptide sequence starting at amino acid position 115 in a Delta variant (Table 4.4D; $_{115}$TNILL**NVPL** to $_{115}$TNILL**NVTL).** HLA-B58:01 was identified as the HLA type that restricted most T cell epitopes identified in this study.

B cell epitopes identified across the M protein sequences are summarised in table 4.5. Twelve potential B cell epitopes were identified. One B epitope region are linked to position 5, 6, 7 and 9 where positions 5 to 8 (NGTI) is a predicted N-linked glycosylation site. Epitopes were also identified at positions 11-19 (EELKKLLEQ), 134-137 (LESE), 180-189 (KLGASQRVAG), 190-207 (DSGFAAYSRYRIGNYKLN) and 209-217 (DHSSSSDNIA). The positions of 71, 160,165 and 166 showed that the Beta and Omicron VOCs share the same predicted epitope at those positions.

Table 4.4A The HLA restrictions of the M 9-mer peptide epitopes detected by NetTepi. The changes in the sequences are in bold. Epitopes starting at amino acid positions 1 to 30.

Protein amino acid position at start of each epitope		**Epitopes**	**Restriction**
1	Wuhan, Beta, Delta Omicron	MADSNGTIT/VEELKK	HLA-A11:01
	Beta, Delta	M**(S/T)** DSNGTIT/VEELKK	HLA-A01:01
	Delta	MA**H**SNGTITV/VEELKK	HLA-B39:01
	Beta	MADSN(**V**/**C**) TIT/VEELKK	HLA-A11:01
	Beta	MADSNGTIT/VEEL**R**K	HLA-A11:01
	Beta	MADSNGT**(L/I)** T/VEELKK	HLA-A11:01, HLA-A03:01
	Beta	MADSNGTIT/(**I/A**) EELKK	HLA-A11:04
	Beta	MADSNGTIT/E**D**LKK	HLA-A11:01, HLA-A03:01
13	Wuhan, Beta, Delta, Omicron	K/KLLEQWNL/VIGF	HLA-A02:01
	Beta	K/**E**LLEQWNL/VIGF	HLA-B40:01
	Beta	K/KL**I**EQWNL/VIGF	HLA-A02:01
	Beta	K/KL**F**EQWNL/VIGF	HLA-A02:01, HLA-A24:02
	Delta	K/**Q**LLWQWNL/VIGF	HLA-A02:01
	Delta, Omicron	K/KLLE**E**WNL/VIGF	HLA-A02:01, HLA-B40:01
	Beta	K/KLLEQWNL/VIG**L**	HLA-B39:01
21	Wuhan; Beta; Delta; Omicron	LVIGFLFLTW/ICLLQF	HLA-A02:01; HLA-A24:02; HLA-B58:01
	Beta	LVIG**L**LFLTW/ICLLQF	HLA-B58:01
	Beta, Delta	LVIGFLF**F**TW/ICLLQF	HLA-A02:01, HLA-A24:02, HLA-B58:01
	Beta	LVIGF**I**FLTW/ICLLQF	HLA-A02:01, HLA-A24:02, HLA-B58:01
	Delta	LVIGFLFLTW/I**F**LLQF	HLA-A02:01, HLA-A24:02, HLA-B58:01
	Delta	LVIGFLFLT**C**/ICLLQF	HLA-A02:01, HLA-A24:02, HLA-B58:01
30	Wuhan; Beta Delta; Omicron	WICLLQFAY	HLA-A01:01, HLA-B35:01
	Beta	WICLLQ**I**AY	HLA-A01:01
	Beta	**C**ICLLQFAY	HLA-A01:01
	Delta	WI**F**LLQFAY	HLA-A01:01, HLA-B15:01

Table 4.4B The HLA restrictions of the M 9-mer peptide epitopes detected by NetTepi**.** The changes in the sequences are in bold. Epitopes starting at amino acid positions 34 to 64.

Protein amino acid position at start of each epitope		**Epitopes**	**Restriction**
34	Wuhan, Beta, Delta, Omicron	LQFAYANRN/FLY	HLA-A01:01, HLA-A24:02, HLA-B15:01, HLA-B35:01, HLA-B58:01
	Beta	LQ**I**AYANR**K**/FLY	HLA-A1:01, HLA-A24:02, HLA-A11:01, HLA-B58:01, HLA-B35:01
	Beta	LQFAY**S**RNR/FLY	HLA-A01:01, HLA-A11:01, HLA-24:02, HLA-B15:01, HLA-B35:01, HLA-B58:01
40	Wuhan, Beta, Delta, Omicron	NRNRFLYII/KLIFLW	HLA-A24:02, HLA-B15:01, HLA-B27:05, HLA-B39:01, HLA-B58:01
	Beta	NR**K**RFLYII/KLIFLW	HLA-B27:05, HLA-B39:01
	Beta, Delta	NRNRFLYI(**V/T**)/KLIFLW	HLA-A24:02, HLA-B27:05, HLA-B39:01, HLA-B58:01
	Beta	NRNRFLYII/KL**V**FLW	HLA-24:02, HLA-B15:01, HLA-58:01
49	Wuhan, Beta, Delta, Omicron	KLIFLWLLW/PVTL	HLA-A02:01, HLA-A24:02, HLA-B58:01
	Beta	KLIFL**C**LLW/PVTL	HLA-B58:01
	Beta	KL**V**FLCLLW/PVTL	HLA-A02:01, HLA-A24:02
54	Wuhan, Beta, Delta, Omicron	WLLWPVTLA/CFV	HLA-A02:01, HLA-B35:01
	Beta	WLLWPVTL**V**/CFV	HLA-A02:01, HLA-24:02, HLA-B35:01
	Beta, Delta, Omicron	WLLWPVTL**T**/CFV	HLA-A24:02, HLA-B07:02, HLA-B35:01
	Beta	WLLWPVTLA/CF**L**	HLA-B07:02, HLA-B35:01, HLA-B39:01
60	Wuhan, Beta, Delta, Omicron	TLACFVLAA/V	HLA-A02:01
	Beta	TLACF**L**LAA/V	HLA-A02:01
	Beta	TLACFLLAA/**F**	HLA-B15:01, HLA-B35:01, HLA-B58:01
64	Wuhan, Beta, Delta, Omicron	FVLAAVYRI/NW	HLA-A02:01, HLA-B58:01
	Beta	FVLA**S**VYRI/NW	HLA-A02:01, HLA-B58:01
	Beta	FVLAA**F**YRI/NW	HLA-B58:01
	Beta	FVLAA**I**YRI/NW	HLA-A02:01, HLA-B58:01
	Beta	F**L**LAAVYRI/NW	HLA-A02:01, HLA-B58:01
	Beta	FVLAAVYRI/**K**W	HLA-A03:01, HLA-A11:01, HLA-B58:01

Table 4.4C The HLA restrictions of the M 9-mer peptide epitopes detected by NetTepi.

The changes in the sequences are in bold font. Epitopes starting at amino acid positions 70 to 91.

Protein amino acid position at start of each epitope		**Epitopes**	**Restriction**
70	Wuhan, Beta, Delta, Omicron	YRINWITGG/I	HLA-B27:05
	Beta	YRI**K**WITGG/I	HLA-B27:05
	Beta	YRINWI**A**GG/I	HLA-B27:05
	Beta	YRINWIT**C**G/I	HLA-B27:05, HLA-B58:01
	Beta, Delta	YRINWITG**V**/I	HLA-B27:05, HLA-B07:02, HLA-B39:01
	Delta	YRIN**L**ITGG/I	HLA-B27:05
75	Wuhan	ITGG/IAIAMA/CLV	
	Beta	I**A**GGI/AIAMA/CLV	HLA-B35:01
	Beta	ITGG/AIAMA/**F**LV	HLA-B15:01, HLA-B58:01
	Beta	ITGG/AIA**I**A/CLV	HLA-B58:01
	Beta	ITGG/I**S**IAMA/CLV	HLA-B58:01
81	Wuhan, Beta, Delta, Omicron	IAMACLVGL/MW	HLA-B15:01, HLA-B58:01
	Beta	IAMA**F**LVGL/MW	HLA-B15:01, HLA-B35:01, HLA-B58:01
	Beta	I**S**MACLVGL/MW	HLA-B15:01, HLA-B58:01
	Delta	**T**AMACLVG**F**/MW	HLA-B15:01, HLA-B35:01, HLA-B58:01
	Beta	**T**AMA**F**LVGL/MW	HLA-B39:01
	Beta	IAM**S**CLVGL/MW	HLA-B15:01, HLA-B58:01
	Beta	IA**I**ACLVGL/MW	HLA-A26:01, HLA-B58:01
85	Wuhan, Beta, Delta, Omicron	CLVGLMWLS/YFIA	HLA-A01:01, HLA-A02:01, HLA-24:02
	Beta	**F**LVGLMWLS/YFIA	HLA-A02:01
	Delta	FLVG**F**MWLS/YFIA	HLA-A01:01, HLA-A02:01, HLA-B15:01, HLA-B35:01
91	Wuhan, Beta, Delta, Omicron	WLSY/FIASF/RLFAR	HLA-A03:01, HLA-A11:01, HLA-A24:02, HLA-A26:01, HLA-B15:01
	Beta	WLSY/FI**S**SF/RLFAR	HLA-A03:01, HLA-A11:01, HLA-A24:02, HLA-B15:01
	Delta	WLSY/FIASF/RLF**MR**	HLA-A26:01
	Delta	WLSY/FIASF/RLF**V**R	HLA-A02:01

Table 4.4D The HLA restrictions of the M 9-mer peptide epitopes detected by NetTepi**.** The changes in the sequences are in bold font. Epitopes starting at amino acid positions 99 to 169.

Protein amino acid position at start of each epitope		**Epitopes**	**Restriction**
99	Wuhan, Beta, Delta, Omicron	FRLFARTRS/MWSF/NPET	HLA-A02:01, HLA-A03:01, HLA-A24:02, HLA-B27:05, HLA-B07:02, HLA-B15:01, HLA-B58:01
	Delta	FRLF**M**RTRS/MWSF/NPET	HLA-A24:02, HLA-B27:05, HLA-B07:02, HLA-B15:01, HLA-B58:01
	Delta	FRLF**V**RTRS/MWSF/NPET	HLA-A24:02, HLA-B27:05, HLA-B07:02, HLA-B15:01
115	Wuhan, Beta, Delta, Omicron	TNILLNVPL	HLA-B39:01
	Beta	T**K**ILLNVPL	HLA-B39:01
	Beta	T**D**ILLNVPL	HLA-B39:01
	Delta	TNILLNV**T**L	HLA-B39:01
121	Wuhan	V/PLHGTILT/R	HLA-B35:01
	Beta	V/PLHGTI**V/T**	HLA-B35:01, HLA-B07:02
	Beta	V/PL**Y**GTILT/R	HLA-A03:01
	Delta	V/**T**LHGTILT/R	HLA-A03:01
133	Wuhan, Beta, Delta, Omicron	LE/SELVIGA/VILR	HLA-A11:01, HLA-B40:01
	Beta	LE/**I**ELVIGA/VILR	HLA-B40:01
	Beta	LE/**N**ELVIGA/VILR	HLA-B40:01
	Beta	LE/SELVIGA/V**V**LR	HLA-A11:01, HLA-B40:01
	Delta	LE/SELVIG**V**/VILR	HLA-A11:01, HLA-B40:01
141	Wuhan, Beta, Delta, Omicron	AVILRGHLR/IAGHHL	HLA-A11:01, HLA-B07:02
	Beta	AV**V**LRGHLR/IAGHHL	HLA-A11:01
	Delta	**V**VILRGHLR/IAGHHL	HLA-A11:01
	Delta	AVILRG**Y**LR/IAGHHL	HLA-A11:01, HLA-B07:02, HLA-B15:01
	Delta	AVILRGHLR/IAGH**Y**L	HLA-B07:02
163	Wuhan	LPKEITVAT/SRTL	HLA-B07:02, HLA-B35:01, HLA-B58:01
	Beta	LPKEITVAT/SR**ML**	HLA-A26:01
169	Wuhan, Beta, Delta, Omicron	V/ATSRTLSY/YKL	HLA-A01:01, HLA-B15:01, HLA-B27:05, HLA-B35:01, HLA-B39:01, HLA-B58:01
	Beta	V/ATSR**M**LSY/YKL	HLA-B15:01, HLA-B27:05, HLA-B39:01
	Delta	V/ATSRTLSY/Y**R**L	HLA-B27:05, HLA-B39:01

Table 4.4E The HLA restrictions of the M 9-mer peptide epitopes detected by NetTepi. The changes in the sequences are in bold font. Epitopes starting at amino acid positions 178 to 208.

Protein amino acid position at start of each epitope		**Epitopes**	**Restriction**
178	Wuhan, Beta, Delta, Omicron	YKL/GASQRV/AGDSGF	HLA-B39:01
	Delta	Y**R**L/GASQRV/AGDSGF	HLA-B39:01, HLA-B27:05
	Beta	YKL/GASQRV/AG**Y**SGF	HLA-B15:01, HLA-B27:05
187	Wuhan, Beta, Delta, Omicron	AGDSGFAAY/S	HLA-A01:01
	Beta	AG**Y**SGFAAY/S	HLA-B15:01, HLA-B35:01
	Beta	AG**E**SGFAAY/S	HLA-B40:01
195	Wuhan, Beta, Delta, Omicron	YSRYRIGNY/KL	HLA-A24:02; HLA-A26:01, HLA-B15:01, HLA-B27:05
208	Wuhan, Beta, Delta, Omicron	DH/SSSSDNI/ALLV	HLA-A01:01, HLA-B39:01
	Beta	**Y**H/SSSSDNI/ALLV	HLA-B39:01
	Beta	DH/S**R**SSDNI/ALLV	HLA-B39:01, HLA-B58:01
	Beta	DH/SS**N**SDNI/ALLV	HLA-A01:01, HLA-B39:01
	Beta	DH/SSSS**G**NI/ALLV	HLA-A01:01, HLA-B39:01
	Beta	DH/SSSSDN**L**/ALLV	HLA-A01:01
	Delta	DH/SSS**G**SDNI/ALLV	HLA-A01:01

4.3.2 Immune escape mutations in E protein

The T cell epitopes of the E protein are shown in table 4.6. A total of 3 CTL epitopes ($_{1}$MYSFVSEET/GTLIV, $_{11}$T/LIVNSVLLF/LAFVV and $_{18}$LFLAFVVFL/L) with escape mutations were identified (Table 4.6A). Five CTL epitopes ($_{22}$FVV/FLLVTLAIL/TALR, $_{30}$LAI/LTALRLCAY/C, $_{37}$RLC/AYCCNIVNV/SLVK, $_{48}$VS/LVKPSFYVY/SRVK and $_{58}$YSRVKNLNS/SRVPDLLV) present in E with the relevant escape mutations are shown in table 4.6B. A N-linked glycosylation loss in $_{58}$YSRVKNLNS/**S**RVPDLLV epitope is linked to mutation (YSRVKNLNS/**(F/P)**RVPDLLV) linked to HLA-03:01; HLA-A11:01; HLA-B07:02 restriction.

Table 4.5 The B epitopes in the M proteins detected by the Bepipred tool.

Reference	Epitopic region	Targeted positions	Epitope	Present in
Wuhan	5-9	5, 6, 7, 9	NGTIT	Beta, Delta, Omicron
Wuhan	11-19	11,12,15,18,19	EELKKLLEQ	Beta, Delta, Omicron
		74	N	Beta, Omicron
Wuhan		41	N	Beta, Delta, Omicron
Wuhan	134-137	134,135,136,137	LESE	Beta, Delta, Omicron
		160	D	Beta, Omicron
Wuhan	162-163	162,163	KD	Beta, Delta, Omicron
		165	P	Beta, Omicron
		166	K	Beta, Omicron
Wuhan	180-189	180,182,183,184,185,186,188,189	KLGASQRVAG	Beta, Delta, Omicron
Wuhan	190-207	190,191,200,201,202,203,205,207	DSGFAAYSRYRIGNYKLN	Beta, Delta, Omicron
Wuhan	209-218	209,210,211,212,213,214,215,216,218	DHSSSSDNIA	Beta, Delta, Omicron

Table 4.6A The HLA restriction of the E 9-mer peptide epitopes detected by NetTepi.

The changes in the sequences are in bold font. Epitopes starting at amino acid positions 1 to 18.

Protein amino acid position at start of each epitope	**Strains**	**Epitope**	**Restriction**
1	Wuhan, Beta, Delta	MYSFVSEET/GTLIV	HLA-A02:01; HLA-B35:01; HLA-40:01
	Beta, Delta, Omicron	MYSFVSEE(**I/V)**/GTLIV	HLA-A02:01; HLA-A26:01; HLA-B35:01; HLA-B39:01; HLA-B40:01
	Beta	MYSFVSEET/G**A**LIV	HLA-A26:01; HLA-B35:01, HLA-B39:01; HLA-B40:01
	Beta	MYSFVSEET/**(S/C)** TLIV	HLA-A02:01; HLA-A26:01; HLA-B35:01, HLA-B39:01; HLA-B40:01
	Beta, Omicron	MYSF**I**SEET/GTLIV	HLA-A26:01; HLA-B35:01, HLA-B39:01; HLA-B07:02
		MYSFVSEET/GTLI(**L/I**)	HLA-B40:01; HLA-B39:01
11	Wuhan, Beta, Delta, Omicron	T/LIVNSVLLF/LAFVV	HLA-A02:01; HLA-B15:01; HLA-B58:01
	Beta	**A**/LIVNSVLLF/LAFVV	HLA-A02:01
	Delta	T/LIVN**I**VLLF/LAFVV	HLA-A02:01; HLA-A26:01; HLA-B15:01; HLA-B58:01
	Beta	T/LIVNSVL**(R/P)** F/LAFVV	HLA-A02:01; HLA-A26:01; HLA-B15:01; HLA-B27:05; HLA-B35:01; HLA-B58:01
	Beta, Delta	T/LI(**I/L**) NSVLLF/LAFVV	HLA-A02:01; HLA-A26:01; HLA-B15:01; HLA-B58:01
	Beta, Delta	T/LIVNSVLLF/**(F/I/V**) AFVV	HLA-A02:01; HLA-A24:02; HLA-B15:01; HLA-B58:01
	Beta	T/LIVNSVLLF/L**V**FVV	HLA-A02:01; HLA-B15:01
	Beta, Delta	T/LIVNSVLLF/**(P/F/I**) AFVV	HLA-A02:01; HLA-A24:02; HLA-B15:01; HLA-B35:01
	Delta	T/LIVNSVLLF/LAF**M**V	HLA-A02:01; HLA-B15:01
18	Wuhan	LFLAFVVFL/L	HLA-A02:01
	Beta	LF**P**AFVVFL/L	HLA-A24:02; HLA-B07:02; HLA-B35:01
	Beta, Delta	LF (**I/V/F**) AFVVFL/L	HLA-A02:01; HLA-A24:02; HLA-A26:01
	Beta	**R**FLAFVVFL/L	HLA-A24:02
	Beta	LFLAFVV**S**L/L	HLA-A02:01
	Beta	LFL**V**FVVFL/L	HLA-A02:01
	Delta	LFLAF**M**VFL/L	HLA-A02:01

Table 4.6B The HLA restriction of the E 9-mer peptide epitopes detected by NetTepi. The changes in the sequences are in bold font. Epitopes starting at amino acid position 22 to 58.

Protein amino acid position at start of each epitope	**Strains**	**Epitope**	**Restriction**
22	Wuhan, Beta, Delta, Omicron	FVV/FLLVTLAIL/TALR	HLA-A02:01, HLA-A24:02
	Beta	FVV/FLLV**I**LAIL/TALR	HLA-A02:01; HLA-A03:01; HLA-A24:02
	Beta	FVV/FLLVTLAI**I**/TALR	HLA-A02:01; HLA-A03:01
		FVV/FLLVTLAIL/T**V**LR	HLA-B58:01
	Beta	FVV/**S**LLVTLAIL/TALR	HLA-A02:01; HLA-B39:01; HLA-B58:01
	Delta	F**M**V/FLLVTLAIL/TALR	HLA-A02:01; HLA-B15:01; HLA-B39:01
30	Wuhan, Beta, Delta, Omicron	LAI/LTALRLCAY/C	HLA-A01:01, HLA-A26:01; HLA-B15:01; HLA-B58:01
	Beta	LAI/LTALRL**F**AY/C	HLA-A01:01; HLA-B15:01; HLA-B35:01
	Beta	LAI/LT**V**LRLCAY/C	HLA-A01:01; HLA-B58:01
	Beta	LAI/**I**TALRLCAY/C	HLA-A01:01; HLA-B58:01
	Beta	LAI/LTALRLC**V**Y/C	HLA-A01:01
	Beta	LAI/LTALRLCAY/**F**	HLA-A24:02, HLA-B58:01
37	Wuhan, Beta, Delta, Omicron	RLC/AYCCNIVNV/SLVK	HLA-A02:01; HLA-A03:01; HLA-A11:01
	Beta	RLC/AY**F**CNIVNV/SLVK	HLA-A02:01; HLA-A024:02
	Beta	RLC/AYCCNI**F**NV/SLVK	HLA-A02:01; HLA-A03:01; HLA-A11:01; HLA-A024:02
	Beta	RLC/**V**YCCNIVNV/SLVK	HLA-A02:01; HLA-A024:02
	Beta	RLC/AYCCN**V**VNV/SLVK	HLA-A03:01; HLA-A11:01
48	Wuhan, Beta, Delta, Omicron	VS/LVKPSFYVY/SRVK	HLA-A01:01; HLA-A02:01; HLA-A03:01; HLA-B15:0:1; HLA-B35:01; HLA-B58:01
	Beta	VS/LVKPSFY(**F/I**) Y/SRVK	HLA-A02:01; HLA-A03:01; HLA-B15:01; HLA-B35:01
	Delta	VS/LV**R**PSFYVY/SRVK	HLA-A01:01; HLA-A02:01; HLA-B15:01; HLA-B35:01; HLA-B58:01
	Beta	VS/LVKP**F**FYVY/SRVK	HLA-A01:01; HLA-A02:01; HLA-B15:01; HLA-B35:01; HLA-B58:01
58	Wuhan	YSRVKNLNS/SRVPDLLV	
	Beta	YSRVKNLN**F**/SRVPDLLV	HLA-A11:01; HLA-A24:02; HLA-B15:01
	Delta	YSRVKNLNS/(**F/P**) RVPDLLV	HLA-03:01; HLA-A11:01; HLA-B07:02

The B cell epitopes of the E protein are shown in table 4.7. A total of seven potential B cell epitopes were identified (Table 4.7) in context of the Wuhan reference strain for the Beta, Delta and Omicron sequences. B epitopes (SEE) are linked to position 6-8, 60-64 (SRVKN) and 66-

71 (NSSRVP). The positions 66 to 69 (NSSR) is a predicted N-linked glycosylation site. The Beta and Delta VOCs have two additional predicted epitopes compared to Wuhan and Omicron. The Delta variant has predicted epitope associated with amino acids at position 5 and 9. The Beta variant has predicted epitopes associated with amino acid positions 62 and 65.

Table 4.7 The B epitopes of the E peptides detected by BepiPred. The B epitope are relevantly conserved.

Reference	Epitopic region	Targeted position	Epitope	Present in
		5	V	Delta
Wuhan	6-8	6,7,8	SEE	Beta, Delta, Omicron
		9	T	Delta
Wuhan	60-64	60,61,63,64	SRVKN	Beta, Delta, Omicron
		62	V	Beta
		65	L	Beta
Wuhan	66-71	66,67,68,69,70,71	NSSRVP	Beta, Delta, Omicron

4.4 Accounting for SARS-CoV-2 strain diversity to derive candidate vaccine immunogens

The sequence datasets of 138 unique E gene sequences and 560 unique M gene sequences from South African SARS-CoV-2 strains were also used to derive mosaic peptide cocktails which account for amino acid sequence diversity within each dataset. The natural protein sequences identified that best represented the mosaic peptide pools derived from each of the South African M and E datasets was identical to the Wuhan reference strain’s sequences. Positional epitope coverage (which indicates the proportion of sequences in the dataset that have the specific epitopes present) across the M protein sequence alignment for 9-mer peptide epitopes was 90% to 100% across the M protein except at amino acid positions 60 to 70 where exact match coverage drops to below 60% (Figure 4.1a). The unique 9-mer counts of the E peptide in Figure 4.1b show that the exact 9-mer coverage by position is 80% to 90% for most of the E protein except at the C terminal end where it drops to around 60%.

In figure 4.2 an overview of the positional epitope coverage for each strain’s M and E sequences are provided. Yellow indicates the perfect matches per 9-mer peptide and dark brown indicates complete mismatches or very low peptide coverage.

The unique 9-mer counts of the E peptides in Figure 4.2 show that there is no match in the ends of the peptide of the Beta sequences and at the beginning of the Omicron sequences. The M peptides 9-mer coverage by position has a high number of exact matches save for a decrease in exact match coverage by 0.5 in the alignment positions around 82 as shown in Figure 4.1a. The unique 9-mer counts of the M peptide in Figure 4.2 show that the areas of the peptide are non-silent mismatch mutation does not match for the 9-mer epitope sequence.

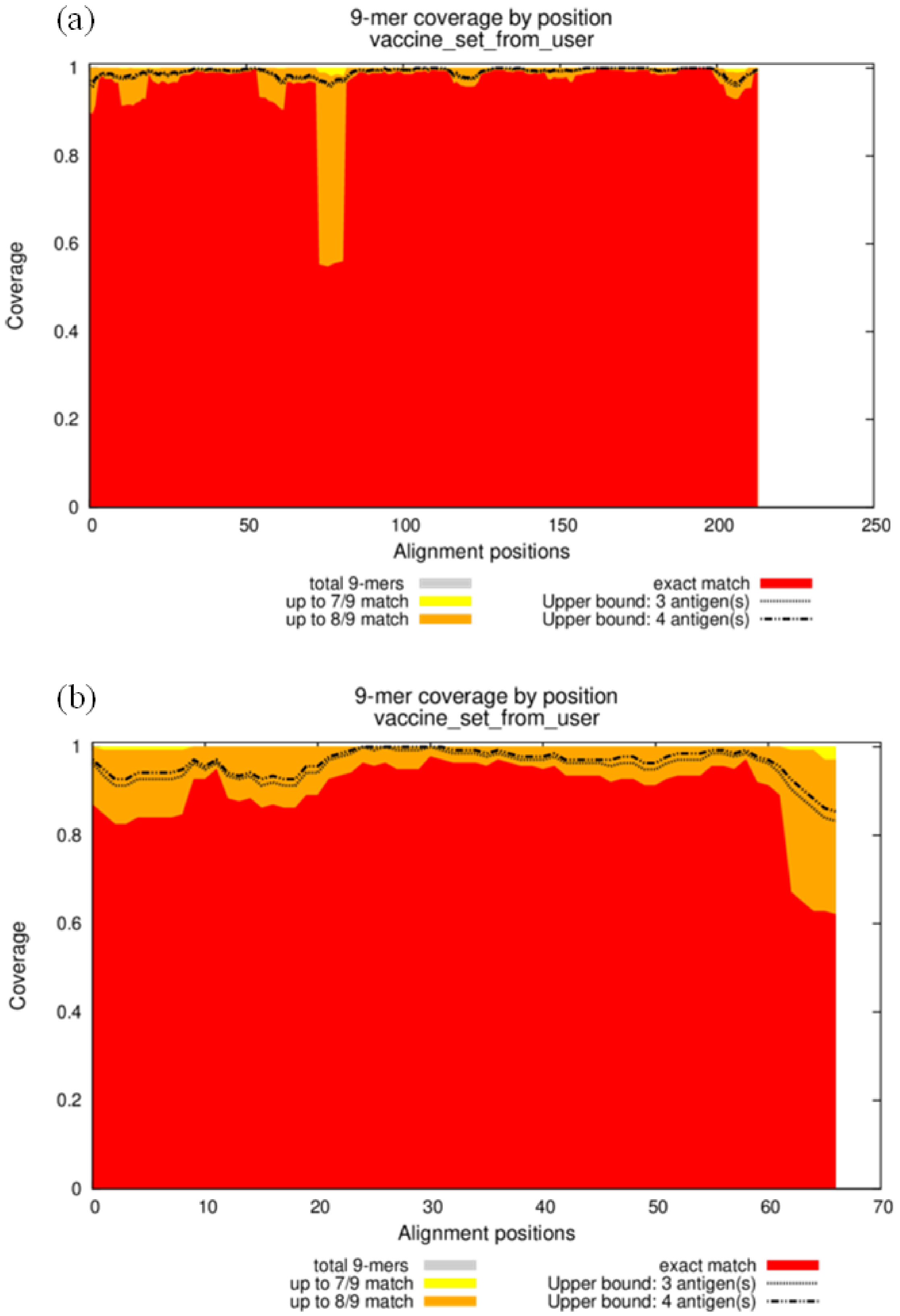


Figure 4.1 Graphical display of predicted positional 9-mer epitope coverage for unique (a) M and (b) E protein sequences from South African SARS-CoV-2 strains.

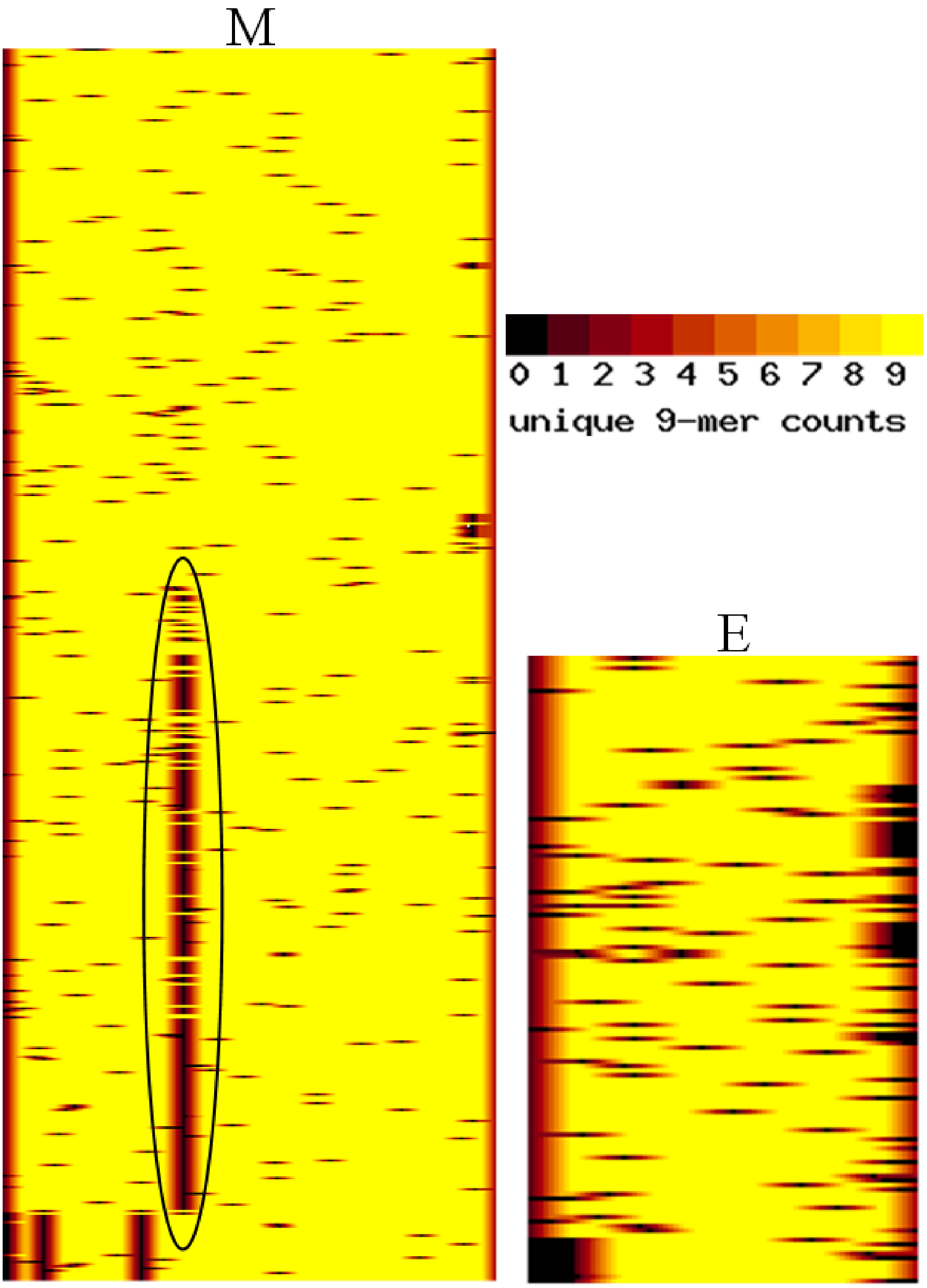


Figure 4.2 The unique 9-mer counts of M and E peptides for each aligned peptide sequence. The circled area for the M protein indicates poor CTL epitope coverage associated with the amino acid mutations in Beta and Delta variants’ protein sequences.

CHAPTER 5

DISCUSSION

The results from the phylogenetic analysis of the M and E genes showed that some of the SARS-CoV-2 variants formed distinct clusters. The E gene tree (Figure 3.1) showed that the Beta and Delta variants lie on the same branch as the reference strain showing that the sequences are conserved. Only the Omicron variant formed a distinct cluster in the phylogenetic tree. The Omicron variants all possess an ACC to ATT mutation at positions 25 to 27 of the gene sequence which causes a threonine to isoleucine mutation in the protein. A threonine to isoleucine change makes the protein region more hydrophobic (Rehman *et al.*, 2020). It has been documented that mutations found in the E protein may led to changes in the structural conformation of the protein there by altering the functions of the virus with regards to the replication, propagation, pathogenesis and viral assembly (Rahman *et al.*, 2021). These mutations in the C-terminal domain include L39M, A41V, C43F, A41S, C43R, C44Y, C43S and N45R. Structural changes due to mutations in the E protein's transmembrane domain and C-terminal domain have also been observed to affect viral replication and pathogenesis for severe acute respiratory syndrome-associated coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) (Rahman *et al.*, 2021).

The M gene tree (Figure 3.1) showed that only the Beta variant was on the same branch as the reference strain. The Delta variants form a cluster with a mutation of ATT to ACC at the positions of 244 to 246 which caused an isoleucine to threonine change. An isoleucine to threonine substitution causes a modification of non-polar to polar amino acid thereby leading to an increase in stability in the viral structure (Marques-Pereira *et al.*, 2022). The isoleucine to threonine mutation is also involved with the putative glucose transport transmembrane helices of the M protein (Shen *et al.*, 2021). The Omicron variant forms a cluster with three mutations, a GAT to CGC at position 7 to 9 causing an aspartic acid to glycine mutation, a CAG to GAA at position 55 to 57 causing a glutamine to glutamic acid mutation and a GCG to ACC mutation at position 187 to 189 causing an alanine to threonine mutation. An aspartic acid to glycine mutation has only been document to occur in the Omicron BA.1 lineages and can possibly create a N-myristoylation site at the position which will increase protein-protein interactions (Hossain *et al.*, 2022). N-myristoylation cause cotranslational lipidic modifications at the alpha-amino groups of the N-terminal residues to increases specific protein-protein interactions (Udenwobele *et al.*, 2017, Wang *et al.*, 2021). This is thought promote viral infection, assembly, budding and host interactions. The glutamine to glutamic acid mutation at position 19 of the M protein that may influence SARS-CoV-2 RNA packaging was only observed in the Omicron variant (Khairnar *et al.*, 2022). The alanine to threonine mutation at position 63 may impact the stability of the M protein and thus promote pathogenesis or immune escape of Omicron variants (Hossain *et al.*, 2022).

The results from the HIV sequence database tools show a detailed report on the mutations of the SARS-CoV-2 variants. SARS-CoV-2 variants demonstrating a disproportionate bias for T to U mutations has been reported previously (Kosuge *et al.*, 2020). This may explain why so many of the mutation seen in the highlighter nucleotide mismatch plots (Figures 3.3) are a change to a T nucleotide. These point mutations are linked to an increased ability to stimulate inflammatory responses (Kosuge *et al.*, 2020). Among the E gene sequences the Omicron variants from 2021 had a consistent mutation from a C to T at position 26 (T9I amino acid change) which is not seen in the Beta and Delta variants. This mutation is in the N-terminal functional region of the E protein. The M gene of the Beta and Delta variants from 2021 have a consistent point mutation from T to a C nucleotide in position 245 (I82T amino acid change; Figure 3.5). The Delta variant sequences have two mutations, a C55G and G187A nucleotide changes (Figure 3.5), which is also seen in the Omicron variant.

The analysis of the mismatch, silent and non-silent mutations show consistent trends in the variants. The E gene sequences of the Beta variants have a CCG to CTG mutation at position 217 (Figure 3.5) which result in proline to leucine changes at position 71. This mutation is thought to affect interaction with host proteins called PDZ (Postsynaptic density protein of 95 kDa [PSD95]/ Drosophila disc large tumour suppressor/ zomula occludens-1 protein [Zo-1]) through the Zo-1 binding domain to promote viral entry. Mutations in the E protein may disrupt binding to PDZ and help with viral escape from neutralising antibodies (Zhu *et al.*, 2022). The proline to leucine mutation at position 71 is also associated with an increase in disease severity and death rates associated with SAR-CoV-2 Beta variants (Mohammad *et al.*, 2021). In figure 3.3 the ACA to ATA mutation in E genes which results in a threonine to isoleucine change at position 26 in Omicron variants affects the ion channel properties and attenuates virus production and virulence (Xia *et al.*, 2022). The M gene mismatch mutations of the late Beta and Delta variants show the T to C mutation which is the isoleucine to threonine mutation at position 82. The mismatch mutations in the Omicron variants show the three nucleotide mutations of a to G in position 8, C to G in position 55 and a G to A at position 187 which cause the respective mutations of aspartic acid to glycine, glutamine to glutamic acid and alanine to threonine. The non-silent and silent mutation shown graphically in figures 3.5 and 3.6 mirror the consistent nucleotide changes in figure 3.3 and 3.4, respectively. Presently, information in peer-reviewed literature on the direct implications of mutations in E or M in immune evasion are lacking.

The predicted T and B cell epitope analysis is critical for determining the diversity of the antigen among the South African variants. The most prominent HLA type restriction of the M epitopes is HLA-B58:01 and HLA-A02:01 is the most prominent HLA type restriction for the E epitopes. The HLA-A*02:05/HLA-B*58:01 is the most prevalent HLA allele combination in the black African population with a frequency of 5.1% in South Africa (Paximadis *et al.*, 2012). HLA-A*02:01 is one of the most frequent class I alleles and is present in 40% of the Caucasoid population (Ishii, 2020) which resembles the most frequent HLA alleles restrictions reported in this study for M and E epitopes identified. The M gene of one of the Delta variant gains a N-linked glycosylation site ($_{115}$TNILL**NVPL** to $_{115}$TNILL**NVTL**; Table 4.4D) which may cause immune evasion. Unfortunately, due to the focus on the S protein as vaccine antigen together with the much smaller size of M and E viral surface glycoproteins their possible contribution to immune escape are not being considered. Presently, it is not known how widespread this escape mutation may be among global SARS-CoV-2 variants. N-linked glycans can provide protection form neutralizing antibodies thereby contributing to immune evasion which has been document in SARS-CoV (Vigerust and Shepherd, 2007). The E gene from one of the Delta variant sequences losses a N-linked glycosylation site, this can possibly affect viral entry of the virus as glycosylation is an important factor for host recognition and penetration (Gong *et al.*, 2021)..The B cell epitope analysis showed that there was no variation between the Wuhan reference strain and South African SARS-CoV-2 variants' M and E epitope sites identified. This again highlights the more conserved nature of M and E protein peptide epitopes compared the S protein where mutations in B-cell epitopes are frequently observed (Yuan and Li, 2021).

The mosaic vaccine cocktail produced is identical to the Wuhan reference strain peptide. A mosaic protein is an artificial recombinant protein created from a set of reference proteins and are generally used as vaccine candidates (Fischer *et al.*, 2007). Mosaic vaccine often outperforms conventional vaccination in infection prevention, improved vaccine efficacy and lower individual risks (McLeod *et al.*, 2021). The fact that the mosaic vaccine cocktails' best natural sequence is identical to the Wuhan reference strain could mean that a candidate vaccine would induce immune responses to all strains. Other research into the use of mosaic vaccines in the spike protein have shown that a mosaic vaccine has the potential to elicit neutralizing antibodies with a broad effectiveness against SARS-CoV-2 variants (Kang *et al.*, 2022).

The positional epitope coverage by the mosaic peptides of the E peptide shows high coverage with exact matches of CTL epitopes and for most of the protein alignment positions coverage is above 80%. Areas of lower coverage are associated with the emergence of Omicron VOC. The positional epitope coverage (which indicates the proportion of sequences in the dataset that have the specific epitopes present) in the mosaic peptide cocktail for the M protein also shows a high coverage for exact matched 9-mer peptides but lower frequency for these exact matched 9-mers were seen at around position 82 in the protein sequence. This reduced coverage was linked to the I82T mutation in Beta and Delta variants as also reported by Sahni *et al.* (2022) in SARS-CoV-2 variants from patients with vaccine breakthrough infections. The high coverage shows that mosaic peptide-based vaccines will induce immune responses associated with CTL epitopes across the M and E peptides of SARS-CoV-2 strains. However, it is important to note that the composition of the mosaic peptide pools for optimal T cell epitope coverage for both M and E still identified the Wuhan reference strain's sequences as the best natural proteins despite the amino acid changes observed. However, the same can't be said for the S protein sequences that are also under investigation in our laboratory (unpublished data, K. Subramoney).

The results from the VaxiJen confirmed that the variant's M and E proteins are indeed antigens that can be used as potential targets for the adaptive immune responses. The AllerTop results showed that both the M and E protein are non-allergens and therefore we do not expect the development of allergic reactions in vaccines if used as vaccine targets. The Toxinpred analysis showed that certain parts of the peptides were toxic and should not be considered as is for epitope-based vaccines (Tables 4.1 and 4.2). The toxic M epitopic regions is in the transmembrane domain 1 of the protein. Toxic effects of the expressed E gene were shown in cell cultures and the E protein on its own is believed to cause cell damage associated with acute respiratory distress syndrome in infected persons (Xia *et al.*, 2022). Therefore, toxic peptide region (ALRLCAYCCNIVNIVNVSLV) in the C-terminal domain of the E protein should not be used as is or must be excluded or modified for improved vaccine safety. We therefore conclude that SARS-CoV-2 E protein should not be used whole as vaccine target, but non-toxic peptides could be included in epitope-based vaccines.

This study does have limitations in that only South African SARS-CoV-2 viruses' M and E genes sequence data were analysed. Therefore, our study findings may not by representational in an international context. We also did not determine the global frequency of the specific mutations highlighted in this study.

Conclusion

Genomic data for South African SARS-CoV-2 M and E genes show that there is a potential for the development of M and E protein vaccine due to our finding that the best natural sequences for the respective mosaic peptide cocktails still resemble the Wuhan reference strain's sequences. However, amino acid mutations that can impact structural changes in M and E proteins from Omicron variants were observed. Further research on immune responses to the M and E protein are needed to verify the suitability of including these two proteins in vaccine design strategies.

REREFENCES

ALANAGREH, L. A., ALZOUGHOOL, F. & ATOUM, M. 2020. The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. *Pathogens,* 9**,** 331.

ASTUTI, I. & YSRAFIL 2020. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. *Diabetes & metabolic syndrome,* 14**,** 407-412.

AYYAGARI, V. S., T C, V., K, A. P. & SRIRAMA, K. 2020. Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. *Journal of biomolecular structure & dynamics***,** 1-15.

BHATTACHARYA, S., BANERJEE, A. & RAY, S. 2021. Development of new vaccine target against SARS-CoV2 using envelope (E) protein: An evolutionary, molecular modeling and docking based study. *International journal of biological macromolecules,* 172**,** 74-81.

CASCELLA, M., RAJNIK, M., ALEEM, A., DULEBOHN, S. & DI NAPOLI, R. 2022. Features, evaluation, and treatment of coronavirus (COVID-19). *StatPearls*.

CHOI, J. Y. & SMITH, D. M. 2021. SARS-CoV-2 Variants of Concern. *Yonsei Med J,* 62**,** 961-968.

CUBUK, J., ALSTON, J. J., INCICCO, J. J., SINGH, S., STUCHELL-BRERETON, M. D., WARD, M. D., ZIMMERMAN, M. I., VITHANI, N., GRIFFITH, D., WAGONER, J. A., BOWMAN, G. R., HALL, K. B., SORANNO, A. & HOLEHOUSE, A. S. 2021. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications,* 12**,** 1936.

DAI, L. & GAO, G. F. 2021. Viral targets for vaccines against COVID-19. *Nature Reviews Immunology,* 21**,** 73-82.

DESINGU, P. A., NAGARAJAN, K. & DHAMA, K. 2022. Emergence of Omicron third lineage BA.3 and its importance. *Journal of Medical Virology,* 94**,** 1808-1810.

DUAN, L., ZHENG, Q., ZHANG, H., NIU, Y., LOU, Y. & WANG, H. 2020. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Front Immunol,* 11**,** 576622.

FISCHER, W., PERKINS, S., THEILER, J., BHATTACHARYA, T., YUSIM, K., FUNKHOUSER, R., KUIKEN, C., HAYNES, B., LETVIN, N. L., WALKER, B. D., HAHN, B. H. & KORBER, B. T. 2007. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nat Med,* 13**,** 100-6.

FLORES-VEGA, V. R., MONROY-MOLINA, J. V., JIMÉNEZ-HERNÁNDEZ, L. E., TORRES, A. G., SANTOS-PRECIADO, J. I. & ROSALES-REYES, R. 2022. SARS-CoV-2: Evolution and Emergence of New Viral Variants. *Viruses,* 14.

FONTANET, A., AUTRAN, B., LINA, B., KIENY, M. P., KARIM, S. S. A. & SRIDHAR, D. 2021. SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet,* 397**,** 952-954.

GAO, T., GAO, Y., LIU, X., NIE, Z., SUN, H., LIN, K., PENG, H. & WANG, S. 2021. Identification and functional analysis of the SARS-COV-2 nucleocapsid protein. *BMC Microbiology,* 21**,** 58.

GONG, Y., QIN, S., DAI, L. & TIAN, Z. 2021. The glycosylation in SARS-CoV-2 and its receptor ACE2. *Signal Transduction and Targeted Therapy,* 6**,** 396.

GORBALENYA, A. E., BAKER, S. C., BARIC, R. S., DE GROOT, R. J., DROSTEN, C., GULYAEVA, A. A., HAAGMANS, B. L., LAUBER, C., LEONTOVICH, A. M., NEUMAN, B. W., PENZAR, D., PERLMAN, S., POON, L. L. M., SAMBORSKIY, D. V., SIDOROV, I. A., SOLA, I., ZIEBUHR, J. & CORONAVIRIDAE STUDY GROUP OF THE INTERNATIONAL COMMITTEE ON TAXONOMY OF, V. 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology,* 5**,** 536-544.

GORKHALI, R., KOIRALA, P., RIJAL, S., MAINALI, A., BARAL, A. & BHATTARAI, H. K. 2021. Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins. *Bioinformatics and biology insights,* 15**,** 11779322211025876-11779322211025876.

GOTTLIEB, R. L., VACA, C. E., PAREDES, R., MERA, J., WEBB, B. J., PEREZ, G., OGUCHI, G., RYAN, P., NIELSEN, B. U., BROWN, M., HIDALGO, A., SACHDEVA, Y., MITTAL, S., OSIYEMI, O., SKARBINSKI, J., JUNEJA, K., HYLAND, R. H., OSINUSI, A., CHEN, S., CAMUS, G., ABDELGHANY, M., DAVIES, S., BEHENNA-RENTON, N., DUFF, F., MARTY, F. M., KATZ, M. J., GINDE, A. A., BROWN, S. M., SCHIFFER, J. T. & HILL, J. A. 2021. Early Remdesivir to Prevent Progression to Severe Covid-19 in Outpatients. *New England Journal of Medicine,* 386**,** 305-315.

GÜNER, R., HASANOĞLU, I. & AKTAŞ, F. 2020. COVID-19: Prevention and control measures in community. *Turkish journal of medical sciences,* 50**,** 571-577.

GUO, G., YE, L., PAN, K., CHEN, Y., XING, D., YAN, K., CHEN, Z., DING, N., LI, W. & HUANG, H. 2020. New insights of emerging SARS-CoV-2: epidemiology, etiology, clinical features, clinical treatment, and prevention. *Frontiers in Cell and Developmental Biology,* 8**,** 410.

HARAPAN, H., ITOH, N., YUFIKA, A., WINARDI, W., KEAM, S., TE, H., MEGAWATI, D., HAYATI, Z., WAGNER, A. L. & MUDATSIR, M. 2020. Coronavirus disease 2019 (COVID-19): A literature review. *Journal of Infection and Public Health,* 13**,** 667-673.

HARDER, T., KÜLPER-SCHIEK, W., REDA, S., TRESKOVA-SCHWARZBACH, M., KOCH, J., VYGEN-BONNET, S. & WICHMANN, O. 2021. Effectiveness of COVID-19 vaccines against SARS-CoV-2 infection with the Delta (B.1.617.2) variant: second interim results of a living systematic review and meta-analysis, 1 January to 25 August 2021. *Euro Surveill,* 26.

HARRISON, A. G., LIN, T. & WANG, P. 2020. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends in immunology,* 41**,** 1100-1115.

HARVEY, W. T., CARABELLI, A. M., JACKSON, B., GUPTA, R. K., THOMSON, E. C., HARRISON, E. M., LUDDEN, C., REEVE, R., RAMBAUT, A., PEACOCK, S. J., ROBERTSON, D. L. & CONSORTIUM, C.-G. U. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology,* 19**,** 409-424.

HELMY, Y. A., FAWZY, M., ELASWAD, A., SOBIEH, A., KENNEY, S. P. & SHEHATA, A. A. 2020. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. *Journal of clinical medicine,* 9**,** 1225.

HERNANDEZ ACOSTA, R. A., ESQUER GARRIGOS, Z., MARCELIN, J. R. & VIJAYVARGIYA, P. 2022. COVID-19 Pathogenesis and Clinical Manifestations. *Infect Dis Clin North Am,* 36**,** 231-249.

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, S. 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,* 184**,** 4848-4856.

HOSSAIN, A., AKTER, S., RASHID, A. A., KHAIR, S. & ALAM, A. 2022. Unique mutations in SARS-CoV-2 Omicron subvariants' non-spike proteins: Potential impacts on viral pathogenesis and host immune evasion. *Microb Pathog,* 170**,** 105699.

HU, B., GUO, H., ZHOU, P. & SHI, Z.-L. 2021. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology,* 19**,** 141-154.

ISHII, T. 2020. Human Leukocyte Antigen (HLA) Class I Susceptible Alleles Against COVID-19 Increase Both Infection and Severity Rate. *Cureus,* 12**,** e12239.

ISLAM, M. R., SHAHRIAR, M. & BHUIYAN, M. A. 2022. The latest Omicron BA.4 and BA.5 lineages are frowning toward COVID-19 preventive measures: A threat to global public health. *Health Science Reports,* 5**,** e884.

KANG, Y.-F., SUN, C., SUN, J., XIE, C., ZHUANG, Z., XU, H.-Q., LIU, Z., LIU, Y.-H., PENG, S., YUAN, R.-Y., ZHAO, J.-C. & ZENG, M.-S. 2022. Quadrivalent mosaic HexaPro-bearing nanoparticle vaccine protects against infection of SARS-CoV-2 variants. *Nature Communications,* 13**,** 2674.

KE, Z., OTON, J., QU, K., CORTESE, M., ZILA, V., MCKEANE, L., NAKANE, T., ZIVANOV, J., NEUFELDT, C. J., CERIKAN, B., LU, J. M., PEUKES, J., XIONG, X., KRÄUSSLICH, H.-G., SCHERES, S. H. W., BARTENSCHLAGER, R. & BRIGGS, J. A. G. 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature,* 588**,** 498-502.

KEVADIYA, B. D., MACHHI, J., HERSKOVITZ, J., OLEYNIKOV, M. D., BLOMBERG, W. R., BAJWA, N., SONI, D., DAS, S., HASAN, M., PATEL, M., SENAN, A. M., GORANTLA, S., MCMILLAN, J., EDAGWA, B., EISENBERG, R., GURUMURTHY, C. B., REID, S. P. M., PUNYADEERA, C., CHANG, L. & GENDELMAN, H. E. 2021. Diagnostics for SARS-CoV-2 infections. *Nature Materials,* 20**,** 593-605.

KHAIRNAR, P., SONI, M., HANDA, M., RIADI, Y., KESHARWANI, P. & SHUKLA, R. 2022. Recent highlights on Omicron as a new SARS-COVID-19 variant: evolution, genetic mutation, and future perspectives. *Journal of drug targeting,* 30**,** 603-613.

KHATEEB, J., LI, Y. & ZHANG, H. 2021. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. *Critical Care,* 25**,** 244.

KLEIN, S., CORTESE, M., WINTER, S. L., WACHSMUTH-MELM, M., NEUFELDT, C. J., CERIKAN, B., STANIFER, M. L., BOULANT, S., BARTENSCHLAGER, R. & CHLANDA, P. 2020. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun,* 11**,** 5885.

KOSUGE, M., FURUSAWA-NISHII, E., ITO, K., SAITO, Y. & OGASAWARA, K. 2020. Point mutation bias in SARS-CoV-2 variants results in increased ability to stimulate inflammatory responses. *Scientific Reports,* 10**,** 17766.

KOYAMA, T., MIYAKAWA, K., TOKUMASU, R., S, S. J., KUDO, M. & RYO, A. 2022. Evasion of vaccine-induced humoral immunity by emerging sub-variants of SARS-CoV-2. *Future Microbiol,* 17**,** 417-424.

KUMAR, S., NYODU, R., MAURYA, V. K. & SAXENA, S. K. 2020. Morphology, Genome Organization, Replication, and Pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Coronavirus Disease 2019 (COVID-19): Epidemiology, Pathogenesis, Diagnosis, and Therapeutics***,** 23-31.

KUMAR, V., SINGH, J., HASNAIN, S. E. & SUNDAR, D. 2021. Possible link between higher transmissibility of B.1.617 and B.1.1.7 variants of SARS-CoV-2 and increased structural stability of its spike protein and hACE2 affinity. *bioRxiv***,** 2021.04.29.441933.

KUZMIN, A., OREKHOV, P., ASTASHKIN, R., GORDELIY, V. & GUSHCHIN, I. 2021. Structure and dynamics of the SARS-CoV-2 envelope protein monomer. *bioRxiv***,** 2021.03.10.434722.

LARICCIA, V., MAGI, S., SERFILIPPI, T., TOUJANI, M., GRATTERI, S. & AMOROSO, S. 2020. Challenges and Opportunities from Targeting Inflammatory Responses to SARS-CoV-2 Infection: A Narrative Review. *Journal of Clinical Medicine,* 9**,** 4021.

LARSEN, M. V., LUNDEGAARD, C., LAMBERTH, K., BUUS, S., LUND, O. & NIELSEN, M. 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics,* 8**,** 424.

LAWTON, G. 2021. Sights set on universal vaccine. *New scientist (1971),* 249**,** 8-9.

LIU, J., SUN, Y., QI, J., CHU, F., WU, H., GAO, F., LI, T., YAN, J. & GAO, G. F. 2010. The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *The Journal of infectious diseases,* 202**,** 1171-1180.

LIU, Y., WANG, K., MASSOUD, T. F. & PAULMURUGAN, R. 2020. SARS-CoV-2 Vaccine Development: An Overview and Perspectives. *ACS Pharmacology & Translational Science,* 3**,** 844-858.

LIVINGSTON, E. H., MALANI, P. N. & CREECH, C. B. 2021. The Johnson & Johnson Vaccine for COVID-19. *JAMA,* 325**,** 1575-1575.

LOTFI, M., HAMBLIN, M. R. & REZAEI, N. 2020. COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clinica chimica acta; international journal of clinical chemistry,* 508**,** 254-266.

LU, S., XIE, X.-X., ZHAO, L., WANG, B., ZHU, J., YANG, T.-R., YANG, G.-W., JI, M., LV, C.-P., XUE, J., DAI, E.-H., FU, X.-M., LIU, D.-Q., ZHANG, L., HOU, S.-J., YU, X.-L., WANG, Y.-L., GAO, H.-X., SHI, X.-H., KE, C.-W., KE, B.-X., JIANG, C.-G. & LIU, R.-T. 2021. The immunodominant and neutralization linear epitopes for SARS-CoV-2. *Cell reports,* 34**,** 108666-108666.

MADHI, S. A., BAILLIE, V., CUTLAND, C. L., VOYSEY, M., KOEN, A. L., FAIRLIE, L., PADAYACHEE, S. D., DHEDA, K., BARNABAS, S. L. & BHORAT, Q. E. 2021. Efficacy of the ChAdOx1 nCoV-19 covid-19 vaccine against the B. 1.351 variant. *New England Journal of Medicine*.

MAJAM, M., FISCHER, A., PHIRI, J., VENTER, F. & LALLA-EDWARD, S. T. 2021. International citizen project to assess early stage adherence to public health measures for COVID-19 in South Africa. *PLOS ONE,* 16**,** e0248055.

MARCONI, V. C., RAMANAN, A. V., DE BONO, S., KARTMAN, C. E., KRISHNAN, V., LIAO, R., PIRUZELI, M. L. B., GOLDMAN, J. D., ALATORRE-ALEXANDER, J., DE CASSIA PELLEGRINI, R., ESTRADA, V., SOM, M., CARDOSO, A., CHAKLADAR, S., CROWE, B., REIS, P., ZHANG, X., ADAMS, D. H. & ELY, E. W. 2021. Efficacy and safety of baricitinib for the treatment of hospitalised adults with COVID-19 (COV-BARRIER): a randomised, double-blind, parallel-group, placebo-controlled phase 3 trial. *Lancet Respir Med,* 9**,** 1407-1418.

MARQUES-PEREIRA, C., PIRES, M. N., GOUVEIA, R. P., PEREIRA, N. N., CANICEIRO, A. B., ROSÁRIO-FERREIRA, N. & MOREIRA, I. S. 2022. SARS-CoV-2 Membrane Protein: From Genomic Data to Structural New Insights. *Int J Mol Sci,* 23.

MARTÍNEZ-FLORES, D., ZEPEDA-CERVANTES, J., CRUZ-RESÉNDIZ, A., AGUIRRE-SAMPIERI, S., SAMPIERI, A. & VACA, L. 2021. SARS-CoV-2 Vaccines Based on the Spike Glycoprotein and Implications of New Viral Variants. *Frontiers in immunology,* 12**,** 701501-701501.

MATHURIA, J. P., YADAV, R. & RAJKUMAR 2020. Laboratory diagnosis of SARS-CoV-2 - A review of current methods. *J Infect Public Health,* 13**,** 901-905.

MCCALLUM, M., BASSI, J., MARCO, A. D., CHEN, A., WALLS, A. C., IULIO, J. D., TORTORICI, M. A., NAVARRO, M.-J., SILACCI-FREGNI, C., SALIBA, C.,

AGOSTINI, M., PINTO, D., CULAP, K., BIANCHI, S., JACONI, S., CAMERONI, E., BOWEN, J. E., TILLES, S. W., PIZZUTO, M. S., GUASTALLA, S. B., BONA, G., PELLANDA, A. F., GARZONI, C., VAN VOORHIS, W. C., ROSEN, L. E., SNELL, G., TELENTI, A., VIRGIN, H. W., PICCOLI, L., CORTI, D. & VEESLER, D. 2021. SARS-CoV-2 immune evasion by variant B.1.427/B.1.429. *bioRxiv*, 2021.03.31.437925.

MCLEOD, D. V., WAHL, L. M. & MIDEO, N. 2021. Mosaic vaccination: How distributing different vaccines across a population could improve epidemic control. *Evolution Letters,* 5**,** 458-471.

MOHAMMAD, T., CHOUDHURY, A., HABIB, I., ASRANI, P., MATHUR, Y., UMAIR, M., ANJUM, F., SHAFIE, A., YADAV, D. K. & HASSAN, M. I. 2021. Genomic Variations in the Structural Proteins of SARS-CoV-2 and Their Deleterious Impact on Pathogenesis: A Comparative Genomics Approach. *Front Cell Infect Microbiol,* 11**,** 765039.

MOHAPATRA, R. K., KANDI, V., SARANGI, A. K., VERMA, S., TULI, H. S., CHAKRABORTY, S., CHAKRABORTY, C. & DHAMA, K. 2022. The recently emerged BA.4 and BA.5 lineages of Omicron and their global health concerns amid the ongoing wave of COVID-19 pandemic - Correspondence. *Int J Surg,* 103**,** 106698.

MWANGI, P., OKENDO, J., MOGOTSI, M., OGUNBAYO, A., ADELABU, O., SONDLANE, H., MAOTOANA, M., MAHOMED, L., MOROBADI, M. D., VAWDA, S., VON GOTTBERG, A., BHIMAN, J., TEGALLY, H., WILKINSON, E., GIANDHARI, J., PILLAY, S., NAIDOO, Y., RAMPHAL, U., DE OLIVEIRA, T., BESTER, A., GOEDHALS, D. & NYAGA, M. 2022. SARS-CoV-2 variants from COVID-19 positive cases in the Free State province, South Africa from July 2020 to December 2021. *Frontiers in Virology,* 2.

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 therapies: Structural genomics approach. *Biochimica et biophysica acta. Molecular basis of disease,* 1866**,** 165878-165878.

NIKNAM, Z., JAFARI, A., GOLCHIN, A., DANESH POUYA, F., NEMATI, M., REZAEI-TAVIRANI, M. & RASMI, Y. 2022. Potential therapeutic options for COVID-19: an update on current evidence. *European Journal of Medical Research,* 27**,** 6.

OTTO, S. P., DAY, T., ARINO, J., COLIJN, C., DUSHOFF, J., LI, M., MECHAI, S., VAN DOMSELAAR, G., WU, J., EARN, D. J. D. & OGDEN, N. H. 2021. The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Current biology : CB,* 31**,** R918-R929.

PAXIMADIS, M., MATHEBULA, T. Y., GENTLE, N. L., VARDAS, E., COLVIN, M., GRAY, C. M., TIEMESSEN, C. T. & PUREN, A. 2012. Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. *Human Immunology,* 73**,** 80-92.

RAHMAN, H. S., AZIZ, M. S., HUSSEIN, R. H., OTHMAN, H. H., SALIH OMER, S. H., KHALID, E. S., ABDULRAHMAN, N. A., AMIN, K. & ABDULLAH, R. 2020a. The transmission modes and sources of COVID-19: A systematic review. *International Journal of Surgery Open,* 26**,** 125-136.

RAHMAN, M. S., HOQUE, M. N., ISLAM, M. R., AKTER, S., RUBAYET UL ALAM, A. S. M., SIDDIQUE, M. A., SAHA, O., RAHAMAN, M. M., SULTANA, M., CRANDALL, K. A. & HOSSAIN, M. A. 2020b. Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: an in silico approach. *PeerJ,* 8**,** e9572-e9572.

RAHMAN, M. S., HOQUE, M. N., ISLAM, M. R., ISLAM, I., MISHU, I. D., RAHAMAN, M. M., SULTANA, M. & HOSSAIN, M. A. 2021. Mutational insights into the envelope protein of SARS-CoV-2. *Gene Rep,* 22**,** 100997.

RANDOLPH, H. E. & BARREIRO, L. B. 2020. Herd Immunity: Understanding COVID-19. *Immunity,* 52**,** 737-741.

RAPIN, N., LUND, O., BERNASCHI, M. & CASTIGLIONE, F. 2010. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. *PloS one,* 5**,** e9862-e9862.

REHMAN, S., MAHMOOD, T., AZIZ, E. & BATOOL, R. 2020. Identification of novel mutations in SARS-COV-2 isolates from Turkey. *Arch Virol,* 165**,** 2937-2944.

RUBIN, R. 2022. Baricitinib Is First Approved COVID-19 Immunomodulatory Treatment. *JAMA,* 327**,** 2281-2281.

SANYAOLU, A., OKORIE, C., MARINKOVIC, A., HAIDER, N., ABBASI, A. F., JAFERI, U., PRAKASH, S. & BALENDRA, V. 2021. The emerging SARS-CoV-2 variants of concern. *Ther Adv Infect Dis,* 8**,** 20499361211024372.

SAXENA, A. 2020. Drug targets for COVID-19 therapeutics: Ongoing global efforts. *Journal of biosciences,* 45**,** 87.

SCHÜTZ, D., RUIZ-BLANCO, Y. B., MÜNCH, J., KIRCHHOFF, F., SANCHEZ-GARCIA, E. & MÜLLER, J. A. 2020. Peptide and peptide-based inhibitors of SARS-CoV-2 entry. *Adv Drug Deliv Rev,* 167**,** 47-65.

SHAHCHERAGHI, S. H., AYATOLLAHI, J., ALJABALI, A. A., SHASTRI, M. D., SHUKLA, S. D., CHELLAPPAN, D. K., JHA, N. K., ANAND, K., KATARI, N. K., MEHTA, M., SATIJA, S., DUREJA, H., MISHRA, V., ALMUTARY, A. G., ALNUQAYDAN, A. M., CHARBE, N., PRASHER, P., GUPTA, G., DUA, K., LOTFI, M., BAKSHI, H. A. & TAMBUWALA, M. M. 2021. An overview of vaccine development for COVID-19. *Therapeutic Delivery,* 12**,** 235-244.

SHEN, L., BARD, J. D., TRICHE, T. J., JUDKINS, A. R., BIEGEL, J. A. & GAI, X. 2021. Emerging variants of concern in SARS-CoV-2 membrane protein: a highly conserved target with potential pathological and therapeutic implications. *Emerg Microbes Infect,* 10**,** 885-893.

TAVILANI, A., ABBASI, E., KIAN ARA, F., DARINI, A. & ASEFY, Z. 2021. COVID-19 vaccines: Current evidence and considerations. *Metabolism Open,* 12**,** 100124.

TEGALLY, H., WILKINSON, E., ALTHAUS, C. L., GIOVANETTI, M., SAN, J. E., GIANDHARI, J., PILLAY, S., NAIDOO, Y., RAMPHAL, U. & MSOMI, N. 2021a. Rapid replacement of the Beta variant by the Delta variant in South Africa. *medRxiv*.

TEGALLY, H., WILKINSON, E., GIOVANETTI, M., IRANZADEH, A., FONSECA, V., GIANDHARI, J., DOOLABH, D., PILLAY, S., SAN, E. J. & MSOMI, N. 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*.

TEGALLY, H., WILKINSON, E., GIOVANETTI, M., IRANZADEH, A., FONSECA, V., GIANDHARI, J., DOOLABH, D., PILLAY, S., SAN, E. J. & MSOMI, N. 2021b. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature,* 592**,** 438-443.

TEGALLY, H., WILKINSON, E., GIOVANETTI, M., IRANZADEH, A., FONSECA, V., GIANDHARI, J., DOOLABH, D., PILLAY, S., SAN, E. J., MSOMI, N., MLISANA, K., VON GOTTBERG, A., WALAZA, S., ALLAM, M., ISMAIL, A., MOHALE, T., GLASS, A. J., ENGELBRECHT, S., VAN ZYL, G., PREISER, W., PETRUCCIONE, F., SIGAL, A., HARDIE, D., MARAIS, G., HSIAO, N.-Y., KORSMAN, S., DAVIES, M.-A., TYERS, L., MUDAU, I., YORK, D., MASLO, C., GOEDHALS, D., ABRAHAMS, S., LAGUDA-AKINGBA, O., ALISOLTANI-

DEHKORDI, A., GODZIK, A., WIBMER, C. K., SEWELL, B. T., LOURENÇO, J., ALCANTARA, L. C. J., KOSAKOVSKY POND, S. L., WEAVER, S., MARTIN, D., LESSELLS, R. J., BHIMAN, J. N., WILLIAMSON, C. & DE OLIVEIRA, T. 2021c. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature,* 592**,** 438-443.

TEIJARO, J. R. & FARBER, D. L. 2021. COVID-19 vaccines: modes of immune activation and future challenges. *Nature Reviews Immunology,* 21**,** 195-197.

THAKUR, S., SASI, S., PILLAI, S. G., NAG, A., SHUKLA, D., SINGHAL, R., PHALKE, S. & VELU, G. S. K. 2022. SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines. *Frontiers in Medicine,* 9.

THOMAS, S. 2020. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. *Pathogens & immunity,* 5**,** 342-363.

TIAN, D., SUN, Y., ZHOU, J. & YE, Q. 2021. The Global Epidemic of the SARS-CoV-2 Delta Variant, Key Spike Mutations and Immune Escape. *Front Immunol,* 12**,** 751778.

TROLLE, T. & NIELSEN, M. 2014. NetTepi: an integrated method for the prediction of T cell epitopes. *Immunogenetics,* 66**,** 449-56.

TUEKPRAKHON, A., NUTALAI, R., DIJOKAITE-GURALIUC, A., ZHOU, D., GINN, H. M., SELVARAJ, M., LIU, C., MENTZER, A. J., SUPASA, P., DUYVESTEYN, H. M. E., DAS, R., SKELLY, D., RITTER, T. G., AMINI, A., BIBI, S., ADELE, S., JOHNSON, S. A., CONSTANTINIDES, B., WEBSTER, H., TEMPERTON, N., KLENERMAN, P., BARNES, E., DUNACHIE, S. J., CROOK, D., POLLARD, A. J., LAMBE, T., GOULDER, P., PATERSON, N. G., WILLIAMS, M. A., HALL, D. R., FRY, E. E., HUO, J., MONGKOLSAPAYA, J., REN, J., STUART, D. I. & SCREATON, G. R. 2022. Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell,* 185**,** 2422-2433.e13.

UDENWOBELE, D. I., SU, R. C., GOOD, S. V., BALL, T. B., VARMA SHRIVASTAV, S. & SHRIVASTAV, A. 2017. Myristoylation: An Important Protein Modification in the Immune Response. *Front Immunol,* 8**,** 751.

UTTARILLI, A., AMALAKANTI, S., KOMMOJU, P.-R., SHARMA, S., GOYAL, P., MANJUNATH, G. K., UPADHAYAY, V., PARVEEN, A., TANDON, R., PRASAD, K. S., DAKAL, T. C., SHLOMO, I. B., YOUSEF, M., NEERATHILINGAM, M. & KUMAR, A. 2021. Super-rapid race for saving lives by developing COVID-19 vaccines. *Journal of Integrative Bioinformatics,* 18**,** 27-43.

V'KOVSKI, P., KRATZEL, A., STEINER, S., STALDER, H. & THIEL, V. 2021. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology,* 19**,** 155-170.

VENKATAKRISHNAN, A., ANAND, P., LENEHAN, P. J., SURATEKAR, R., RAGHUNATHAN, B., NIESEN, M. J. & SOUNDARARAJAN, V. 2021. Omicron variant of SARS-CoV-2 harbors a unique insertion mutation of putative viral or human genomic origin.

VENTER, W. D. F., MADHI, S. A., NEL, J., MENDELSON, M., VAN DEN HEEVER, A. & MOSHABELA, M. 2021. *South Africa should be using all the COVID-19 vaccines available to it – urgently*.

VIGERUST, D. J. & SHEPHERD, V. L. 2007. Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol,* 15**,** 211-8.

WANG, B., DAI, T., SUN, W., WEI, Y., REN, J., ZHANG, L., ZHANG, M. & ZHOU, F. 2021. Protein N-myristoylation: functions and mechanisms in control of innate immunity. *Cellular & Molecular Immunology,* 18**,** 878-888.

WEI, C., SHAN, K.-J., WANG, W., ZHANG, S., HUAN, Q. & QIAN, W. 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *Journal of Genetics and Genomics*.

WIBMER, C. K., AYRES, F., HERMANUS, T., MADZIVHANDILA, M., KGAGUDI, P., OOSTHUYSEN, B., LAMBSON, B. E., DE OLIVEIRA, T., VERMEULEN, M., VAN DER BERG, K., ROSSOUW, T., BOSWELL, M., UECKERMANN, V., MEIRING, S., VON GOTTBERG, A., COHEN, C., MORRIS, L., BHIMAN, J. N. & MOORE, P. L. 2021. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nature Medicine,* 27**,** 622-625.

WISE, J. 2021. Covid-19: European countries suspend use of Oxford-AstraZeneca vaccine after reports of blood clots. British Medical Journal Publishing Group.

WOLTER, N., JASSAT, W., WALAZA, S., WELCH, R., MOULTRIE, H., GROOME, M., AMOAKO, D., EVERATT, J., BHIMAN, J., SCHEEPERS, C., TEBEILA, N., CHIWANDIRE, N., PLESSIS, M., GOVENDER, N., ISMAIL, A., GLASS, A., MLISANA, K., STEVENS, W., TREURNICHT, F. & COHEN, C. 2021. *Early assessment of the clinical severity of the SARS-CoV-2 Omicron variant in South Africa*.

WOLTER, N., JASSAT, W., WALAZA, S., WELCH, R., MOULTRIE, H., GROOME, M. J., AMOAKO, D. G., EVERATT, J., BHIMAN, J. N., SCHEEPERS, C., TEBEILA,

N., CHIWANDIRE, N., DU PLESSIS, M., GOVENDER, N., ISMAIL, A., GLASS, A., MLISANA, K., STEVENS, W., TREURNICHT, F. K., SUBRAMONEY, K., MAKATINI, Z., HSIAO, N.-Y., PARBOOSING, R., WADULA, J., HUSSEY, H., DAVIES, M.-A., BOULLE, A., VON GOTTBERG, A. & COHEN, C. 2022. Clinical severity of SARS-CoV-2 Omicron BA.4 and BA.5 lineages compared to BA.1 and Delta in South Africa. *Nature Communications,* 13**,** 5860.

XIA, B., WANG, Y., PAN, X., CHENG, X., JI, H., ZUO, X., JIANG, H., LI, J. & GAO, Z. 2022. Why is the SARS-CoV-2 Omicron variant milder? *Innovation (Camb),* 3**,** 100251.

YANG, W. & SHAMAN, J. 2021. SARS-CoV-2 transmission dynamics in South Africa and epidemiological characteristics of the Omicron variant. *medRxiv***,** 2021.12.19.21268073.

YUAN, X. & LI, L. 2021. Immuno Informatic Analysis of B-Cell Epitope Changes in SARS-Cov-2 Variants with Dominant S-Protein Mutations. *Journal of Bioinformatics and Systems Biology,* 04.

ZENG, W., LIU, G., MA, H., ZHAO, D., YANG, Y., LIU, M., MOHAMMED, A., ZHAO, C., YANG, Y., XIE, J., DING, C., MA, X., WENG, J., GAO, Y., HE, H. & JIN, T. 2020. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochemical and biophysical research communications,* 527**,** 618-623.

ZHAO, Z., ZHOU, J., TIAN, M., HUANG, M., LIU, S., XIE, Y., HAN, P., BAI, C., HAN, P., ZHENG, A., FU, L., GAO, Y., PENG, Q., LI, Y., CHAI, Y., ZHANG, Z., ZHAO, X., SONG, H., QI, J., WANG, Q., WANG, P. & GAO, G. F. 2022. Omicron SARS-CoV-2 mutations stabilize spike up-RBD conformation and lead to a non-RBM-binding monoclonal antibody escape. *Nature Communications,* 13**,** 4958.

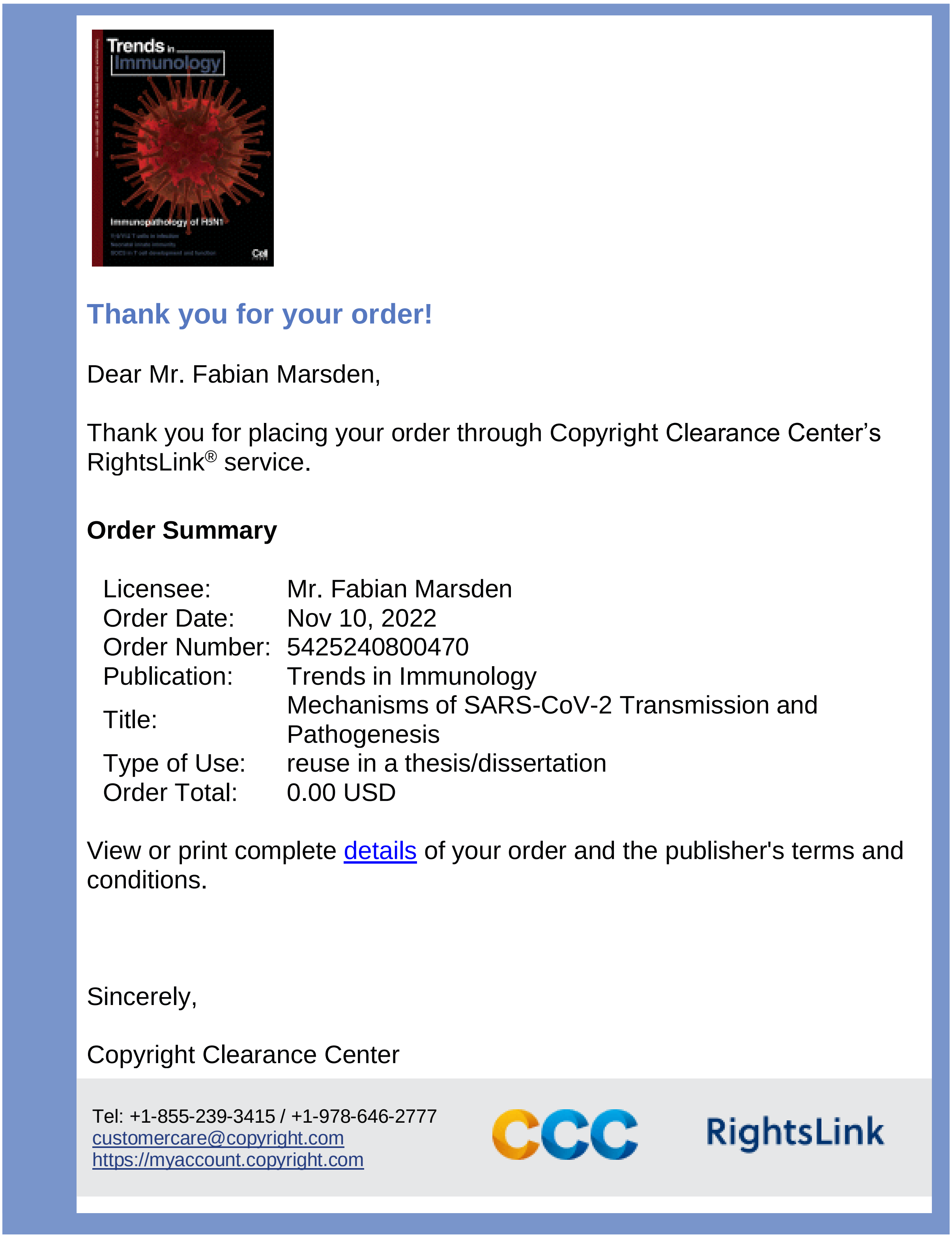
ZHOU, D., DEJNIRATTISAI, W., SUPASA, P., LIU, C., MENTZER, A. J., GINN, H. M., ZHAO, Y., DUYVESTEYN, H. M. E., TUEKPRAKHON, A., NUTALAI, R., WANG, B., PAESEN, G. C., LOPEZ-CAMACHO, C., SLON-CAMPOS, J., HALLIS, B., COOMBES, N., BEWLEY, K., CHARLTON, S., WALTER, T. S., SKELLY, D., LUMLEY, S. F., DOLD, C., LEVIN, R., DONG, T., POLLARD, A. J., KNIGHT, J. C., CROOK, D., LAMBE, T., CLUTTERBUCK, E., BIBI, S., FLAXMAN, A., BITTAYE, M., BELIJ-RAMMERSTORFER, S., GILBERT, S., JAMES, W., CARROLL, M. W., KLENERMAN, P., BARNES, E., DUNACHIE, S. J., FRY, E. E., MONGKOLSAPAYA, J., REN, J., STUART, D. I. & SCREATON,

G. R. 2021. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell*.

ZHU, Y., ALVAREZ, F., WOLFF, N., MECHALY, A., BRÛLÉ, S., NEITTHOFFER, B., ETIENNE-MANNEVILLE, S., HAOUZ, A., BOËDA, B. & CAILLET-SAGUY, C. 2022. Interactions of Severe Acute Respiratory Syndrome Coronavirus 2 Protein E With Cell Junctions and Polarity PSD-95/Dlg/ZO-1-Containing Proteins. *Front Microbiol,* 13**,** 829094.

ADDENDUM

1. Transcript from Rightslink



Thank you for your order!

Dear Mr. Fabian Marsden,

Thank you for placing your order through Copyright Clearance Center’s RightsLink® service.

Order Summary

Licensee:	Mr. Fabian Marsden
Order Date:	Nov 10, 2022
Order Number:	5425240800470
Publication:	Trends in Immunology
Title:	Mechanisms of SARS-CoV-2 Transmission and Pathogenesis
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

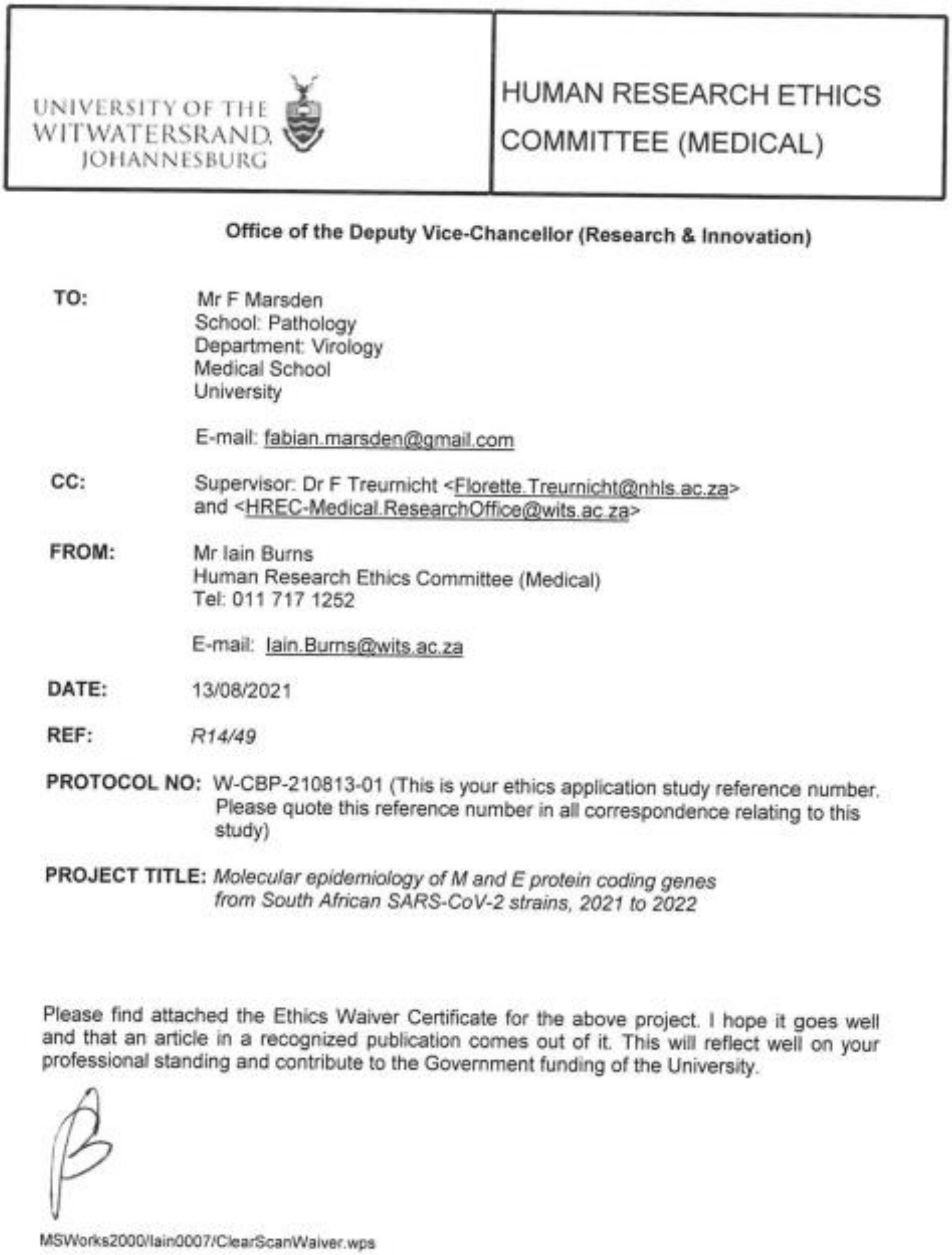
Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

2. SARS-CoV-2 reference sequence dataset (Roemer data) details

Sequence names or Accession numbers	Sequence names or Accession numbers	Sequence names or Accession numbers
ON895103	Scotland/QEUH-14E640D/2021	AUS/VIC8093/2020
ON895548	USA/AK-PHL9083/2021	AUS/VIC10882/2020
ON537316	USA/NJ-PHEL-21-12068/2021	USA/MA-MASPHL-01933/2021
ON544943	USA/MA-CDCBI-CRSP_NPPKJNMG7V6OAGRW/2021	USA/FL-CDC-QDX22678176/2021
ON629031	USA/AK-PHL10845/2021	USA/F57FCBBEAA9D4211B7F162FE6C1B1D7F/2021
ON626380	USA/FL-CDC-LC0099490/2021	OU205023
ON627541	IND/GBRC640a/2021	SYC/G38725/2021
ON627543	USA/IL-CDC-LC0060638/2021	USA/FL-CDC-ASC210334485/2021
ON396327	IND/CWS_0018_47303313/2021	USA/MA-CDCBI-CRSP_MLUB7OEWDYUWSKGQ/2021
ON544567	England/CAMC-13B81A4/2021	USA/CO-CDC-MMB10752210/2021
ON084447	USA/WA-S5413/2021	England/ALDP-20A77A6/2021
IMS-10182-CVDP-83F89C70-C702-461C-BD64-53FADC5A0601	USA/CA-CDC-QDX21613057/2021	USA/CO-CDC-ASC210309969/2021
OV950637 (XL)	USA/ID-CDC-ASC210011921/2021	England/MILK-2759487/2021
OV940149 (XF)	USA/VA-CDC-LC0067096/2021	USA/RI-CDCBI-CRSP_EVAFDCVZGQ4NYXIQ/2021
OW018845 (XE)	OU149858	USA/CA-CDC-FG-161774/2021
OM990851 (XD)	USA/VA-CDC-LC0051792/2021	USA/MA-CDCBI-CRSP_6WUXOVDJ3B5G2RNQ/2021
OK016320 (XB)	USA/NJ-CDC-LC0084532/2021	OU819939
OU219033 (XA)	England/CAMC-139B57E/2021	USA/FL-CDC-ASC210334485/2021
IMS-10004-CVDP-E64B5426-4FB5-4D41-AFEC-77D84720E886	USA/CA-CDPH-3000131458/2021	OU361297
USA/MA-CDCBI-CRSP_URIKNCKBF5674VF4/2021	USA/NC-CDC-STM-000073018/2021	USA/MA-CDCBI-CRSP_EFLUYAQE2ICW2C6E/2021
CAN/ON-NML-249359/2021	USA/NC-CDC-LC0050688/2021	OU221273
BEL/rega-20174/2021	USA/FL-CDC-LC0065484/2021	England/ALDP-1004959/2021
USA/WA-UW-2181/2020	OU517038	AUS/VIC7772/2020
USA/TX-CDC-ASC210279141/2021	USA/MA-CDCBI-CRSP_6XNVDTWZVHP437HT/2021	USA/AR-CDC-ASC210377904/2021
LR991190	USA/CA-CDC-FG-042304/2021	USA/NJ-CDC-ASC210001462/2021
USA/MD-MDH-0082/2020	USA/AZ-CDC-LC0166866/2021	
USA/UNKNOWN-UW-1403/2020	USA/KS-CDC-2-4449463/2021	
USA/UT-03435/2020	USA/FL-CDC-ASC210078123/2021	
Switzerland/BS-UHB-42535085/2021	OU570677	
Switzerland/SO-ETHZ-531441/2020	USA/MA-CDCBI-CRSP_M2WXUWBLS72TQFUR/2021	
USA/CO-CDPHE-2007261261/2020	USA/OKPHLCOV0000666/2021	
Switzerland/BE-ETHZ-630376/2021	USA/MA-CDCBI-CRSP_S637QBAELKOECBRG/2021	
USA/AK-PHL8614/2021	USA/NY-CDC-LC0344242/2021	
USA/AK-PHL8565/2021	USA/VA-CDC-ASC210374497/2021	
USA/CA-CDC-QDX22169637/2021	Scotland/QEUH-27B5322/2021	
USA/TG278209/2020	USA/TX-CDC-ASC210376640/2021	
USA/CO-CDPHE-2003260110/2020	USA/MN-CDC-LC0334883/2021	
England/MILK-74F4E4/2020	USA/DC-CDC-LC0331077/2021	
Switzerland/ZH-ETHZ-450502/2021	OU150047	
England/MILK-516E26/2020	OU187687	
Iceland/3124/2020	OU149610	
England/QEUH-AD0825/2020	OU125106	

3. Ethics Waiver



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

HUMAN RESEARCH ETHICS
COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Innovation)

TO: Mr F Marsden
School: Pathology
Department: Virology
Medical School
University

E-mail: fabian.marsden@gmail.com

CC: Supervisor: Dr F Treurnicht <Florette.Treurnicht@nhls.ac.za>
and <HREC-Medical.ResearchOffice@wits.ac.za>

FROM: Mr Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 13/08/2021

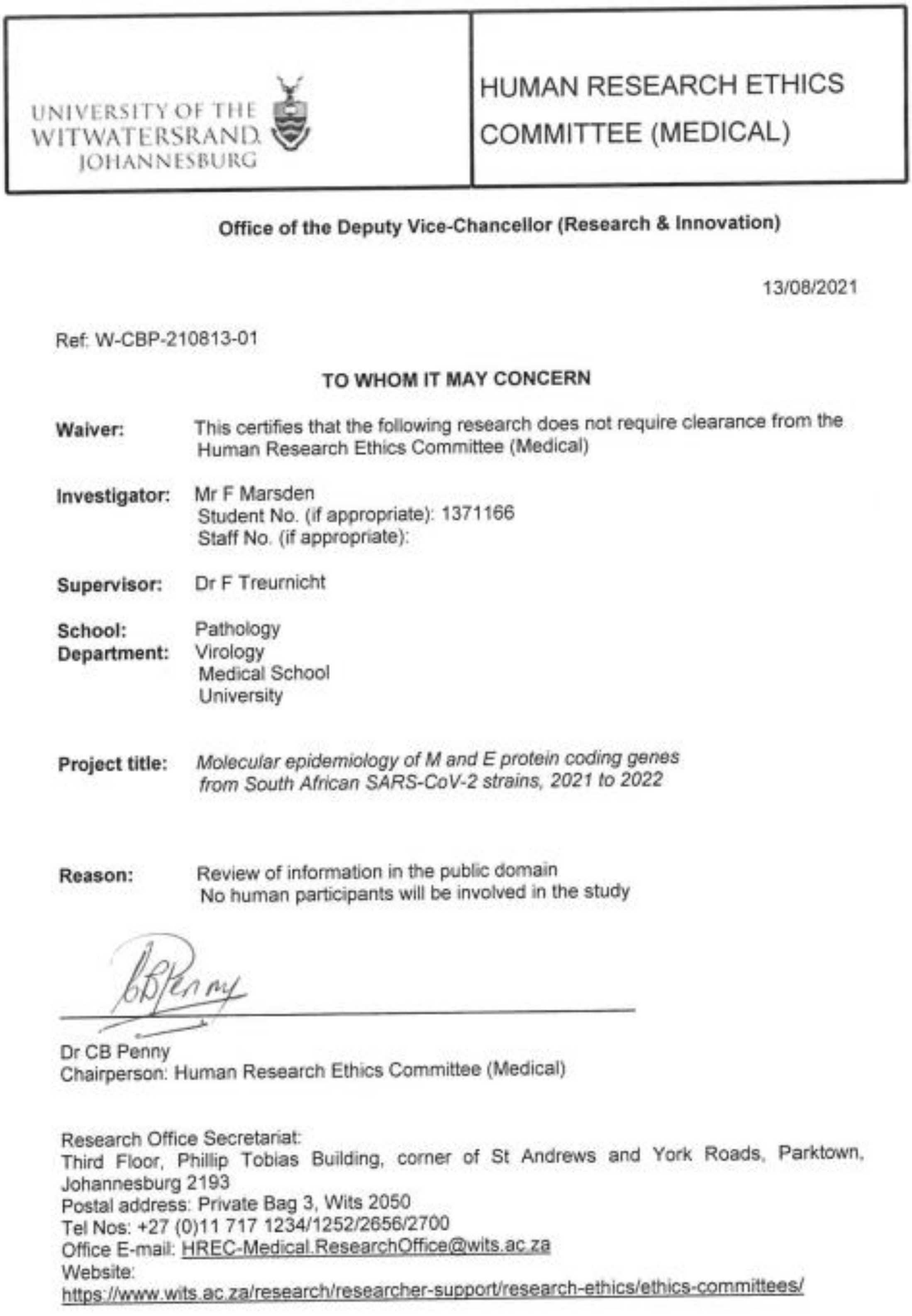
REF: *R14/49*

PROTOCOL NO: W-CBP-210813-01 (This is your ethics application study reference number. Please quote this reference number in all correspondence relating to this study)

PROJECT TITLE: *Molecular epidemiology of M and E protein coding genes from South African SARS-CoV-2 strains, 2021 to 2022*

Please find attached the Ethics Waiver Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to the Government funding of the University.

MSWorks2000/Iain0007/ClearScanWaiver.wps



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

HUMAN RESEARCH ETHICS
COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research & Innovation)

13/08/2021

Ref: W-CBP-210813-01

TO WHOM IT MAY CONCERN

Waiver: This certifies that the following research does not require clearance from the Human Research Ethics Committee (Medical)

Investigator: Mr F Marsden
Student No. (if appropriate): 1371166
Staff No. (if appropriate):

Supervisor: Dr F Treurnicht

School: Pathology
Department: Virology
Medical School
University

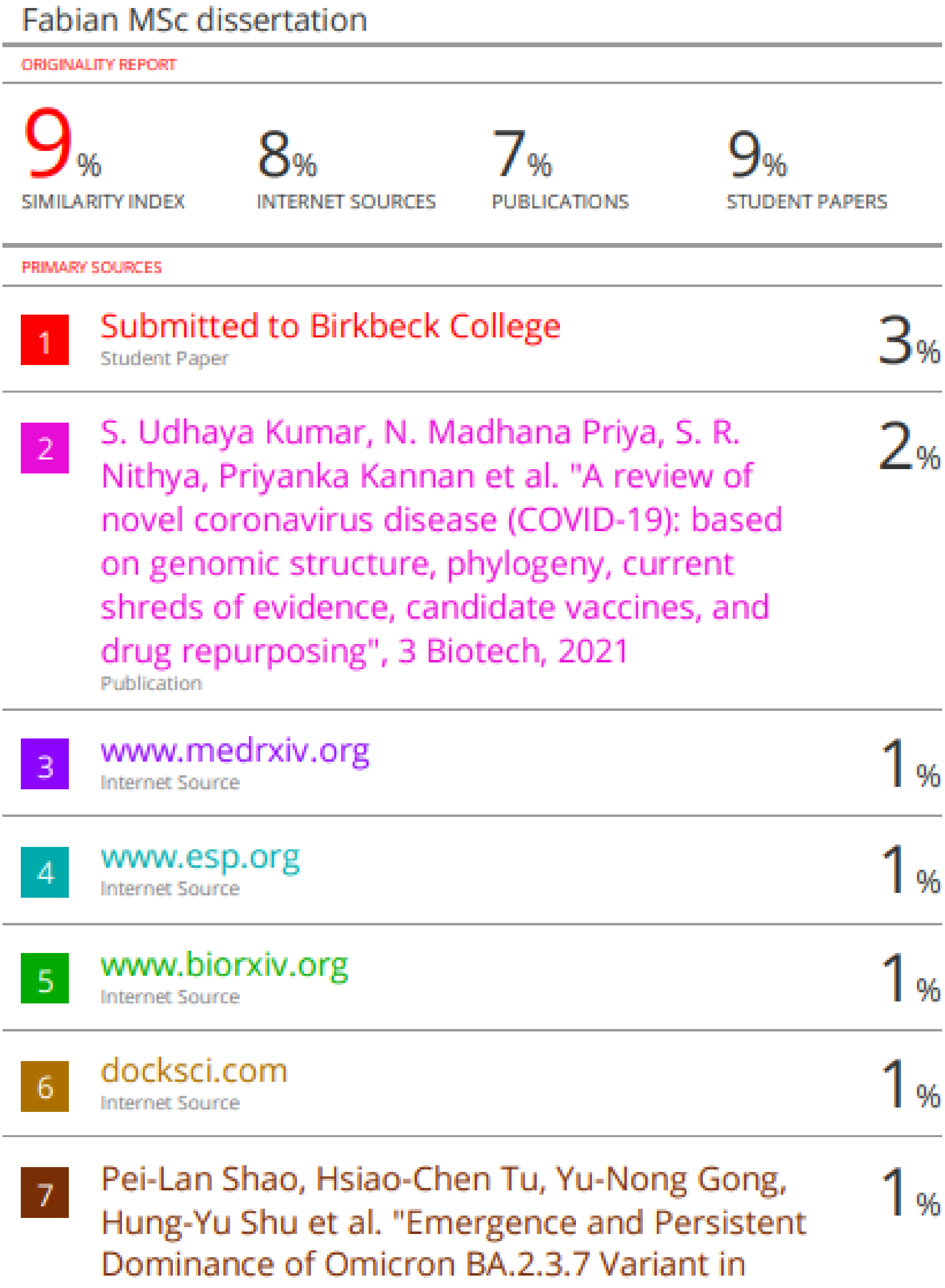
Project title: *Molecular epidemiology of M and E protein coding genes from South African SARS-CoV-2 strains, 2021 to 2022*

Reason: Review of information in the public domain
No human participants will be involved in the study

Dr CB Penny
Chairperson: Human Research Ethics Committee (Medical)

Research Office Secretariat:
Third Floor, Phillip Tobias Building, corner of St Andrews and York Roads, Parktown, Johannesburg 2193
Postal address: Private Bag 3, Wits 2050
Tel Nos: +27 (0)11 717 1234/1252/2656/2700
Office E-mail: HREC-Medical.ResearchOffice@wits.ac.za
Website:
https://www.wits.ac.za/research/researcher-support/research-ethics/ethics-committees/

3. Turnitin Results (excluding <1% similarity)



Fabian MSc dissertation

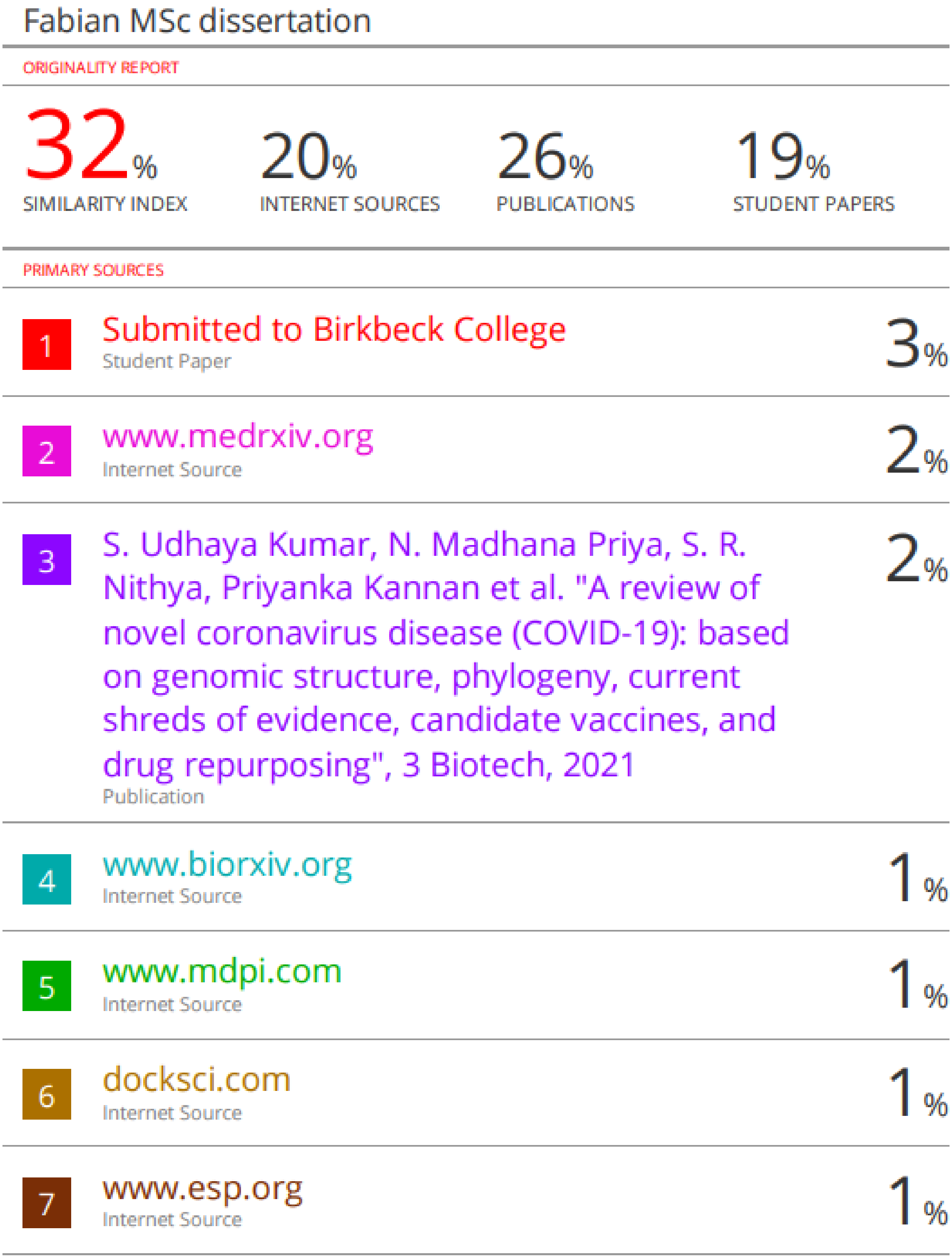
ORIGINALITY REPORT

9%	8%	7%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Birkbeck College Student Paper	3%
2	S. Udhaya Kumar, N. Madhana Priya, S. R. Nithya, Priyanka Kannan et al. "A review of novel coronavirus disease (COVID-19): based on genomic structure, phylogeny, current shreds of evidence, candidate vaccines, and drug repurposing", 3 Biotech, 2021 Publication	2%
3	www.medrxiv.org Internet Source	1%
4	www.esp.org Internet Source	1%
5	www.biorxiv.org Internet Source	1%
6	docksci.com Internet Source	1%
7	Pei-Lan Shao, Hsiao-Chen Tu, Yu-Nong Gong, Hung-Yu Shu et al. "Emergence and Persistent Dominance of Omicron BA.2.3.7 Variant in	1%

4. Turnitin Results (including <1% similarity)



Fabian MSc dissertation

ORIGINALITY REPORT

32%	20%	26%	19%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Birkbeck College Student Paper	3%
2	www.medrxiv.org Internet Source	2%
3	S. Udhaya Kumar, N. Madhana Priya, S. R. Nithya, Priyanka Kannan et al. "A review of novel coronavirus disease (COVID-19): based on genomic structure, phylogeny, current shreds of evidence, candidate vaccines, and drug repurposing", 3 Biotech, 2021 Publication	2%
4	www.biorxiv.org Internet Source	1%
5	www.mdpi.com Internet Source	1%
6	docksci.com Internet Source	1%
7	www.esp.org Internet Source	1%

5. GISAID Acknowledgement and supplemental table

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

SUPPLEMENTAL TABLE

Data Availability

GISAID Identifier: EPI_SET_221130zm

doi: 10.55876/gis8.221130zm

All genome sequences and associated metadata in this dataset are published in GISAID's EpiCoV database. To view the contributors of each individual sequence with details such as accession number, Virus name, Collection date, Originating Lab and Submitting Lab and the list of Authors, visit 10.55876/gis8.221130zm

Data Snapshot

- EPI_SET_221130zm is composed of 26,734 individual genome sequences.
- The collection dates range from 2020-03-06 to 2021-12-31;
- Data were collected in 1 countries and territories;
- All sequences in this dataset are compared relative to hCoV-19/Wuhan/WIV04/2019 (WIV04), the official reference sequence employed by GISAID (EPI_ISL_402124). Learn more at https://gisaid.org/WIV04.