Molecular epidemiology of SARS-CoV-2 in Healthcare Workers at Chris Hani Baragwanath Academic Hospital in South Africa

Jeanine du Plessis

Master of Science in Medicine (Virology)

Student number: 807629

Supervisors: Dr V. Baillie,

Prof S. Madhi, Prof M. Nunes

Candidate's Declaration

I, **Jeanine du Plessis (807629)**, am a student registered for the degree of **Master of Science in Medicine (Virology)** in the academic year **2022.** I confirm that this dissertation is my own, unassisted work and has not been submitted to any other academic institution.

2023/06/12

Candidate's Signature Date

Abstract

South Africa has experienced multiple waves of infections by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), which causes coronavirus disease 2019 (COVID-19). Healthcare workers (HCWs) are at an increased risk of SARS-CoV-2 infections as a result of multiple exposures to infected individuals and contaminated work environments. HCWs may also contribute to nosocomial transmission of SARS-CoV-2 in hospital environments, including between themselves and patients.

We conducted surveillance for SARS-CoV-2 infections in HCW at Chris Hani Baragwanath Academic Hospital (CHBAH). The HCWs were tested weekly, irrespective of symptomatology, for SARS-CoV-2 infections using a nucleic acid amplification test (NAAT). Testing for SARS-CoV-2 was undertaken over three wave periods spanning from April 2020 to September 2021. Furthermore, the HCWs were encouraged to come forth for testing at any time if they displayed symptoms of a respiratory illness or were in close contact with a confirmed case of COVID-19. The NAAT reactive samples were sequenced to characterise the molecular epidemiology of positive cases and better understand virus transmission between HCWs at CHBAH. We defined HCW-HCW transmission if HCWs tested positive for SARS-CoV-2 within 3 to 10 days of one another, worked in the same or related departments and if HCW sequences were isogenic. However, we could not exclude that the infections could be due to potential viral transmission from patients to HCWs or that the HCW contracted the virus elsewhere.

Overall, 193 (35.2% of 549) HCWs had a reactive SARS-CoV-2 NAAT during the study. Of the 193 samples from HCW in which SARS-CoV-2 was detected, 85.0% (N=164/193) were successfully sequenced. We identified nine putative transmission clusters among HCWs using phylogenetic characterisation and single nucleotide polymorphism (SNP) analysis, including four clusters in the first and third waves and one cluster in the second wave of SARS-CoV-2 infections. Twenty-five percent of infections among the HCW (N=41/164) were part of these clusters of infections based on phylogenetic analysis. However, based on the HCW-HCW transmission

definitions, only 11 (26.8% of 41) of the infections were associated with probable transmission between HCW.

We report modest transmission of SARS-CoV-2 infections between HCWs in our setting based on phylogenetic sequencing. The early detection of SARS-CoV-2 infection due to the active surveillance, could have led to a lower risk of SARS-CoV-2 infections among HCWs as those who were infected were identified timeously and went into isolation thereafter. The use of genomic sequencing enabled us to quantify the role of transmission between HCWs, in the overall burden of SARS-CoV-2 infections that affected this population.

Acknowledgements

I would like to thank the following individuals for their contributions in the completion of this dissertation.

1. To my supervisor Vicky, there are no words to describe the massive amount of respect that I have for you. Thank you for guiding me during the pandemic, encouraging me during this entire process, and editing this dissertation even when you were on maternity leave. You continue to inspire me as a powerhouse female in science, thank you.
2. To my co-supervisors Prof Madhi and Prof Nunes, thank you for providing me the opportunity to embark on this educational journey and enabling me to learn so much throughout the process.
3. To my lab partner Lara, thank you for supporting me throughout the experimental element of my project and for helping me find samples. Thank you for being at every journal club presentation and always bringing a positive energy, especially during the difficult times of this process.
4. To my siblings Charl, Tess, Wy, Brandon, and Tanja, thank you for always encouraging me to finish my MSc and for always being my biggest supporters in life.
5. To my partner Andre, a gigantic thank you for helping me with the bioinformatics in the project and allowing me to use your laptop during this process. Thank you for supporting me, inspiring me, and standing behind me in the moments of frustration. Your diligence and hard work is infectious. Thank you for being the voice of reason throughout this MSc.
6. To my parents Johan and Theresia, you are the reason why I love the things that I do. Thank you for cultivating my curiosity, for teaching me that anything you set your mind to is achievable, and for listening even in the moments when you don't understand. I appreciate the years you both have stood behind me and encouraged me to be the scientist that I am today.

Contents

List of Figures

List of Tables

List of Abbreviations

ACE 2	Angiotensin-converting enzyme 2
BAM	Basic alignment map
bp	basepair
C	Cluster
CDC	Centers for Disease Control and Prevention
cDNA	complementary DNA
CHBAH	Chris Hani Baragwanath Academic Hospital
COG-UK	COVID-19 Genomics UK Consortium
CoV	Coronavirus
COVID-19	Coronavirus Disease 2019
Ct	Cycle threshold
del	deletion
DNA	Deoxyribonucleic acid
DPP4	Dipeptidyl transferase 4
E	Envelope
EUA	Emergency use authorization
GISAID	Global initiative of sharing all influenza data
HCoV	Human coronavirus
HCW	Healthcare Worker
HREC	Human Research Ethics Committee
IC	Ectodomain
ICU	Intensive care unit
IPC	Infection prevention control
KZN	KwaZulu Natal
M	Membrane
mAb	Monoclonal antibody
MERS-CoV	Middle Eastern Respiratory Syndrome Coronavirus
N	Nucleocapsid
N1/N2	Nucleocapsid genes 1/2
NA	Nucleic acid
NAAT	Nucleic acid amplification test
NS	Nasal swab
nsp	non-structural protein
NTD	N-terminal domain
ORF	Open reading frame
Pangolin	Phylogenetic Assignment of Named Global Outbreak Lineages
PCR	Polymerase chain reaction
PPE	Personal protective equipment
QC	Quality control
RBD	Receptor binding domain
RBM	Receptor binding motif
RNA	Ribonucleic acid
RSV	Respiratory Syncytial Virus
Rt	Reproductive number
S	Spike
S1/S2	Spike subunit 1/2
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus

SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SGTF	S-gene target failure
SNAP	Swift Biosciences' Normalase Amplicon Panel
SNP	Single nucleotide polymorphisms
TB	Tuberculosis
TM	Transmembrane
TNA	Total nucleic acid
TRS	Transcription regulatory sequence
TTI	Test, trace and isolate
UK	United Kingdom
VIDA	Vaccines and Infectious Disease Analytics
VOC	Variant of concern
VOI	Variant of interest
WGS	Whole genome sequencing
WHO	World Health Organisation

1.Introduction

1.1 Coronavirus Disease 2019 (COVID-19)

In December 2019, an outbreak of viral pneumonia was identified in Wuhan, China (1), which was determined to be caused by a novel coronavirus (CoV) species named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (1, 2) . The disease caused by SARS-CoV-2 is known as coronavirus disease 2019 (COVID-19). Since the start of the COVID-19 pandemic in March 2020, there was universal spread of the virus with 674 million confirmed cases and 6.8 million confirmed deaths attributed to the virus as of 23rd February 2023 (2). Sero-surveys have indicated there has been substantial under-diagnosis of infections, with SARS-CoV-2 seropositivity reportedly being ~90%, even in unvaccinated individuals (3).

Infections by SARS-CoV-2 can be asymptomatic, whilst COVID-19 ranges from being mild illness to being fatal (4). Risk factors associated with severe COVID-19 include age >65 years, immunosuppressive conditions, diabetes, hypertension and obesity (5). Further, underlying communicable diseases such as human immunodeficiency virus (HIV) and tuberculosis (TB) has been associated with higher case fatality risk among hospitalized patients in low to middle income countries, including in South Africa (6). South Africa has a HIV prevalence between 6.8-20.4% (7, 8) and in 2020 accounted for 3.6% of the global burden of TB (9).

1.2 Biological and epidemiological comparison of coronaviruses

Coronaviruses are single-stranded, enveloped, positive sense RNA viruses that form part of the Coronaviridae family and infect humans and animals (5). There are four known genera of coronaviruses, two of which are capable of infecting humans, namely *alphacoronaviruses* and *betacoronaviruses*. Infections of humans by coronavirus may manifest with a range of symptoms, including involvement of the gastrointestinal, cardiac and respiratory tracts systems (5). The *deltacoronaviruses* and *gammacoronaviruses* primarily infect avian species (5).

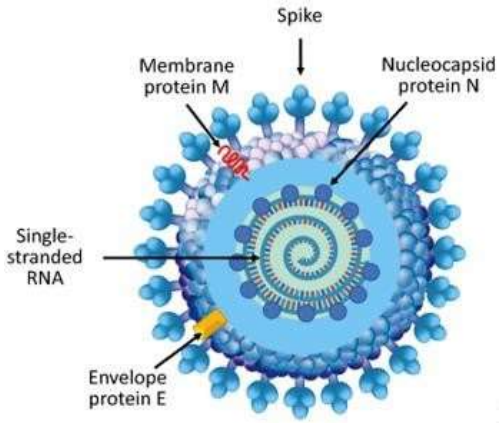


Figure 1. Schematic of a coronavirus (CoV). Membrane, Envelope and Spike structural proteins are embedded in a viral envelope which contains the nucleocapsid protein that forms a framework around single-stranded RNA. Figure from Rossi *et al* (10).

Coronaviruses have the largest genomes (approximately 26.4-31.7Kb) of all known RNA viruses, with a guanine-cytosine (GC) ratio varying between 32-43% (11). The CoVs' genome consists of at least six open reading frames (ORFs); replicase proteins are encoded by ORF1a/b and the four structural proteins are namely membrane (M), envelope (E), nucleocapsid (N) and spike (S) proteins (12). The viral genome is bound to the nucleocapsid, which is embedded in a lipid envelope in which the M, E and S proteins are found; Figure 1 (10).

The M and E proteins maintain the shape, size and integrity of the virus structure and the S protein interacts with the host cellular receptors. (13) The S protein is responsible for the characteristic coronavirus crown appearance (14). The S protein allows entry into host cells through membrane fusion, it consists of two subunits namely S1 and S2; Figure 2 (15, 16). The S1 subunit recognises and binds to the host receptors. Once S1 is bound to a host receptor the S2 subunit is exposed, cleaved by proteases and thus enables membrane fusion of host and virus (15). The S proteins differ in structure among the different coronavirus species and is a virulence factor (17).

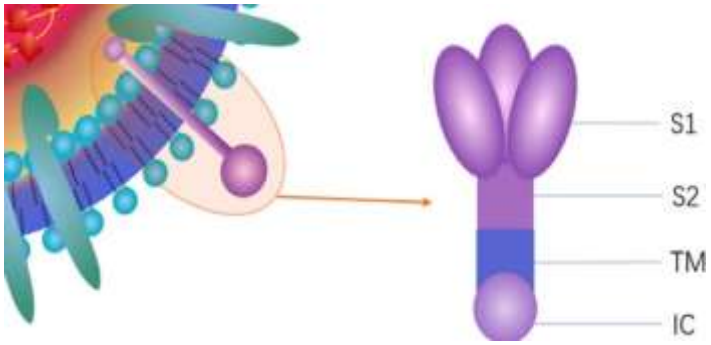


Figure 2. SARS-CoV-2 spike protein structure. The spike protein is characterised by two subunits, subunit 1 (S1) facilitates receptor recognition and contains the receptor binding domain (RBD); subunit 2 (S2) mediates virus-host cell entry through fusion following protease activation. Figure from Heald-Sargent *et al* (16).

1.3 Human coronaviruses

Human coronaviruses (HCoV) were first discovered in 1965 and are frequently associated with mild upper-respiratory tract viral infections. The endemic HCoV include 229E, OC43, HKU and NL63; are detected globally (5, 17). The origin of most HCoV trace back to zoonotic hosts especially bats (18). Usually, HCoV are non-pathogenic in their natural reservoir; however, after interspecies transmission they may become more pathogenic (13).

In 2002/2003, severe acute respiratory syndrome CoV (SARS-CoV) was detected in Guangdong Province, China and resulted in 774 deaths with a case fatality risk of 9-11% (2). SARS-CoV spread to 26 countries, but was subsequently contained after 4 months, with no new reported cases since 2004. In 2012, Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV) was detected in Saudi Arabia and resulted in 858 deaths (15), with a high case fatality rate (35%) (2,16). Since the discovery of MERS-CoV, it has spread to 27 countries with intermittent detection of few additional cases that is ongoing (52).

In December 2019, a group of people who were somewhat associated with the Haunan seafood market in China presented with an unidentified form of viral pneumonia (1). The unknown virus was genetically sequenced from lower respiratory tract specimens and was later termed SARS-CoV-2 (1). Viral transmission occurred internationally within months of its first identification despite containment efforts initiated by public health authorities in China. On the 30th of January 2020 the World

Health Organization (WHO) declared SARS-CoV-2 a Public Health Emergency of International Concern (1), and subsequently declared a COVID-19 pandemic on March 11, 2020.

1.4 SARS-CoV, MERS-CoV and SARS-CoV-2 epidemiology

The human host receptor for MERS-CoV is dipeptidyl transferase 4 (DPP4) (19), which is restricted to bronchiolar epithelial cells (20). In contrast, SARS-CoV and SARS-CoV-2 use the angiotensin-converting enzyme 2 (ACE2) receptors (21), which are located on the surface of most host epithelial cells (22). Although the S1 subunit is conserved between SARS-CoV and SARS-CoV-2, SARS-CoV-2 has a greater affinity for ACE2 due to variations in the receptor binding domain (RBD) and cleavage of the S1-S2 site during biosynthesis (18, 23). These differences in affinity for the host receptors allows SARS-CoV-2 to rapidly spread among humans with a reproductive rate (R0) ranging from 1.4 to 6.49 (23). The R0 value is an epidemiological concept used to assess the spread of infectious diseases in a susceptible population with no immunity and is influenced by a combination of several factors: the number of daily contacts an infectious person has, the likelihood of transmission during such contacts (depends on the social habits of a population) and the number of days the individual is infectious (depends on the biological characteristics of the disease) (20). An R0 greater than 1 implies that transmission of infection in populations with no immunity will spread exponentially (24-26). A population, however may develop a level of immunity to disease through the evolution of infection (natural acquired immunity) and through vaccine-induced immunity; which is then measured as the effective reproductive rate (Rt) (20, 27-29).

In individuals infected with SARS-CoV-2, the highest viral load is detected around or immediately prior to symptom onset (30); as such, traditional symptom-based public health regulations did not prove as effective to prevent SARS-CoV-2 outbreaks. Furthermore, SARS-CoV-2 is efficiently transmitted in the pre-symptomatic phase and by asymptomatic individuals (31). Test-trace and isolate (TTI) strategies prove to be taxing in the mitigation of the spread of SARS-CoV-2 due to pre and asymptomatic transmission of the virus (32).

1.5 SARS-CoV-2 mutations and variants

Whole genome sequencing (WGS) of the viral genome has been performed internationally to characterise SARS-CoV-2 (33). Sequencing of SARS-CoV-2 has yielded important findings including the detection of novel variants of the virus (33). Due to a proofreading ability to correct replication errors enzymatically, CoVs usually have a decreased evolutionary rate in comparison to other RNA viruses such as Respiratory Syncytial Virus (RSV) subtype A and B. Coronaviruses compared with RSV have approximated mutation rates of 2.1×10^{-3} substitutions per site per year (s/s/y) and 3.03×10^{-3} s/s/y, respectively (34). Nevertheless, there has been significantly higher than expected mutation rates for the whole genome of SARS-CoV-2, which is estimated at 6.677×10^{-4} s/s/y (35). The increased mutation rates is possibly due to evolving immune pressure, and has been associated variably with heightened transmissibility and virulence (31). SARS-CoV-2 also mutates continuously during human-to-human transmission (23), and particularly when there is prolonged shedding in the same host such as immunocompromised individuals . Mutations that result in competitive advantage in terms of transmission, immune escape and viral replication typically outlast those that reduce viral fitness and tend to become the predominant form in the viral population (33, 34).

The need for public databases arose during the progression of the pandemic as the amount of genomic data related to SARS-CoV-2 exponentially increased. Web resources including Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin) and NextClade were widely used by researchers worldwide for the lineage assignment, analysis and visualization of SARS-CoV-2 data (36). These databases utilize a phylogenetic framework to classify lineages into a virus nomenclature that is ubiquitously used by researchers. In the SARS-CoV-2 pandemic a universally recognized nomenclature system was imperative in the real-time description of the virus epidemiology (36). Additionally, defined descriptions for SARS-CoV-2 lineages enabled links between various geographical regions to be inferred through shared viral genomes (36).

The WHO, COVID-19 Genomics UK Consortium (COG-UK) and US Centers for Disease Control and Prevention (CDC) have developed a classification for SARS-CoV-2 variants that have emerged during the progression of the pandemic (37-39); i.e. variants of interest (VOI) and variants of concern (VOC). VOC are defined as strains which have evidence of being more transmissible, capable of causing severe disease and result in reduction of neutralization antibody activity induced by prior infection or prototype vaccines (38, 40). Variants of interest refer to mutations that are considered to be of potential importance but do not yet meet the criteria to be termed a VOC (39, 40). Five VOC have been recorded to date by WHO, namely Alpha (B.1.1.7), Beta (B.1.352), Gamma (P.1), Delta (B.1.617.2) and Omicron (B.1.1.529) (Table 1).

Table 1. Classifications of Variants of Concern (VOCs) and Variants of Interest (VOI) including common amino acid mutations (February 2023) (43).

WHO Classification	PANGO lineage	VOC/VOI	Initial country of identification	Date of first detection	*18	*95	*144	*242	*417	*452	*478	*484	*501	#nsp6:106	#N:203/204
					^Leucine	^Threonine	^Tyrosine	^Leucine	^Lysine	^Leucine	^Threonine	^Glutamic Acid	^Asparagine	^Serine	^Arginine/ Glycine
Alpha	B.1.1.7	VOC	UK	Dec-20			del						Tyrosine	del	Lysine/Arginine
Beta	B.1.351	VOC	South Africa	Sep-20	Phenylalanine			del	Asparagine			Lysine	Tyrosine	del	
Gamma	P.1	VOC	Brazil	Dec-20	Phenylalanine				Threonine			Lysine	Tyrosine	del	Lysine/Arginine
Delta	B.1.617.2	VOC	India	Dec-20		Isoleucine				Arginine	Lysine				Methionine
Omicron	B.1.1.529	VOC	South Africa	Nov-21			del				Lysine	Alanine	Tyrosine		
Epsilon	B.1.427/9	VOI	USA	Jul-20						Arginine					
Eta	B.1.525	VOI	Nigeria	Dec-20			del					Lysine		del	
Theta	P.3	VOI	Philippines	Feb-21								Lysine			Lysine/Arginine
lota	B.1.526	VOI	USA	Nov-20		Isoleucine	del					Lysine		del	
Kappa	B.1.617.1	VOI	India	Dec-20		Isoleucine				Arginine		Glutamine			Methionine
Zeta	P.2	VOI	Brazil	Apr-20								Lysine			Lysine/Arginine
Lambda	C.37	VOI	Peru	Aug-20						Glutamine				del	Lysine/Arginine

*Nucleotide positions in the spike protein.
Non-spike mutations in the nonstructural protein 6 (nsp6) and nucleocapsid genes.
^ Amino acid residue present in the reference sequence.
Deletions are indicated as 'del'.

The first major SARS-CoV-2 genetic mutation detected in early March 2020 was the spike D614G mutation, which rapidly resulted in the mutated virus becoming the dominant virus globally within months of being identified (34). The D614G mutation showed convergent evolution, as there was independent development of D614G in all subsequent emerging variants of SARS-CoV-2 (Table 2). The D614G mutated virus was linked to an increase in transmissibility through enhanced binding affinity to ACE2 receptors (34).

In September 2020, the B.1.1.7 variant, also referred to as Alpha VOC, was identified in England and rapidly spread throughout the United Kingdom (UK) (34). By late December 2020 it accounted for 28% of all active cases in the UK and was estimated to be 56% more transmissible than the wild-type virus (34). The genetic mutations associated with the Alpha VOC is shown in Table 1. Davies *et al* reported that Alpha VOC was 43-90% more transmissible than ancestral strain (41); which was likely due to a greater number of open receptor binding domains (RBDs) from the D614G mutation coupled with the N501Y mutation (41). Increased receptor affinity of SARS-CoV-2 results in higher virus load in the upper respiratory tract, as was demonstrated for Alpha VOC (41). S- gene target failure (SGTF) was observed for the Alpha VOC, which is attributed to N-terminal domain (NTD) deletions at position 69-70, which prevents amplification of specific genomic regions that are normally used in some diagnostic nucleic acid amplification test (NAAT) assays (42).

In September 2020, the B.1.351 variant was identified in South Africa, which was later classified as the Beta VOC (43). The mutations associated with Beta VOC are detailed in Section 1.6, which describes the molecular evolution of SARS-CoV-2 in South Africa.

In early January 2021, the discovery of a new SARS-CoV-2 variant, later designated as Gamma VOC, was reported in Japan. The Gamma VOC that likely emerged from Brazil around November 2020 (44); has three mutations in the RBD; i.e. K417T, E484K and N501Y (Figure 3) and is distinguishable from the ancestral strain by 12 amino acid mutations in the S (44). E484K and N501Y mutations have been frequently linked with increased binding affinity of the S to ACE2 (44). Mutations

such as E484K are capable of evading neutralizing antibody activity induced by wild-type infection and first generation COVID-19 vaccines (45) (Table 2). The antibody evasiveness induced by E484K is through the stabilization of the RBD-down conformation which inhibits binding of neutralizing antibodies (46). Due to advantageous conformational changes coupled with antibody escape, continuous monitoring of E484 is important, as emphasized by the evolution of the E484Q mutation seen in the Delta variant (46).

In late 2020, the Delta VOC was identified in Maharashtra, India and resulted in a resurgence of COVID-19 cases and deaths globally (47). The Delta VOC was less sensitive to ancestral virus infection and vaccine induced neutralizing antibody activity compared with strains bearing the D614G mutation (47). The Delta VOC demonstrated enhanced replication and cellular entry ability compared with the B.1.617.1 (Kappa) variant (47). The Delta VOC has a characteristic spike receptor-binding motif (RBM) namely L452R (Figure 3), which is associated with increased infectivity and diminished effect of neutralizing antibodies (Table 2) (47). During a time of mixed variant circulation, the ChAdOx-1 vaccine showed reduced efficiency against infection with the Delta variant in comparison with non-Delta variants (47).

On the 24th November 2021 a new variant was reported by South Africa, later designated as the Omicron VOC (B.1.1.529) (48, 49). The Omicron VOC contains the largest number of mutation sites in comparison with other SARS-CoV-2 variants thus far (50). Omicron VOC has furin cleavage site mutations that enhance its infectivity (Table 2). The origin of Omicron is still largely unknown and since the discovery of Omicron, multiple sublineages of the variant have been identified including BA.1, BA.2 and BA.3 (51). The BA.1 sublineage was responsible for the surge of infections resulting in South Africa's fourth pandemic wave; however, BA.2 became the dominant circulating sublineage in the latter stages of the fourth wave (52, 53). A shift to the dominant BA.2 lineage was not associated with substantial high rates of COVID-19 hospitalizations or deaths in South Africa, albeit being associated with severe disease in many other countries (52, 53). Two additional lineages of Omicron with identical spike proteins are the more recently identified BA.4 and BA.5 sub-lineages (54). The BA.4 and BA.5 lineages share similar genetic traits to BA.2 apart from a 69-70 deletion in the spike protein, which allowed for preliminary

identification via S-gene target failure (54). BA.4 and BA.5 have accounted for 50% of all the sequenced SARS-CoV-2 cases in South Africa since April 2022 (55). The infections associated with Omicron and its sublineages has proven to be less severe than those associated to Delta, despite the variants capability to evade immune responses (56).

The perplexing scenarios resulting from the continued emergence of novel SARS-CoV-2 variants cannot be mitigated as long as transmission of SARS-CoV-2 persists, which emphasizes the importance of ongoing surveillance and sequencing of SARS-CoV-2.

Table 2. Mutations identified in various VOC/VOI and the associated functionality of each mutation.

Mutation	Location	Functionality	Variants found in
D614G	RBD	Increased transmissibility	All VOC and VOI
N501Y	RBD	Increased affinity to ACE2	Alpha, Beta, Gamma, Omicron
		Increased viral replication in human upper-airway cells	
		Reduced susceptibility to convalescent and vaccine plasma	
E848	RBD	Evade immune responses	Beta, Gamma, Eta, Iota, Theta, Zeta,Kappa, Omicron,Alpha variant sub-lineages
L452R	RBD	Reduced susceptibility to mob's	Delta, Kappa
		Reduced susceptibility to convalescent and vaccine plasma	
		Increased cellular entry into lung organoids	
K417N/T	RBD	Reduced ACE2 binding, evade immune responses	Beta, Gamma, Omicron
Y453F	RBD	Increased affinity to ACE2	Mink lineages, cluster 5
S477N	RBD	Increased affinity to ACE2	20A.EU2, Omicron
S477N	RBD	mAb escape	20A.EU2, Omicron
T478K	RBD	Evade immune responses	Delta, Omicron
Δ69-70	NTD	Increased virus replication	Alpha, Eta
		Increased infectivity	Alpha, Eta
Δ141-146	NTD	Interference of neutralizing activity of NTD-binding antibodies	Alpha, Eta, Lota
Δ242-244	NTD	Interference of neutralizing activity of NTD-binding antibodies	Beta
L18F	NTD	Reduced effect of NTD-neutralizing antibodies	Alpha, Beta
P681H	S1/S2 furin cleavage site	Incorporation of basic aa enhancing S1/S2 cleavage and fusion	Delta, Kappa, Omicron
		Increased infectivity	
P681R	S1/S2 furin cleavage site	Virus tropism by enhanced S1/S2 cleavage in airway epithelial cells, increased infectivity	Theta, Alpha
D3L	N gene	Enhanced transcription regulatory sequence (TRS) of Orf9b	Alpha
Δ106–108	nsp6	Unknown phenotypic consequence	Alpha, Beta, Gamma, Eta, Iota, Lambda

Abbreviations: RBD: Receptor binding domain, ACE: Angiotensin converting enzyme 2, Orf9b: open-reading frame 9b, NTD: N-terminal domain, S1/S2: subunit 1/subunit 2, mAb: monoclonal antibody.

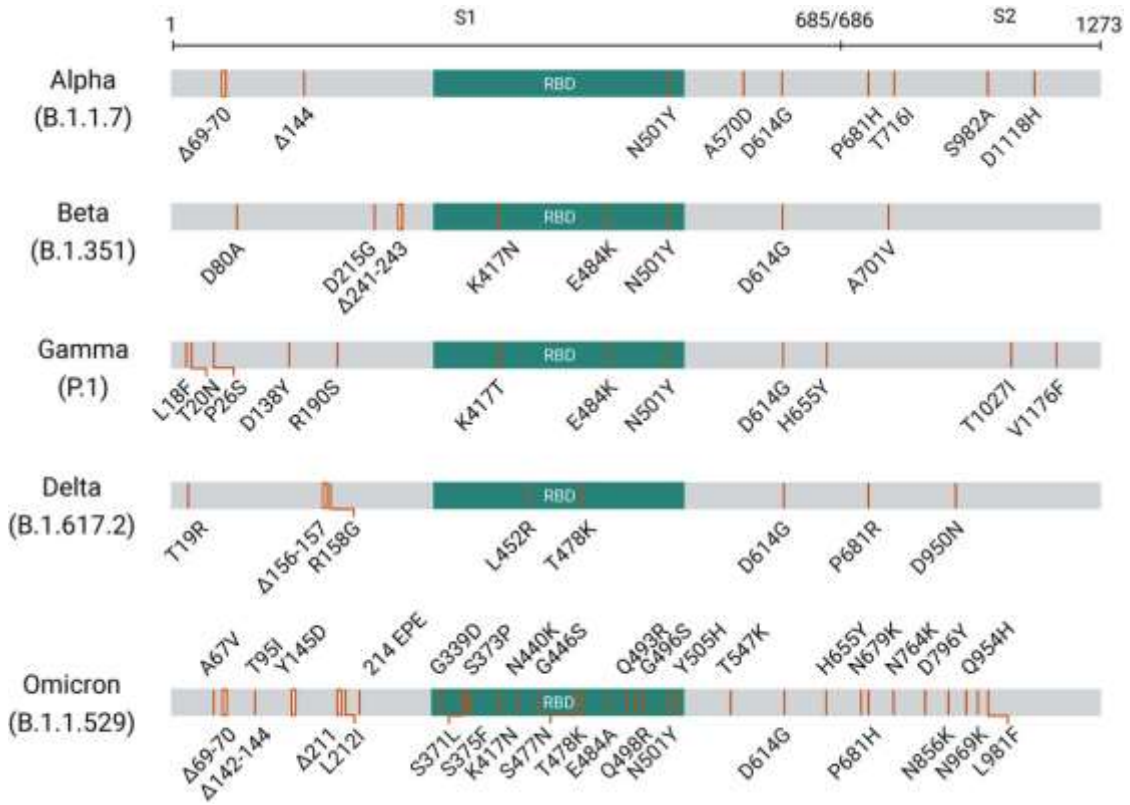


Figure 3.Spike mutations present in SARS-CoV-2 variants of concern (VOC). A visual representation of the shared and unique S mutations that are present in the five VOCs namely Alpha, Beta, Gamma, Delta and Omicron. Figure from Xuemei et al (52).

1.6 Molecular evolution of SARS-CoV-2 in South Africa

On the 5th March 2020 the index case of SARS-CoV-2 was detected in KwaZulu Natal (KZN), South Africa, in an adult patient who had returned from a trip to Italy. By the 26th March 2020, full lockdown restrictions were implemented countrywide in South Africa (28). The majority of infections during the first COVID-19 wave in South Africa (from March to October 2020) were due to circulation of three monophyletic clusters, namely lineages B.1.1.54 (20B), B.1.1.56 (20B) and C.1 (20D) (57, 58).

Phylogeographic analysis revealed that B.1.1.56 was first detected in Durban around mid-March 2020, and rapidly spread throughout the KZN province from June 2020 onwards as lockdown restrictions were eased in South Africa (57). The C.1 lineage was first detected in late May 2020 in Johannesburg (Gauteng Province), and subsequently dispersed to the bordering North West Province. The C.1 lineage virus was associated with large nosocomial

outbreak at the Witrand Hospital, which cares for mentally disabled and psychiatric patients (57, 59, 60). The index case was a patient in ward 03 who tested SARS-CoV-2 positive on the 10th of June 2020 (59)The nosocomial outbreak involved two wards at the hospital, namely ward 03 and 04, and resulted in 100 patients and 6 nurses being infected (60). The B1.1.54 lineage was first detected in mid-May 2020 in KZN (57) and then spread to the North West and Free State provinces (57). Mutations that distinguish the three lineages are shown in Figure 4 (33). All three lineages have mutated from the parental Wuhan lineage (B.1.1) and include the D614G missense mutation (61).

.

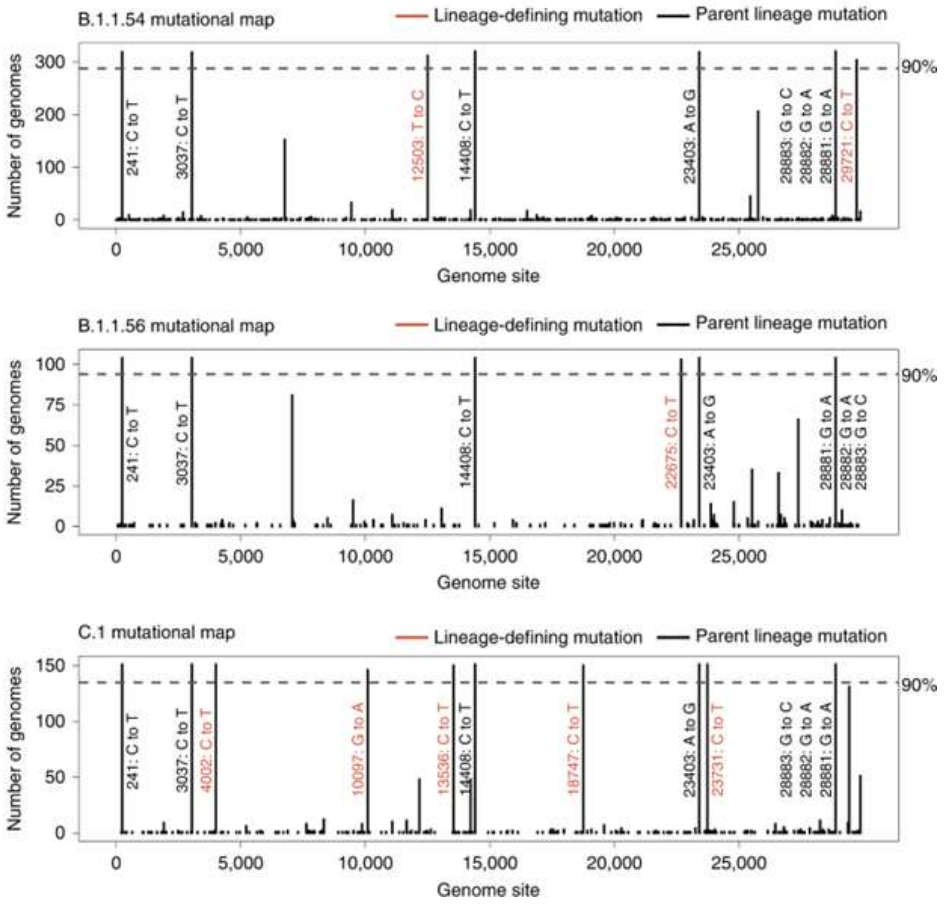


Figure 4.Variant map with the mutations identified during the first wave of infection in South Africa. Lineage-defining mutations against the original parent SARS-CoV-2 genome (B.1.1). Lineage-defining mutations are classified as mutations or single nucleotide polymorphisms (SNPs) that exist in 90% or more of the genomes in that cluster. Figure from Tegally et al (57).

As shown by the variant map (Figure 4), the lineage defining mutations in B1.1.54 are the 29721C>T and the 12503T>C mutations. The 12503T>C mutation is responsible for

substitution of tyrosine (Y) to histidine (H) at position 138 (Y138H) in the non-structural protein 8 (nsp8) (61). In the B.1.1.56, a nonsynonymous mutation attributable to a 22675C>T nucleotide substitution is present (61). Missense mutations in nsp3 (T428I) and 3C-like proteinase (G15S) are observed in the C.1 variant and correspond to nucleotide substitutions 4002C>T and 10097G>A, respectively (61). None of these nonsynonymous mutations have been associated with changes in phenotypic characteristics.

Sequencing of samples collected from patients in the Eastern Cape Province in early September 2020, revealed a novel variant of SARS-CoV-2, that drove the second pandemic wave in South Africa; referred to as the Beta (B.1.351) variant (43, 58). By the end of 2020, Beta accounted for >95% of the SARS-CoV-2 infections in South Africa (Figure 5) (58). This variant has multiple nonsynonymous mutations (K417N, E484K and N501Y) affecting the RBD functional sites (Table 2) (43) and has five additional spike mutations that are of less concern (L18F, D80A, D215G, R246I and A701V) as there is no phenotypic change associated with these mutations (62). Away from the spike region the Beta variant carries the following mutations: SGF 3675-3677 deletion, T2051I, K1655N and P71L (62). Beta was designated as a variant of concern (VOC) due to its increased transmissibility and adaptations to the host immune responses (39, 63). *In vitro* experiments reveal that neutralizing antibodies are not affected by the N501Y mutation (43, 64, 65).

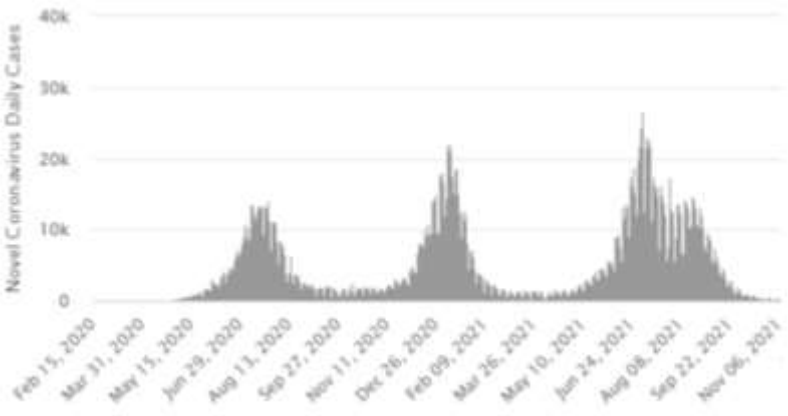


Figure 5. Reported national numbers of daily SARS-CoV-2 infections. The number of daily new SARS-CoV-2 infections from February 2020 to September 2021 demarcating the three waves of the South African epidemic. Figure from Worldometer (66).

In May 2021, the C.1.2 lineage was first identified and was derived from the C.1 lineage. The C.1.2. lineage was highly mutated with an estimated mutation rate of 3.04×10^{-3} nucleotide

changes/site/year. The C.1.2 lineage has mutations similar to those identified in Alpha, Beta and Delta VOC (67). C.1.2 has two additional RBD mutations that have not been detected in other VOIs/VOCs namely Y449H and N440K; these mutations have been associated with neutralizing antibody escape (67, 68) Nevertheless, past infection with Beta and Delta variants demonstrated neutralizing antibody activity against C.1.2 infections (67). Although C.1.2 has a high substitution rate and shares mutations that have been linked to increased transmissibility, variant levels remain low and C.1.2 hasn't been classified as a VOC/VOI (67).

1.7 SARS-CoV-2 infection among Healthcare Workers

Healthcare workers have a higher likelihood of contracting SARS-CoV-2 as a result of exposure to infected individuals and contaminated work environments. Additionally, social distancing practices are often not achievable when caring for patients (69, 70). Thus, HCWs have a disproportionately higher rate of SARS-CoV-2 infection in comparison with the general population (39). The use of personal protective equipment (PPE) can reduce the risk of SARS-CoV-2 infection. Nevertheless, there was a severe shortage of PPE in resource-constrained countries, especially during the first COVID-19 wave. Further, HCWs might also be more likely to have severe COVID-19 due to repeat exposures to large inoculum doses of SARS-CoV-2 (71). In August 2020, South Africa reported over 27 000 cases and 240 deaths among HCWs associated with COVID-19 (40,41). HCWs may also contribute to nosocomial spread of SARS-CoV-2 between themselves, as well as patients. Studies revealed that nosocomial transmission of SARS-CoV-2 from HCW-to-HCW may account for 89% of infections in HCWs during the early stage of the pandemic (69). Patient-to-HCW transmission was determined to be the primary route of infection for the SARS-CoV outbreak among HCWs (72), and patient-to-patient transmission was evident in the MERS-CoV epidemic (72, 73). The number of SARS-CoV-2 cases among HCWs derived from nosocomial transmission or community acquisition is uncertain, as identification of the source of infection is challenging due to the continuous movement of HCWs between a hospital and community setting (73). Contact tracing is resource-intensive, however, it enables establishing the origin and transmission chain of outbreaks with genomic sequencing (75).

Knowledge on the exposure of HCWs to SARS-CoV-2 is important in identifying virus transmission patterns and understanding the effectiveness of infection prevention control (IPC) practices. Epidemiological studies and surveillance of HCWs enable early discovery of suspected cases and can assist in informative decision making to minimize SARS-CoV-2 transmission in healthcare facilities (59).

The objective of our study was to characterise the genome sequences of SARS-CoV-2 identified among HCWs and evaluate whether there were clustering of Single Nucleotide Polymorphisms (SNPs), deletions and mutations among HCWs within specific departments at Chris Hani Baragwanath Academic Hospital (CHBAH).

1.8 Study Justification

SARS-CoV-2 infections have caused 6.8 million known deaths worldwide and continued emergence of new variants is of ongoing health concern (2). Monitoring for the emergence of new SARS-CoV-2 variants could assist in informing public health interventions. Healthcare workers are at increased risk of being exposed to SARS-CoV-2 and can potentially develop more severe disease (71).

Surveillance of SARS-CoV-2 infections in HCWs can assist in characterising the virus transmission patterns and evaluation of the effectiveness of IPC practices. Epidemiological studies and surveillance of HCWs enable early detection of suspected cases and can reduce the contribution of HCW to SARS-CoV-2 transmission in healthcare facilities (59). The use of molecular epidemiology of SARS-CoV-2 specifically in the context of HCWs has proven to be critical in the detection of linked transmission events (76). Furthermore, it has been utilized in the targeting of interventions to reduce HCW-associated SARS-CoV-2 infections which in turn allows lowers the burden on healthcare infrastructure, an imperative factor to consider due to heightened medical needs of the public during a pandemic (76). Molecular epidemiology of SARS-CoV-2 allows for the rapid tracking and subsequent restriction of transmission in both hospital and community environments.

2. Aims and Objectives

2.1 Study aim

To determine the molecular epidemiology of SARS-CoV-2, using next generation sequencing, in HCWs in an effort to understand the contribution of virus transmission between HCWs at CHBAH.

2.2 Study objectives

2.2.1 Primary Objective

Characterise the genome sequences of SARS-CoV-2 identified among HCWs, and evaluate whether there were clustering of Single Nucleotide Polymorphisms (SNPs), deletions and mutations among HCWs within specific departments.

2.2.2 Secondary objectives

- Construct a phylogenetic tree of all generated sequences.
- Characterise the nucleotide variations in viral genomes in comparison to the first SARS-CoV-2 published genome.

3. Methods
3.1 Study Population

A longitudinal surveillance study was undertaken by the Wits Vaccines and Infectious Diseases Analytics (Wits-VIDA) research unit to investigate the epidemiology of SARS-CoV-2 infections in HCW at CHBAH (77). CHBAH is located in the Soweto area south of Johannesburg, South Africa, and is the largest tertiary hospital facility in Africa and ranks as the third largest hospital in the world (56). The first reported hospitalised case of COVID-19 at CHBAH was on the 6th of April 2020 (71).

HCWs working at CHBAH were enrolled into the surveillance study during two enrolment periods: the first from 22nd April to 24th July 2020 (Cohort 1) and, the second from 16th February to 9th October 2021 (Cohort 2). A total of 549 HCWs were enrolled across five different departments namely: Internal Medicine (included the treatment of all SARS-CoV-2 infected patients), General Paediatrics and Neonates, Intensive Care Unit (ICU) (included the treatment of all severe SARS-CoV-2 infected patients), Obstetrics and Gynaecology and at Wits-VIDA (Table 3) (77).

As part of the surveillance, nasal mid-turbinate swabs were collected weekly for SARS-CoV-2 testing using a NAAT (77). During the same time period (April 2020 to September 2021), positive SARS-CoV-2 samples collected from individuals enrolled in a COVID-19 vaccine trial, namely the ChAdOx1 nCoV-19 trial (78-80), were sequenced and used as a comparison of what SARS-CoV-2 sequences were circulating in the local population.

Table 3. Number of study participant enrolled per department in the cohorts comprising of the HCW surveillance study (76).

Cohort 1 Enrolments							Cohort 2 Enrolments							
Department	Auxillary /para-medical	Laboratory staff	Medical doctor	Nursing	Field-workers	Total	Department	Auxillary /para-medical	Laboratory staff	Medical doctor	Nursing	Field-workers	Other	Total
Obstetrics & Gynae			22	1		23	Obstetrics & Gynae			1	2			3
General Paediatrics	2		47	42		91	General Paediatrics	1		12	10		2	24
Internal Medicine	2		38	127		167	Internal Medicine	3		7	31		9	50
Wits-VIDA	5	27	6	19	7	64	Wits-VIDA		3			11	9	23
ICU	5		25	21		51	ICU			2	10			12
							Other*	6	1	2	13	0	19	41

*Other enrolments include support services, occupational therapy, physiotherapy, human resources, and admin departments.
ICU= Intensive Care Units, VIDA= Vaccine & Infectious Disease Analytics.

3.2 Study design

A retrospective exploratory sequencing study on SARS-CoV-2 positive samples collected from HCWs during the course of the first three COVID-19 waves was conducted (66). The study defined wave periods were: June 2020-October 2020 (wave 1), December 2020-March 2021 (wave 2) and April 2021-September 2021 (wave 3) (Figure 5) (66).

As part of the surveillance study, nasal mid-turbinate swabs were collected weekly for SARS-CoV-2 testing using a nucleic acid amplification test (NAAT). Furthermore, HCWs were encouraged to come for testing in between scheduled visits if they manifested any symptoms of a respiratory or gastrointestinal illness, as well as symptoms of fever or impaired sense of taste or smell (81). Additionally, participants could screen in between scheduled visits if there was contact with a confirmed COVID-19 case. Participants that tested positive for COVID-19 on NAAT were notified within 24 hours of sampling and had repeat nasal swabs and blood samples collected approximately every 48-96 hours, until they had at least two consecutive NAAT-negative results (81). We also collected questionnaires outlining demographic, health and behavioural information of the participants. The surveillance study comprised of two enrolment cohorts; enrolment into the first cohort occurred between the 22nd April 2020 to the 24th of July 2020. Enrolments into the longitudinal cohort recommenced on 16th February 2021 and concluded on 10th August 2021 as a result of participants discontinuing the study. The second enrolment cohort a less stringent sampling schedule.

3.3 Sample collection, clinical testing, and viral quantification

Nasal swabs (NS) were stored in universal transport media (UTM®; COPAN, Brescia, Italy). The MagaBio plus virus DNA/RNA purification kit II was used to extract a final volume 80µL of total nucleic acid (TNA) from 300µL of vortexed UTM using the Bioer automated extraction system (Hangzhou Bioer Technology Co. Ltd, China). The presence of SARS-CoV-2 RNA in extracted TNA was determined with reverse-transcriptase NAAT using the CDC emergency use authorization developed assays (82).

NAAT results were considered valid if the housekeeping RNP assay had a cycle threshold (Ct) value ≤33, invalid RNP results (Ct > 33) resulted in a retest and repeat swab taken from the participant. NAAT results were designated positive when both the nucleocapsid genes (N1 and N2) were present with Ct values <40; inconclusive when only N1 or N2 was detected with Ct values <40 and negative if N1 and N2 had Ct values> 40. Nucleic acids extracts that had NAAT Ct values less than 35 were archived at -80°C for downstream SARS-CoV-2 sequencing. One nucleic extract per infection episode was selected for whole genome sequencing, selection was done based on the Ct inclusion criteria for sequencing (Ct<35). In circumstances whereby a participant had more than one nucleic extract that fell within the inclusion criteria the lowest Ct sample was selected for next generation sequencing.

3.4 cDNA synthesis and library preparation

First strand cDNA was synthesized from extracted viral RNA using 50 ng/µl random hexamers and Superscript IV one-step RT-PCR system (Life Technologies, Carlsbad, CA). The quality of synthesized cDNA was tested using the CDC SARS-CoV-2 NAAT assays using TaqMan Gene Expression master mix (Thermo Fisher Scientific); and the Ct's of the cDNA were compared to the original extract Ct. Typically Ct's that were within 3-5 Ct's of the original extract was selected for further processing. If cDNA extracts failed the preliminary quality control (QC) NAAT, the original NS sample was re-extracted and cDNA first strand synthesis was repeated. If subsequent QC NAAT failure occurred cDNA first strand synthesis was redone using an alternative extract from the same participant that had a Ct within the inclusion criteria of the study. Swift Biosciences' Normalase Amplicon (SNAP) and SARS-CoV-2 amplicon panel (Swift Biosciences, Ann Arbor, MI, USA) were used to produce sequencing libraries, manufacturer's recommendations were followed (Table 4). Briefly, 341 primers pairs tiling the length of the SARS-CoV-2 genome were combined with the cDNA samples in order to generate amplicons ranging from 116-300 base pairs according to the cycling parameters listed in Table 4. Samples with cDNA Ct's >29 underwent 24 cycles of amplification as per manufacturers' guidelines regarding low viral load inputs. A single negative template control (NTC) containing nuclease-free water and master mix was included per library preparation run to act as a measure of potential contamination during the laboratory sequencing protocol.

Table 4. cDNA amplification NAAT cycling conditions as per manufacturers' recommendations.

Temperature (°C)	Time (s)	Cycles	
98	30	1	
98	10		
60	60	4	
98	10		
64	60	18	***24**
65	60	1	
4	∞		

***Cycling conditions for samples with low viral load inputs (Ct > 29)**

The resulting NAAT products were cleaned and concentrated using Mag-Bind® TotalPure NGS magnetic beads (Omega Bio-tek Inc, Norcross, USA). A second NAAT was performed to incorporate coded adapters to the amplicons in an indexing step as per the NAAT specifications described in Table 5. Samples that were previously defined as having low viral inputs were cycled 6 times as per the manufacturer's recommendations (Table 5).

Table 5. Indexing NAAT cycling conditions as per SWIFT SNAP manufacturers' recommendations.

Temperature (°C)	Time	Cycles	
37	20min	1	
98	30s	1	
98	10s		
60	30s		
66	60s	9	***6**
4	∞		

***Cycling conditions for samples with low viral load inputs**

3.5 SARS-CoV-2 sequencing

The indexed libraries were cleaned and concentrated with PEG NaCl and the resulting libraries were quantified with the NEBNext library quantification kit as per manufacturers instruction (New England Biolabs, UK). The amplicon size and integrity of the generated library was assessed using a high-sensitivity D1000 TapeStation screen tape (Agilent Technologies) and was determined to be 285 base pairs (bp). Libraries that had a concentration ≥6nM were enzymatically normalized

according to the Swift protocol and all synthesized libraries were pooled. The pool was quantified with the NEBNEXT library quantification kit according to manufacturer's instruction (New England Biolabs, UK) and the D1000 TapeStation screen tape (Agilent Technologies). The quantified library pool was diluted to a final concentration of 100pM and sequenced with a 300 cycle v2 iSeq 100 Reagent Kit on an Illumina iSeq 100 instrument (Illumina, San Diego, CA, USA).

3.6 Genome Assembly and Phylogenetic Analysis

Paired-end fastq reads from the iSeq were *de novo* assembled using Genome Detective 1.126 (https://www.genomedetective.com) (77, 83). Adapters were trimmed and low-quality reads were filtered out using Trimmomatic (84, 85); read quality was visualised using FastQC (85). A report containing the final contig as a FASTA file was generated; the report contained additional information on filtering, assemblage and consensus sequences (85). Generated FASTA files were aligned to the SARS-CoV-2 reference genome (NC_045512.2, ancestral strain) using the multiple sequence alignment software MUSCLE (version 3.8.1551). All SNPs, insertions and deletions were visualised with CodonCodeAligner software (version 10.0.1, https://www.codoncode.com) and confirmed through manual inspection of the sequence BAM (basic alignment map) file. All synonymous mutations were included. Nonsynonymous mutations present in 2 or more departments and or present in more than 5% in a single department were included. The polished consensus sequences were imported into Geneious Prime 2021 (version 11.1.3, https://www.genious.com and NextClade (https://clades.nextstrain.org/) was used to determine the sequence quality, clade calling and the various mutations for each sequence. Genomes that had a genome coverage ≥90% were included in the analysis and any genomes that did not meet the criteria were excluded from the dataset. Sequenced libraries that were within the inclusion criteria but fell within the upper limit of Ct's tested and failed to provide confident lineage results (sequencing coverage <30%) were designated as unclassified. The genomes were uploaded to GISAID (http://gisaid.org) according to the recommended guidelines (Accession numbers: EPI_ISL_14590729-14590731, EPI_ISL_14594649-14594780). SARS-CoV-2 lineages of all sequences were assigned with the PANGOLIN nomenclature tool (https://pangolin.cog-uk.io).

Contemporaneous SARS-CoV-2 sequences from Johannesburg, South Africa (Africa / South Africa / Johannesburg) were incorporated into the phylogenetic tree to contextualize the clustering of HCW sequences with that of the surrounding community. These contemporaneous sequences included the ChAdOx1 nCoV-19 vaccine trial sequences generated in-house as well as those sequences downloaded from GISAID. A total of 25 contemporaneous sequences per month were included in the phylogenetic analysis. Sequences downloaded from GISAID (downloaded 2022-06-21) were included in the phylogenetic analysis according to the following filters: high coverage, complete genomes, and complete metadata. The collection dates of these samples vary from June 2020 to September 2021. All ChAdOx1 nCoV-19 sequences that were generated in-house were used in the generation of the phylogenetic trees across the three wave periods. Phylogenetic newick files were generated using the neighbour-joining method in Molecular Evolutionary Genetics Analysis software (MEGA11) (86). Phylogenetic trees were visualised and annotated using the online Interactive Tree of Life tool (https://itol.embl.de) (87). Notable mutations, SNPs and deletions were determined using Pathogenwatch (https://www.sanger.ac.uk/tool/pathogenwatch/) and nextstrain web-based software (https://clades.nextstrain.org/).

3.7 Identification of Putative Transmission Clusters

Samples were considered isogenic if sequences had ≥1 unique SNP that was not attributable to a lineage defining mutation or SNP (88, 89). Putative transmission clusters were refined from isogenic sequences if these samples originated from HCWs that tested positive for SARS-CoV-2 within 0 to 10 days of each other (90, 91).

Our conclusions on putative transmission clusters and transmission events were based on whether the sequence from the HCW met the SNP and 0 to 10 day criteria and belonged to the same department or related departments. Related departments included General Paediatrics and Neonates with Obstetrics and Gynaecology as well as Internal Medicine with ICU. If genomic similarity was observed in sequences of HCWs from unrelated departments a transmission link was not inferred. Loss or acquisition of a SNP between HCWs from the same or related departments within the 0 to 10 day period inferred a transmission link. If HCWs tested positive between 0 to 2 days of each other and their sequences were isogenic, infection from a same

common source such as an infected patient in the department was inferred. If HCWs tested positive between 3 to 10 days of each other and their sequences were isogenic, HCW-HCW transmission was concluded.

3.8 SARS-CoV-2 cases by department

HCW SARS-CoV-2 cases were stratified according to department during the first wave of infection. Therefore, the detection rate of SARS-CoV-2 was only calculated for the first wave. Detection rate was calculated as a percentage of the number of positive HCW cases in each department per the number of enrolled HCWs in each department in the first wave. Cases were reported as numbers (n) for the second and third infection waves as the surveillance study included HCW enrolment from other hospitals and additional departments within CHBAH. Detection rate of SARS-CoV-2 for the second and third infection waves was not reported, as some participants did not present for routine screening and additional HCWs were enrolled (i.e. second enrolment cohort).

4. Ethical considerations

The HCW surveillance study was approved by the Human Research Ethics Committee (HREC) of the University of the Witwatersrand (reference number 200405), appendix A. All study participants provided written informed consent. An amendment to undertake sequencing of positive samples was approved by HREC and (reference number M211042), appendix B.

5. Results

5.1 Study population

A total of 549 HCWs were enrolled into the surveillance, of whom 197 (35.8%) had at least one NAAT confirmed SARS-CoV-2 infection during the course of the study period (April 2020 to September 2021). The confirmed SARS-CoV-2 cases included 137, 13 and 47 infections during the first, second and third waves, respectively. Only one HCW tested positive in the first and the second infection waves (N=1/549, 0.18%); however, the SARS-CoV-2 specimen from the second wave was not sequenced due to Ct>35. There were no HCWs that tested positive in the third infection wave that were positive for SARS-CoV-2 during the first or second infection waves.

Eighty-three percent (N=164/197) of the positive samples were sequenced; 21 (10.7%) were excluded as the Ct>35 and 12 (6.1%) had insufficient archived volume. Accordingly, 84.7% (N=116/137), 92.3% (N=12/13) and 76.6% (N=36/47) of positive samples were sequenced from the three respective waves. The trajectory of SARS-CoV-2 infections in HCW cohort mirrored that of infections which transpired nationally in South Africa, with peaks in July 2020, January 2021 and June 2021 for the respective waves (Figure 6).

In the ChAdOx1 nCoV19 vaccine trial, 3664 participants were enrolled of whom 562 (15.3%) tested positive for SARS-CoV-2 infection; including 283, 112 and 144 cases during the respective three waves (Figure 6). A total of 79.4% (N=446 /562) of the samples from the ChAdOx1 nCoV19 were sequenced.

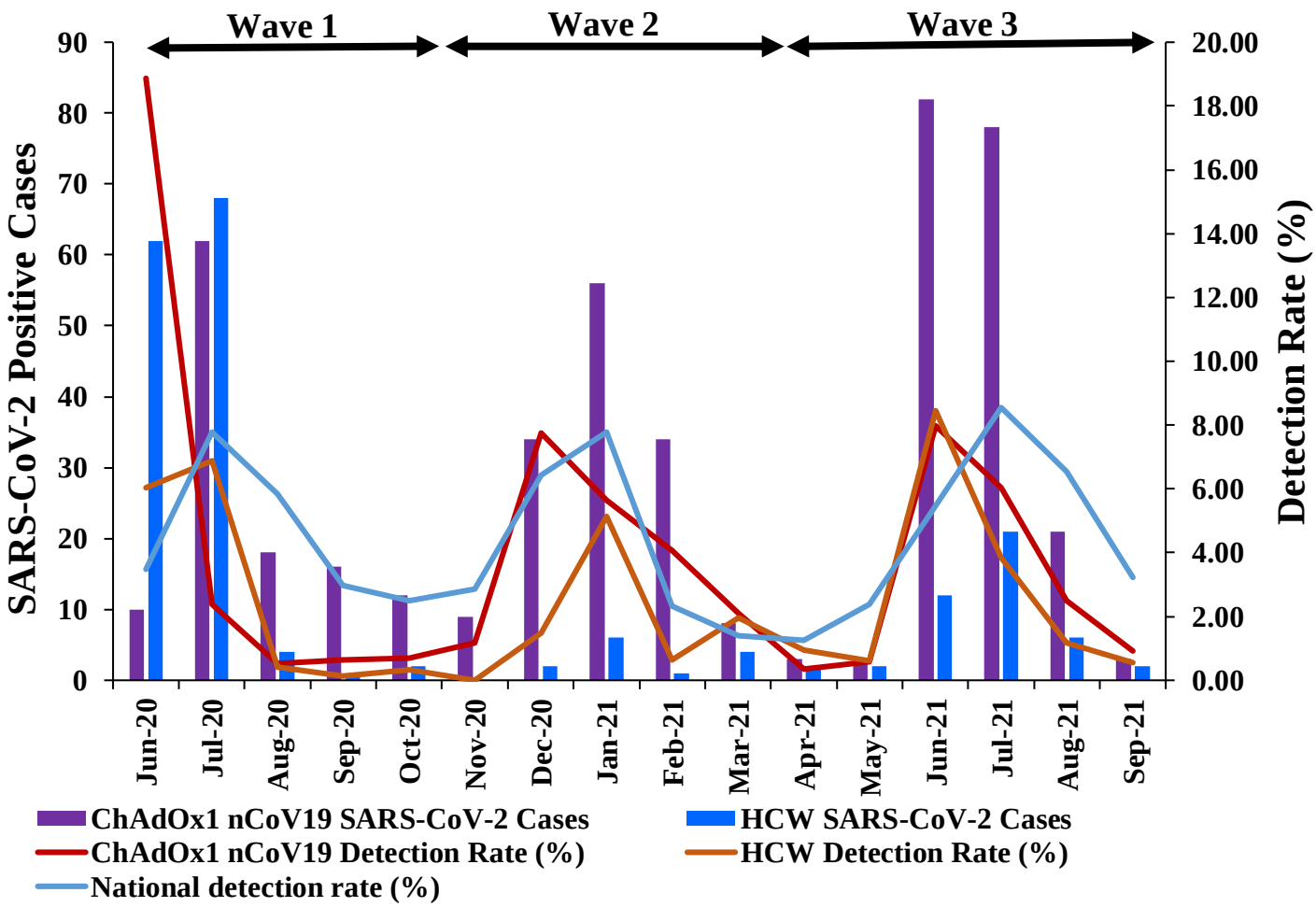


Figure 6. SARS-CoV-2 cases in HCW and ChAdOx1 nCoV-19 vaccine trial samples across three infection waves. Detection rate and number of SARS-CoV-2 cases from June 2020 to September 2021 across the different wave periods from HCWs (Blue) and ChAdOx1 nCoV-19 vaccine trial samples (Purple). Study defined wave periods: Wave 1 (June 2020-October 2020), Wave 2 (December 2020-March 2021) and Wave 3 (April 2021-September 2021). Detection rate for HCWs (Orange), ChadOx1 nCoV-19 (Red) and South Africa (Blue) (92) was calculated as a percentage of the number of positive cases from the total number of samples tested per month.

Similar distributions of SARS-CoV-2 strains were found to be circulating among HCW and the COVID-19 vaccine trial population. Briefly, a total of 8 clades were identified in both the HCW and ChAdOx1 nCoV-19 trial samples, including 20H (Beta), 20J (Delta), 20D C.1.2, 20B, 20A and 20D; with samples from both groups showing a similar distribution of the clades over time (Figure 7).

Among the HCWs, 3 clades were identified in the first wave; the predominate clade being 20B followed by 20D and 20A (Figure 7B). Similarly, in the COVID-19 vaccine trial participants, clade 20B was the most dominant in the first wave. Clade 20H (Beta) dominated in HCW and the vaccine group in the second wave, but was

detected in the vaccine participant samples four months (September 2020) before it was detected among HCW (December 2020). In HCW and the vaccine trial population, the third wave of SARS-CoV-2 infections was the most genotypically diverse with 7 and 4 clades identified, respectively. Clades 20H (Beta), 20J (Delta), 20D C.1.2 and 20B were co-circulating in both groups during the third wave of infection. Regardless, 20J (Delta) was the dominant clade in both groups and accounted for 74.2% (N=23/31) and 87.6% (N=134/153) for the HCW and ChadOx1 nCoV-19 cohort respectively. Three additional clades (20A, 20I (Alpha) and 19A) were identified in the vaccine trial population but not among the HCW during the third infection wave, albeit being infrequent even among the vaccine trial participants (N=1/350, N=1/350, and N=3/350, respectively) (Figure 7B).

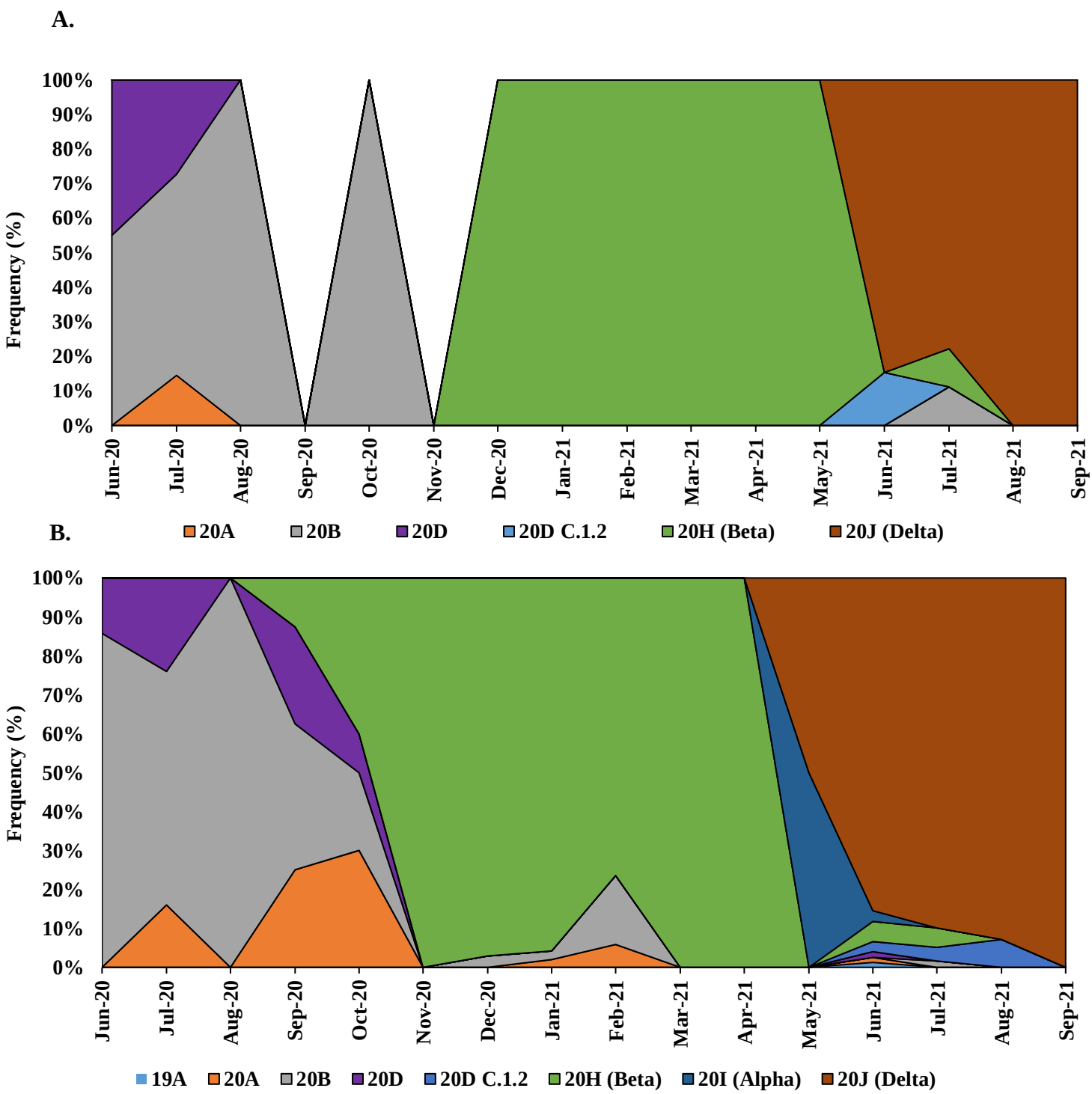
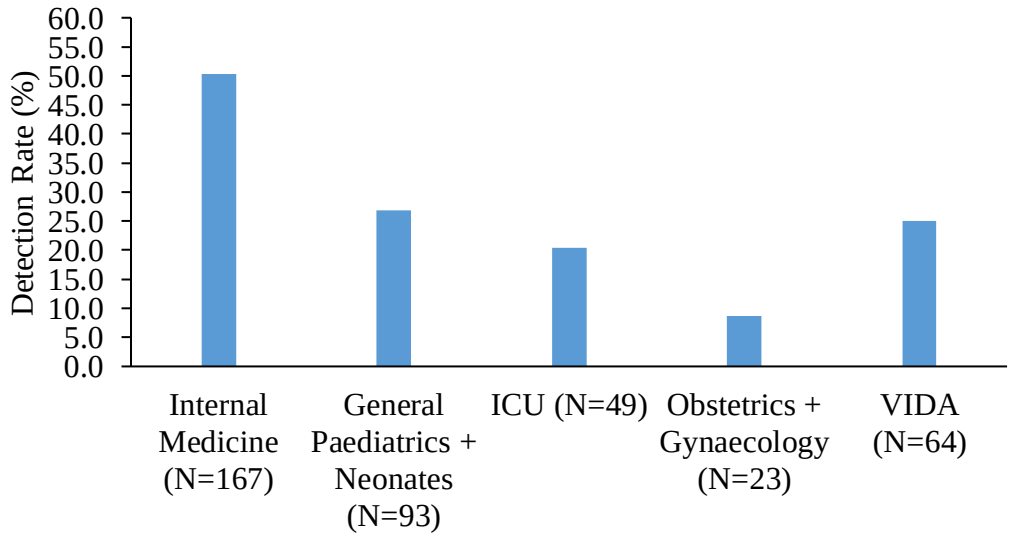


Figure 7. Frequency of monthly SARS-CoV-2 detections. Each panel indicates the frequency of clades commonly identified in HCW (Panel A) and in the same community (Panel B) over the same time period. The colour represents the identified clades including 19A (blue), 20A (orange), 20B (grey), 20D (purple), 20D C.1.2 (light blue), 20H Beta (green), 20I Alpha (dark blue), 21J Delta (brown) and Unclassified (dark grey).

5.2 Molecular Epidemiology of SARS-CoV-2 Cases by Department

The HCWs from the Internal Medicine department had the highest overall positivity for SARS-CoV-2 throughout the study period (N=103/217, 47.5%); followed by the HCWs from the General Paediatrics and Neonates department (N=40/115, 34.8%). The HCWs from the Obstetrics and Gynaecology department accounted for the lowest positivity (N=4/26, 15.4%) (Figure 8A). During the first COVID-19 wave, positivity was 50.3% (N=84/167) among HCWs in Internal Medicine, 26.9% (25/93) in General Paediatrics and Neonates, 25% (16/64) at Wits-Vida, 20.4% (10/49) in ICU HCWs and lowest in HCWs in Obstetrics and Gynaecology (8.7%; n=2/23) (Figure 8A). Detection rate of SARS-CoV-2 for the second and third infection waves was not reported, as study participants were less likely to present for routine screening due to study fatigue. During the second wave of infection SARS-CoV-2 HCW cases were detected in three departments namely Internal Medicine (n=6), Wits-VIDA (n=5) and General Paediatrics and Neonates (n=2) (Figure 8B); no positive cases were detected in the HCWs from the ICU and Obstetrics and Gynaecology department during the second wave. Positive cases detected in the third wave of infection were seen in the HCWs from Internal Medicine (n=13), General Paediatrics and Neonates (n=13), ICU (n=9), Obstetrics and Gynaecology (n=2) and Wits-VIDA (n=7).

A.



B.

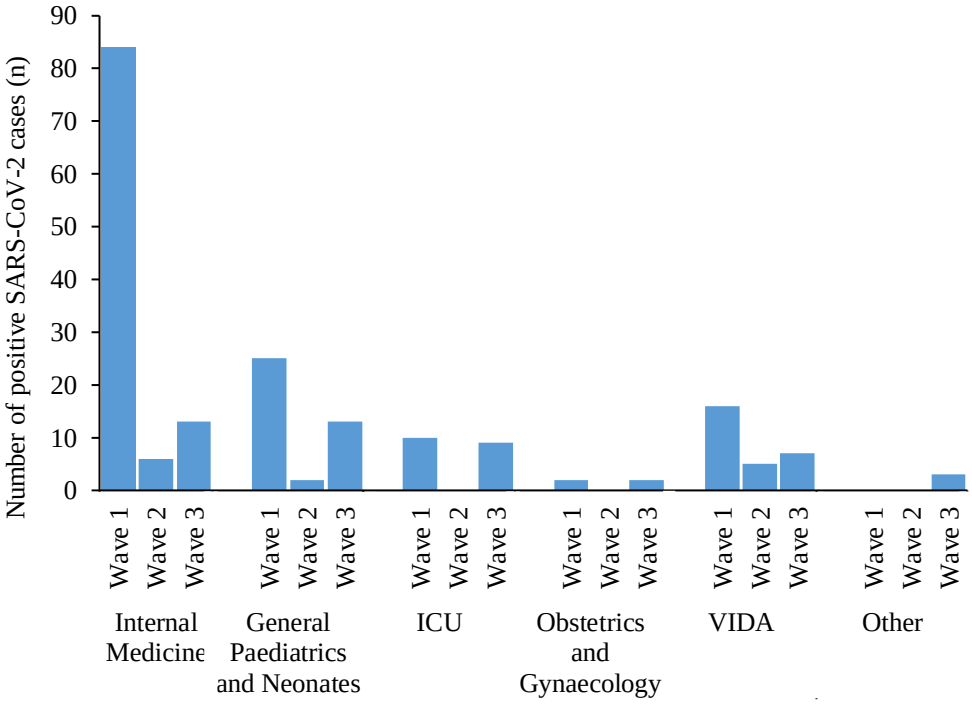


Figure 8.SARS-CoV-2 HCW cases across the five investigated departments during the three infection waves. Panel A represents the detection rate (%) of SARS-CoV-2 cases per department for the first wave of infection. The number positive cases per department from the first, second and third wave of infection is shown in Panel B. Abbreviations: VIDA: Vaccines & Infectious Disease Analytics Unit, ICU: Intensive Care Unit, SARS-CoV-2: Severe Acute Respiratory Symptom Coronavirus 2; N: number of HCWs enrolled in each department in the first wave; n: number of SARS-CoV-2 positive HCWs per department. Other constitutes support services, occupational therapy, physiotherapy, human resources, and admin departments. Study defined wave periods: Wave 1 (June 2020-October 2020), Wave 2 (December 2020-March 2021) and Wave 3 (April 2021-September 2021).

During the first COVID-19 wave, only clade 20B was detected in HCWs from Obstetrics and Gynaecology department (N=2/2, 100.0%) and was also frequently detected at Wits-VIDA (N=9/16, 56.3%) and the Internal Medicine department (N=31/56, 55.4%); while 20A was detected in General Paediatrics and Neonates (N=7/19, 36.8%) and Wits-VIDA (N=1/14, 7.1%). Clade 20D was most commonly detected in the Internal Medicine department (N=25/56, 44.6%), followed by Wits-VIDA (N=4/16, 25.0%) (Figure 9). Sequenced libraries that were within the inclusion criteria but fell within the upper limit of Ct's tested and failed to provide confident lineage results were designated as unclassified.

During the second wave of infection, clade 20H (Beta) was the only clade detected among samples collected from HCWs from any department (Figure 9). Similarly, in the third wave, clade 20J (Delta) was detected in HCWs from all investigated departments and was the dominant circulating clade in the third wave, Obstetrics and Gynaecology (N=1/1, 100.0%), General Paediatric and Neonate department (N=9/10, 90.0%), Internal Medicine (N=6/11, 54.5%), Wits-VIDA (N=3/3, 100.0%) and ICU (N=4/5, 80.0%) (Figure 9). Clade 20D C.1.2 was only detected in HCWs from Internal Medicine (N=1/11, 9.1%) and ICU (N=1/5, 20.0%) (Figure 9A; Figure 9C). Clade 20H (Beta) was detected throughout the third wave in HCWs from two departments namely, Internal Medicine (N=3/11, 27.3%) and General Paediatrics and Neonate (N=1/10, 10.0%) (Figure 9A; Figure 9B). Clade 20B was only detected in HCWs from Internal Medicine (N=1/11, 9.1%).

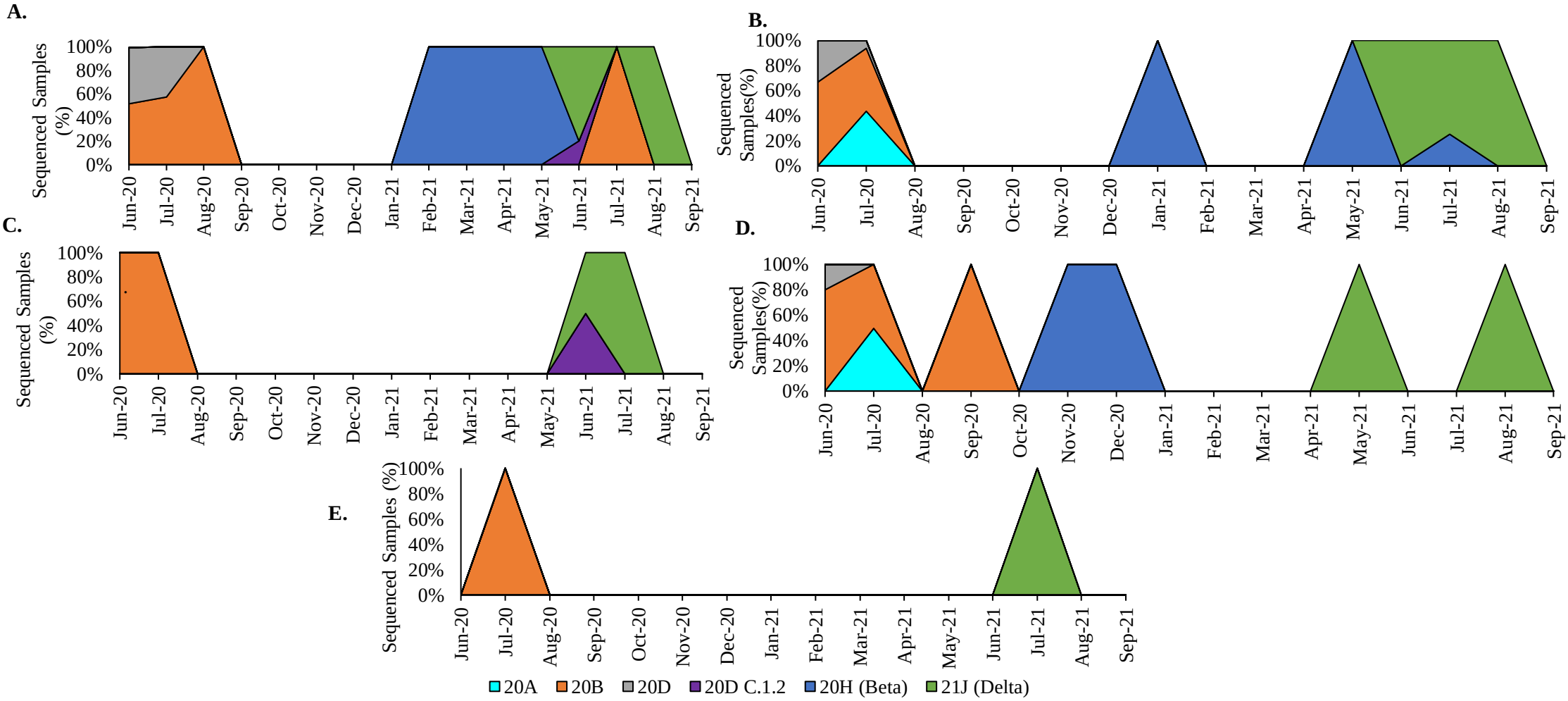


Figure 9. Distribution of clades identified among the HCWs in the different departments across the first three waves of the SARS-CoV-2 pandemic. Internal Medicine (Panel A), General Paediatrics and Neonates (Panel B), Intensive Care Unit (Panel C), Vaccines and Infectious Disease Analytics Unit (Wits-VIDA) (Panel D) and Obstetrics and Gynaecology (Panel E). Wave defined periods are: Wave 1 (June 2020-October 2020), Wave 2 (December 2020-March 2021) and Wave 3 (April 2021-September 2021). The colour represents the identified strains including 20A (light blue), 20B (orange), 20D (grey), 20D C.1.2 (purple), 20H Beta (blue) and 21J Delta (green).

5.3 Phylogenetic Characterisation across Departments

We constructed maximum likelihood phylogenetic trees from the database of HCWs and the in-house COVID-19 vaccine trial SARS-CoV-2 sequences together with contextual GISAID sequences for SARS-CoV-2 sequences isolated in Johannesburg, South Africa during the same time as this study. We classified all infections occurring among HCWs in the same or linked departments that occurred within 3-10 days as probable HCW-HCW transmission; however, we cannot exclude that HCWs could have been infected from the same patient in a ward or contracted the virus elsewhere.

5.3.1 Wave 1

In the maximum likelihood phylogenetic analysis of sequences identified during the first wave of SARS-CoV-2 infections, 100 of 137 sequences from the HCWs were inter-dispersed across the tree which contained 125 additional contemporaneous viruses. The remaining 37 were excluded due to low sequence coverage. We identified four putative transmission clusters which contained HCW sequences from all the different investigated departments, except the Obstetrics and Gynaecology department (Figure 10). Cluster 1 (C1) was comprised of 7 HCWs from the same department, namely General Paediatric and Neonate (Figure 10A); two HCWs (HCW-2075 and HCW-2012) in this cluster tested positive on the same day (D0) and another two within <48 hours of the first cases (HCW-2069 and HCW-2073 respectively). All four of these HCWs had SNPs C355T, C4093T, G213055, C14670T and C21575T (Figure 10B, Figure 10C). The very short infections timelines and shared genomic profile of the sequences suggest that these HCWs likely contracted the virus from the same infected source, possibly a patient in the department. The first plausible HCW-HCW transmission event (HCW-2064) was detected five days after the first case (Figure 10B, infection timeline 1); this sequence is isogenic to the sequences first detected in this cluster thereby indicating virus transmission either from one of previously infected HCW or from the same initial infected source. Two additional HCWs (HCW-2041 and HCW-2011) tested positive 9 days after HCW-2064 tested

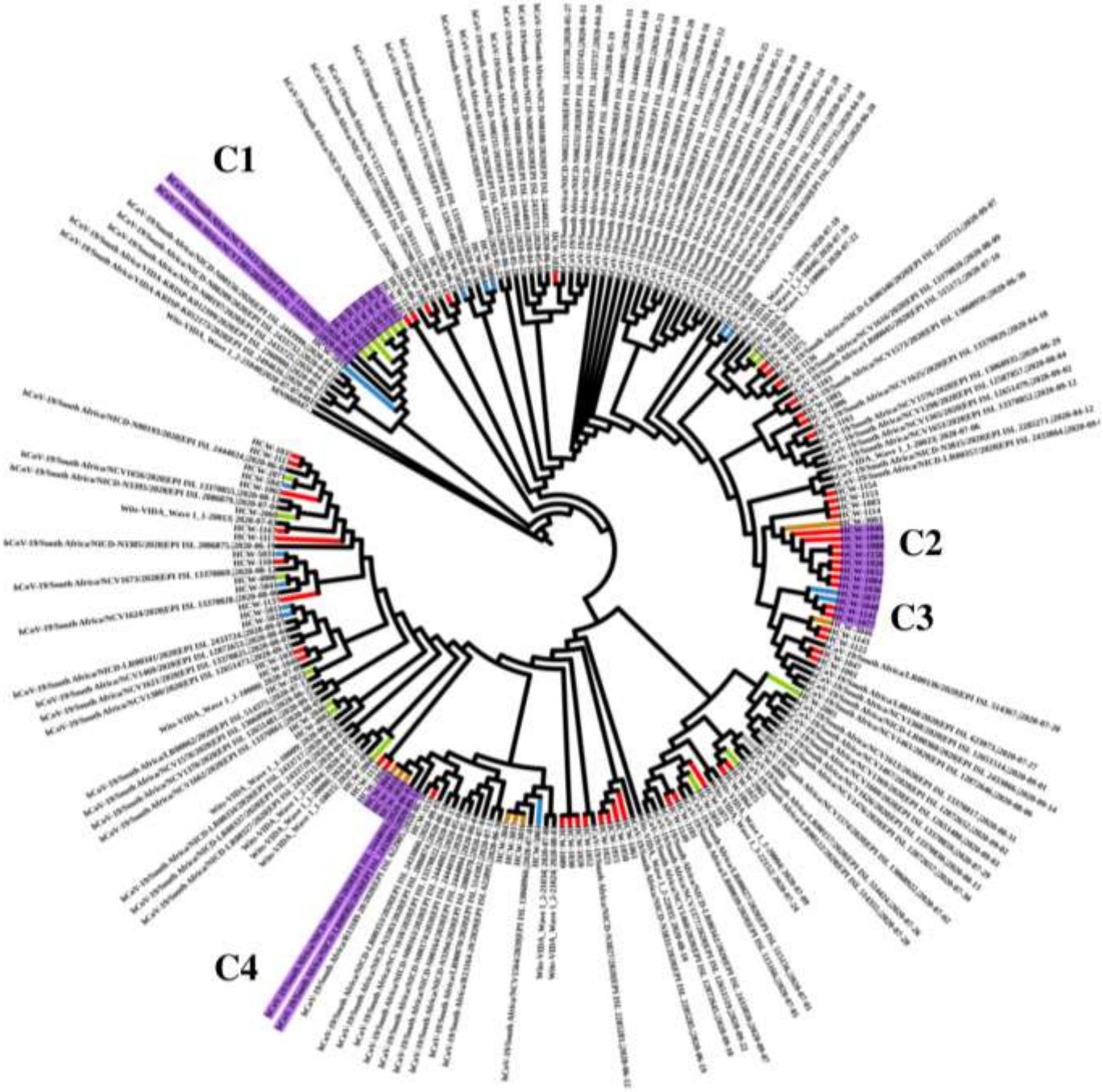
positive (Figure 10B, infection timeline 2) and shared the loss of SNP C21575T suggesting a second HCW-HCW transmission event.

Cluster 2 (C2) comprised of 8 HCWs from the Internal Medicine department (Figure 10A); two HCWs (HCW-1004 and HCW-1046) in this cluster tested positive on the same day (D0) and another two days after the first cases (HCW-1084) (Figure 10B). All three sequences contained SNP G26660T and T23490A, suggesting the possibility that these HCWs were infected by the same patient in the department (Figure 10C). HCW-1004, who tested positive on D0 acquired an additional SNP (C26801T) during the transmission event; the remaining HCW in this cluster (HCW-1036, HCW-1080, HCW-1156, HCW-1032 and HCW-1026) tested positive 9-10 days after this case and shared the same genomic profile including SNP C26801T indicative of HCW-HCW transmission from HCW-1004 to the other HCW in this cluster.

Cluster 3 (C3) comprised of 4 HCWs, two from Wits-VIDA and two from Internal Medicine (Figure 10A) thus two timelines were considered for this cluster as virus transmission between the HCW from Wits-VIDA to Internal Medicine was unlikely. In the first timeline two HCWs from Wits-VIDA tested positive – one on D0 (HCW-5037) and another seven days later (HCW-5048), SNP G25388T was seen in both sequences with HCW-5048 acquiring an additional SNP (A35542T) (Figure 10B and 10C). In the second timeline two HCWs from Internal Medicine tested positive on D0 (HCW-1077) and D9 (HCW-1141), both sequences had SNPs A3542T and G14277T suggesting potential HCW-HCW transmission of the virus.

Cluster 4 (C4) comprised of 3 HCWa from ICU (Figure 10A); briefly all sequences in this cluster were isogenic (Figure 10C). Two HCWs (HCW-3026 and HCW-3043) tested positive on the same day (D0) and the other (HCW-3018) tested positive 8 days later (Figure 10B). The timeframe of each case and the associated incubation period of the virus suggests that HCW-3026 and HCW-3043 were infected by a patient in the department with probable HCW-HCW transmission of the virus to HCW-3018 (D8).

A.



B.

Sample ID	Infection Timeline 1 (Days)	Infection Timeline 2 (Days)
Cluster 1		
HCW-2075	0	
HCW-2012	0	
HCW-2069	1	
HCW-2073	2	
HCW-2064	5	0
HCW-2041	14	9
HCW-2011	14	9
Cluster 2		
HCW-1004	0	
HCW-1046	0	
HCW-1084	2	
HCW-1036	9	
HCW-1080	10	
HCW-1156	10	
HCW-1032	10	
HCW-1026	10	
Cluster 3		
HCW-5037	0	
HCW-5048	7	
HCW-1077	12	0
HCW-1141	21	9
Cluster 4		
HCW-3026	0	
HCW-3043	0	
HCW-3018	8	

C.

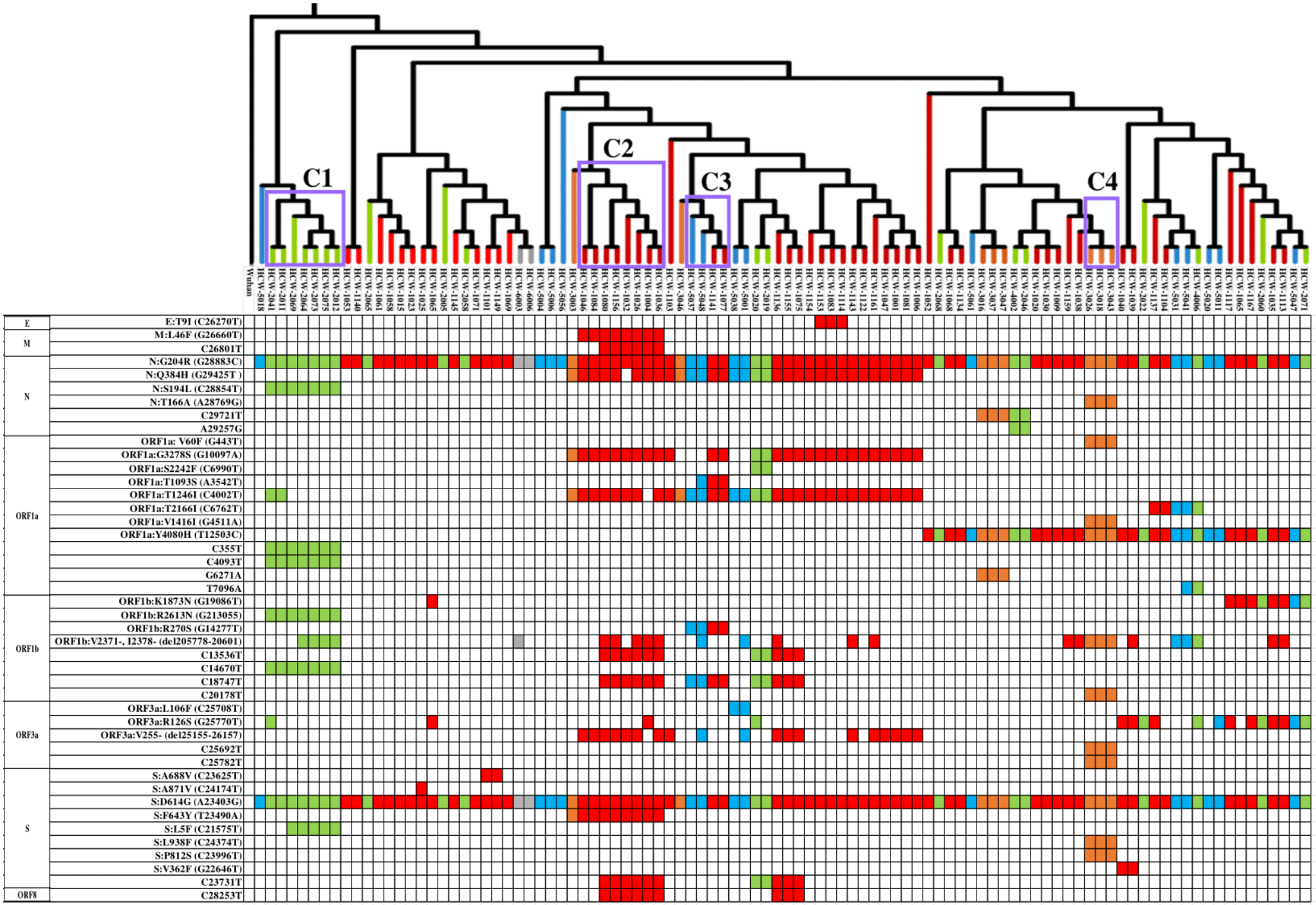
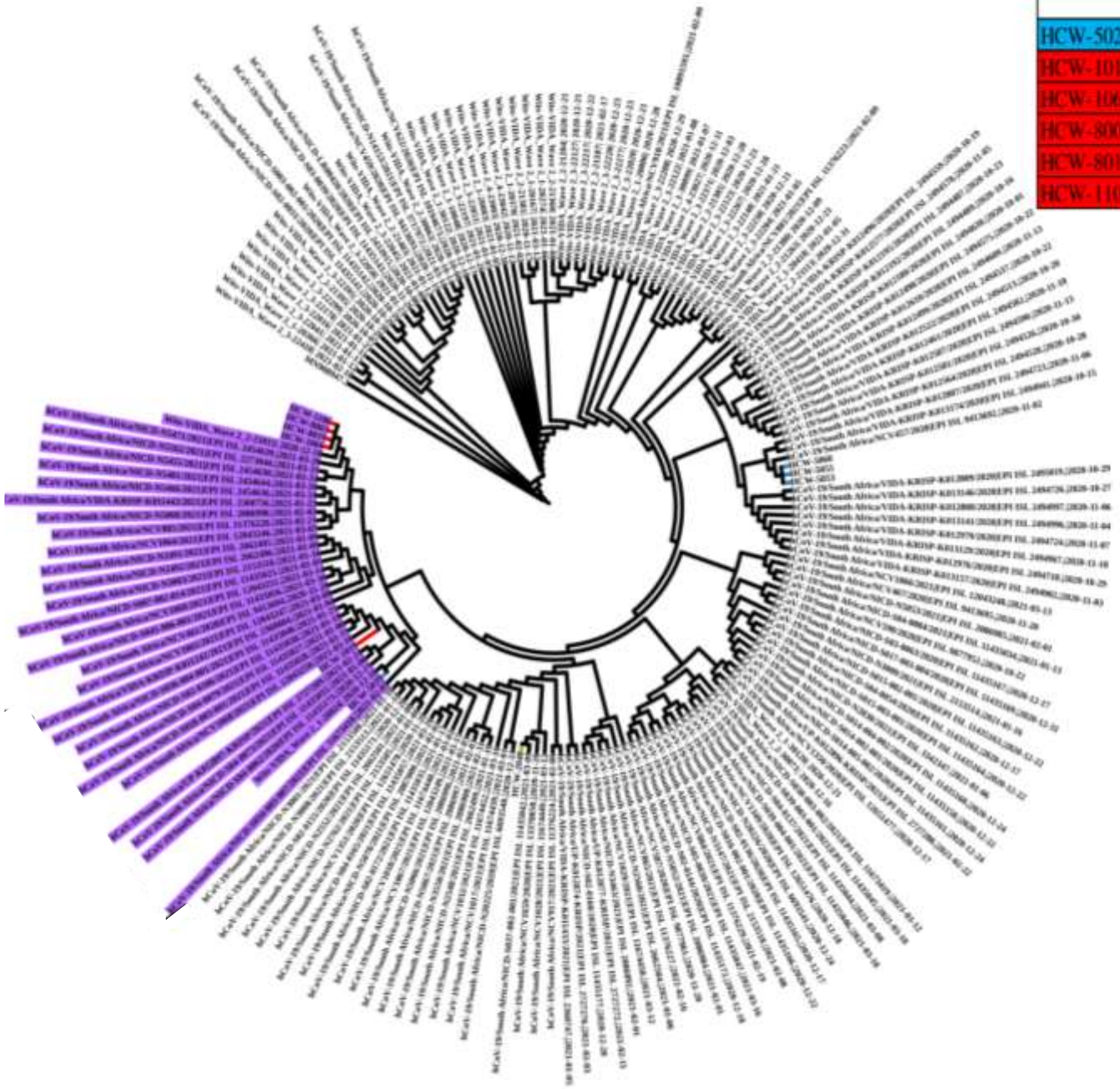

C1
C2
C3
C4
E
M
N
ORF1a
ORF1b
ORF3a
S
ORF8
E:T9I (C26270T)
M:L46F (G26660T)
C26801T
N:G204R (G28883C)
N:Q384H (G29425T)
N:S194L (C28854T)
N:T166A (A28769G)
C29721T
A29257G
ORF1a: V60F (G443T)
ORF1a:G3278S (G10097A)
ORF1a:S2242F (C6990T)
ORF1a:T1093S (A3542T)
ORF1a:T1246I (C4002T)
ORF1a:T2166I (C6762T)
ORF1a:V1416I (G4511A)
ORF1a:Y4080H (T12503C)
C355T
C4093T
G6271A
T7096A
ORF1b:K1873N (G19086T)
ORF1b:R2613N (G213055)
ORF1b:R270S (G14277T)
ORF1b:V2371-, I2378- (del205778-20601)
C13536T
C14670T
C18747T
C20178T
ORF3a:L106F (C25708T)
ORF3a:R126S (G25770T)
ORF3a:V255- (del25155-26157)
C25692T
C25782T
S:A688V (C23625T)
S:A871V (C24174T)
S:D614G (A23403G)
S:F643Y (T23490A)
S:L5F (C21575T)
S:L938F (C24374T)
S:P812S (C23996T)
S:V362F (G22646T)
C23731T
C28253T

Figure 10. Phylogenetic characterisation of community and HCW sequences from the first infection wave. The branches of the tree and mutations are coloured according to department from which the SARS-CoV-2 specimen was identified; Internal Medicine (Red), General Paediatrics and Neonates (Green), Wits-VIDA (Blue), ICU (Orange) and Obstetrics and Gynaecology (Grey). Clusters are indicated in purple (C1-C4). Panel A represents a maximum likelihood phylogenetic tree from the database of HCW, the in-house COVID-19 vaccine trial (Wits-VIDA, Black) and GISAID (EPI_ISL numbered, Black) sequences for SARS-CoV-2 sequences isolated in Johannesburg, South Africa. Two infection timelines (days) for the sequences in each cluster are shown in Panel B. Mutation, SNP and deletion summary for each sequence from the first wave is represented in Panel C. Abbreviations: SNP: Single Nucleotide Polymorphism, ORF: Open-reading frame, M: membrane, N: nucleocapsid, E: envelope, S: spike and del: deletion.

5.3.2 Wave 2

In the maximum likelihood phylogenetic analysis of sequences identified during the second wave of SARS-CoV-2 infections, 10 of 11 sequences from the HCWs were inter-dispersed across the tree which contained 125 additional contemporaneous viruses. One sequence was excluded due to low sequence coverage. We identified one putative transmission cluster, Cluster 5, which contained 6 HCW sequences from VIDA and the Internal medicine department; five HCW (HCW-1019, HCW-1066, HCW-8019, HCW-8003, HCW-1106) were from the Internal Medicine department and the remaining HCW (HCW-5029) was from Wits-VIDA (Figure 11A). All six sequences had the mutation S:L18F (C21614T) and SNP A2692T (Figure 11C). Two HCW sequences within C5 (HCW-5029 and HCW-1019) shared an additional SNP specifically C26645T; however, the second HCW (HCW-1019) tested positive 52 days after the first case (HCW-5029) (Figure 11B, Figure 11C). The remaining four HCWs in this cluster were from the Internal Medicine department (Figure 11A), all of these sequences shared SNPs C29580T and C26010T and mutation ORF3a:W131L (G25784T) (Figure 11C). HCW-8003 tested positive six days (D6) after the first case (HCW-1066) thus indicating likely HCW-HCW (Figure 11B, infection timeline 2); in infection timeline 3 a second HCW-HCW transmission event between HCW-8003 and HCW-8019 was observed. Although the remaining HCW (HCW-1106) is isogenic to the other three HCW sequences (HCW-1066, HCW-8003 and HCW-8019) the timelines associated with detection of the virus is not indicative of a linked transmission event and thus the infection of HCW-1106 was extrapolated to community importation.

A.



B.

Sample ID	Infection Timeline 1 (Days)	Infection Timeline 2 (Days)	Infection Timeline 3 (Days)
Cluster 5			
HCW-5029	0		
HCW-1019	52		
HCW-1066	56	0	
HCW-8003	62	6	0
HCW-8019	69	13	7
HCW-1106	83	27	21

C.

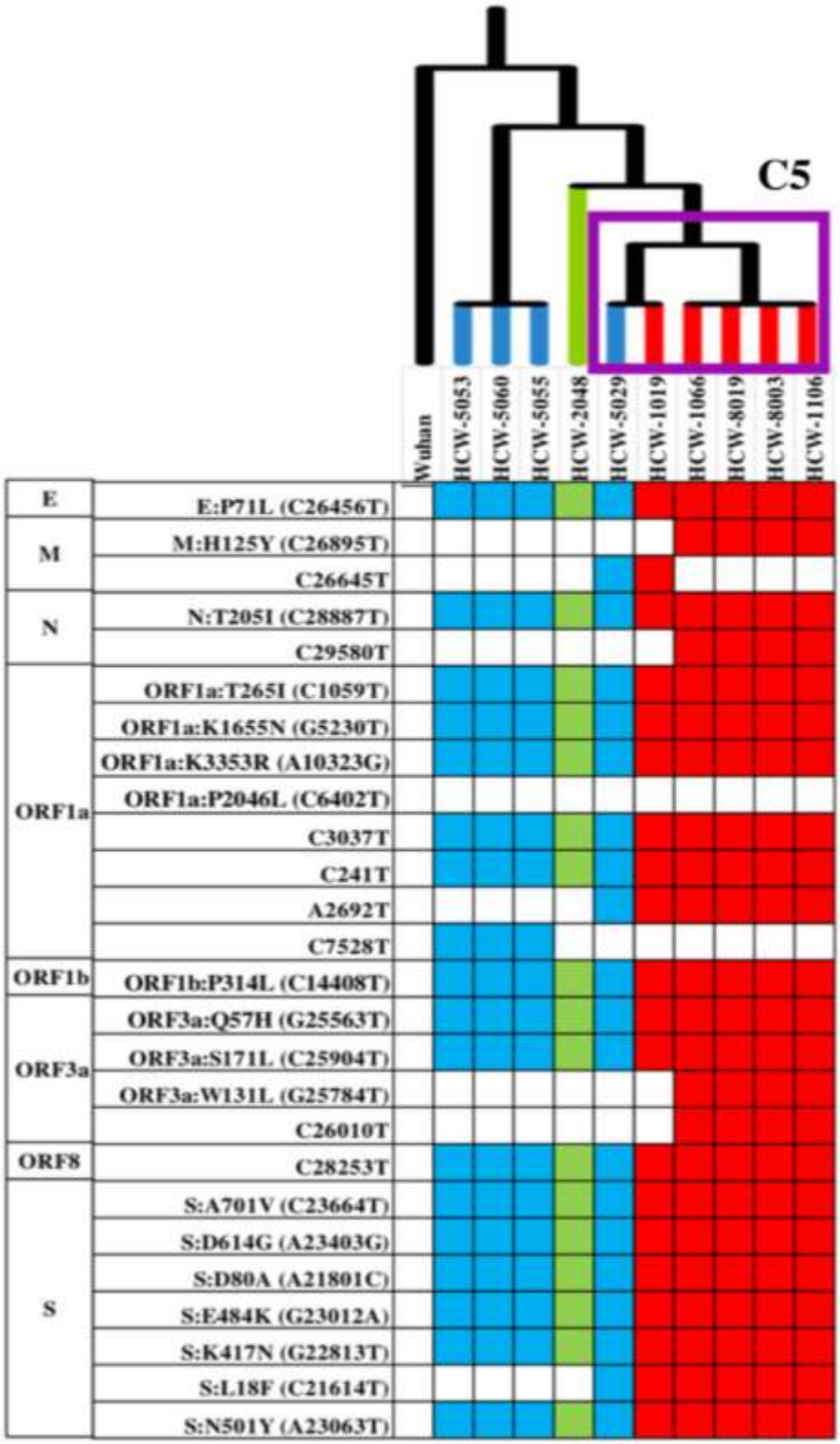


Figure 11.Phylogenetic characterisation of community and HCW sequences from the second infection wave. The branches of the tree and mutations are coloured according to department from which the SARS-CoV-2 specimen was identified; Internal Medicine (Red), General Paediatrics and Neonates (Green) and Wits-VIDA (Blue). Clusters are indicated in purple (C5). Panel A represents a maximum likelihood phylogenetic tree from the database of HCW and the in-house COVID-19 vaccine trial SARS-CoV-2 first wave sequences (Wits-VIDA, Black) together with GISAID (EPI_ISL numbered, Black) sequences for SARS-CoV-2 sequences isolated in Johannesburg, South Africa and three infection timelines (days) for the sequences in each cluster is shown in Panel B. Mutation, SNP and deletion summary for each sequence from the first wave is represented in Panel C. Abbreviations: SNP: Single Nucleotide Polymorphism, ORF: Open-reading frame, M: membrane, N:nucleocapsid, E:envelope, S:spike and del: deletion.

5.3.3 Wave 3

In the maximum likelihood phylogenetic analysis of sequences identified during the third wave of SARS-CoV-2 infections, 33 of 47 sequences from the HCWs were inter-dispersed across the tree which contained 125 additional contemporaneous viruses. The remaining 13 sequences were excluded due to low sequence coverage. We identified four putative transmission clusters which contained HCW sequences from all the investigated departments except Wits-VIDA (Figure 12). Cluster 6 (C6) had 2 HCW from the Internal Medicine department (Figure 12B). These sequences were isogenic and shared SNPs T2526C and C9448T (Figure 12C); HCW-1050 tested positive three days after the first case (HCW-1049) thereby indicative of a linked HCW-HCW transmission episode.

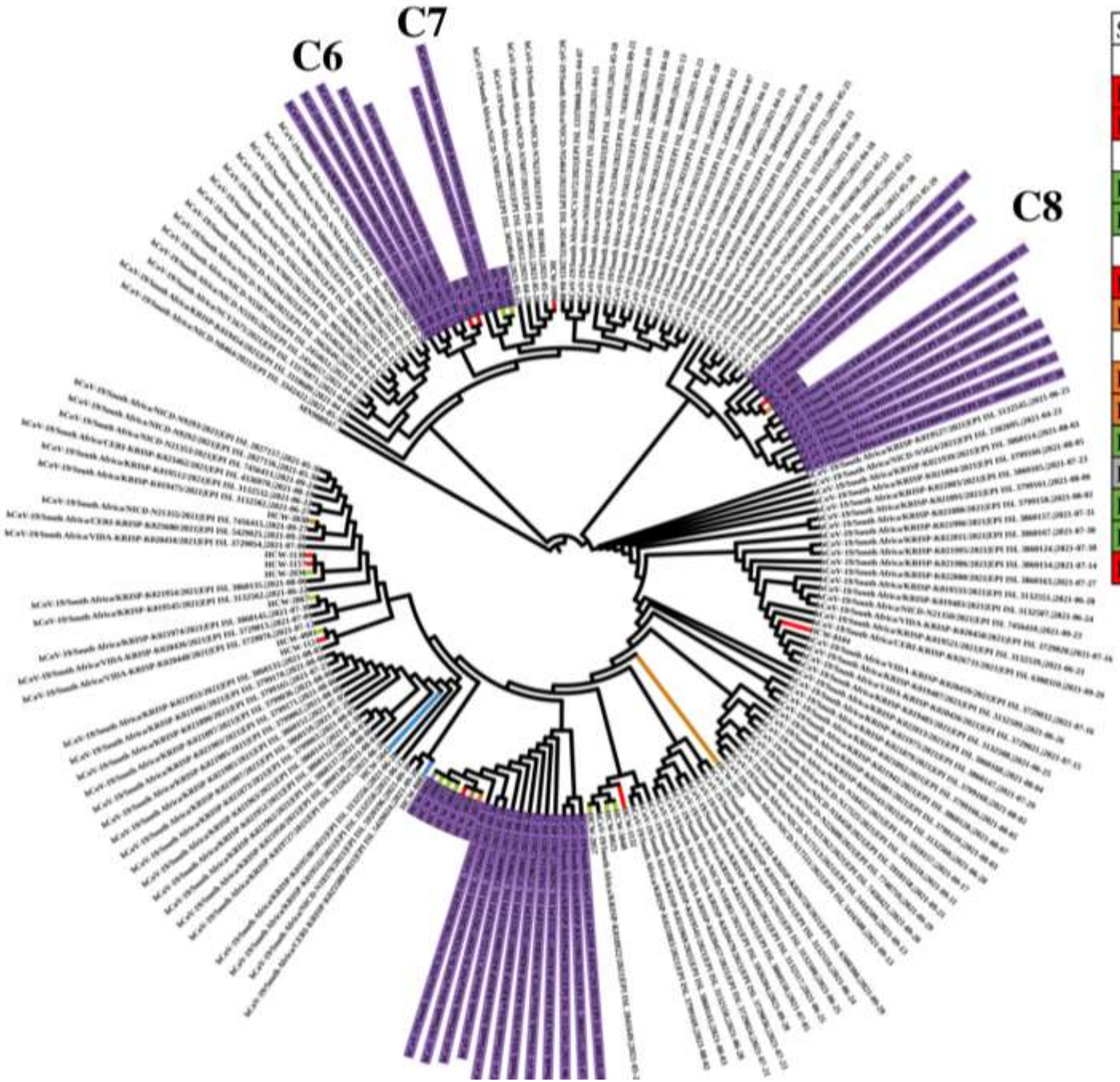
Cluster 7 (C7) was comprised of 2 HCWs from the General Paediatrics and Neonate department (Figure 12A). HCW-8106 tested positive 2 days after the first case in this cluster (HCW-8050) (Figure 12B). Both sequences are isogenic sharing SNPs G3452A, C9166T, G21974T and A22298G suggesting that both HCWs contracted the virus from the same infected patient in the department (Figure 12C).

Cluster 8 (C8) was comprised of 2 HCW, one from ICU (HCW-8083) and one from Internal Medicine (HCW-8074) (Figure 12A). These sequences were isogenic, sharing many of the same SNPs (Figure 12C). The second HCW (HCW-8083) tested positive 5 days after the first case (HCW-8074) (Figure 12B) indicative of a linked HCW-HCW transmission episode.

Cluster 9 (C9) was comprised of 7 HCWs from multiple departments including ICU, Internal Medicine, General Paediatrics and Neonates and Obstetrics and Gynaecology (Figure 12A). Shared SNPs among the sequences were observed including C4438T and C7528T (Figure 12C). Although a shared genomic profile was observed for these samples it is unlikely due to HCW-HCW transmission as the timeline between infections were often >10 days and the HCWs were from departments where cross over was unlikely. These sequences cluster with other contextual sequences and thus we propose community importation of virus into each department rather than linked transmission. A linked transmission event was considered in timeline two, between a

HCW from Obstetrics and Gynaecology (HCW-6022) and a HCW from the Paediatrics and Neonate (HCW-2066) department. HCW-2066 tested positive 3 days (D3) after the first case (HCW-6022).

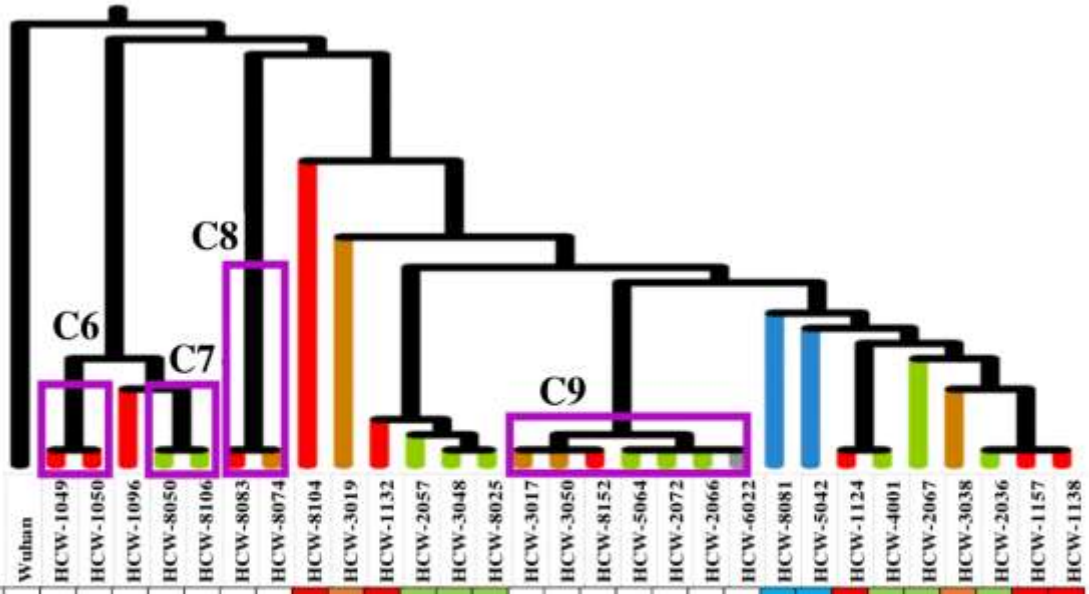
A.



B.

Sample ID	Infection Timeline 1 (Days)	Infection Timeline 2 (Days)
Cluster 6		
HCW-1049	0	
HCW-1050	3	
Cluster 7		
HCW-8050	0	
HCW-8106	2	
Cluster 8		
HCW-8074	0	
HCW-8083	5	
Cluster 9		
HCW-3050	0	
HCW-3017	1	
HCW-2072	14	
HCW-6022	28	0
HCW-2066	31	3
HCW-5064	45	
HCW-8152	40	

C.



M	M:I82T (T26767C)
	C26570T
	C26895T
	C26645T
	C26010T
	C27131T
N	N:G204R (G28883C)
	N:G215C (G28916T)
	N:P13L (C28311T)
	N:R203K (G28881A + G28882A)
	C29580T
ORF1a	ORF1a:A1306S (G4181T)
	ORF1a:E102K (G569A)
	ORF1a:G49D (G411A)
	ORF1a:H2092Y (C6539T)
	ORF1a:P2046L (C6402T)
	ORF1a:S3675-/F3677- (del11288-11296)
	ORF1a:T1637I (C5175T)
	ORF1a:T3646A (A11201G)
	A11332G
	A7699G
	C11674T
	C11902T
	C4438T
	C7528T
	G1264T
	G210T
	C1912T
	C2485T
	C1059T
	G174T
	T5626C
	G3452A
	C9166T
	C9448T
ORF1b	ORF1b:G6625 (G15451A)
	ORF1b:M115I (G13812T)
	ORF1b:P1000L (C16466T)
	ORF1b:P314L (C14408T)
	C13536T
	C13548T
	C18747T
	C19017T
ORF3a	ORF3a:Q57H (G25563T)
	ORF3a:S171L (C25904T)
	ORF3a:S26L (C25469T)
	ORF3a:T60A (A28461G)
	G25874T
	C25413T
ORF6	A27399G
ORF7a	ORF7a:V82A (T27638C)
ORF7b	ORF7b:T40I (C27874T)
ORF8	C28253T
ORF9b	ORF9b:M78T (T28516C)
	ORF9b:P10S (C28311T)
S	S:T19R (C21618G)
	S:L18F (C21614T)
	S:D80A (A21801C)
	S:C136F (G21969T)
	S:G142D (G21987A)
	S:E156-/R158G (del22029-22034)
	S:R190S (G22132T)
	S:D2115G (A22206G)
	S:L242-/A243- (del22289-22294)
	S:K147N (G22813T)
	S:L452R (T22917G)
	S:V449H (T22907C)
	S:T478K (C22995A)
	S:E484K (G23012A)
	S:N501Y (A23063T)
	S:D614G (A23403G)
	S:H655Y (C23525T)
	S:A701V (C23664T)
	S:T716I (C23709T)
	S:D950N (G24410A)
	C22210T
	G21974T
	A22298G

Figure 12. Phylogenetic characterisation of community and HCW sequences from the third infection wave. The branches of the tree and mutations are coloured according to department from which the SARS-CoV-2 specimen was identified; namely Internal Medicine (Red), General Paediatrics and Neonates (Green), Wits-VIDA (Blue), ICU (Orange) and Obstetrics and Gynaecology (Grey). Clusters are indicated in purple (C6-C9). Panel A represents a maximum likelihood phylogenetic tree from the database of HCW and the in-house COVID-19 vaccine trial SARS-CoV-2 first wave sequences (Wits-VIDA, Black) together with GISAID (EPI_ISL numbered, Black) sequences for SARS-CoV-2 sequences isolated in Johannesburg, South Africa and two infection timeline (days) for the sequences in each cluster is shown in Panel B. Mutation, SNP and deletion summary for each sequence from the first wave is represented in Panel C. Abbreviations: SNP: Single Nucleotide Polymorphism, ORF: Open-reading frame, M: membrane, N:nucleocapsid, E:envelope, S:spike and del: deletion.

6. Discussion

This study reports on the genotyping of 164 SARS-CoV-2 samples collected from healthcare workers working at Chris Hani Baragwanath Academic Hospital - a large tertiary hospital which services a population of approximately 1.5 million people in Soweto, South Africa (93). During the study, transmission peaks and clade proportion of SARS-CoV-2 among the HCWs paralleled those observed in the general population in South Africa. Our phylogenetic analysis revealed that the bulk of the HCWs samples (N=122/164, 74.4%) clustered with other contemporaneous strains which were reflective of the community with evidence of only nine putative transmission clusters identified (four in the first and third infection wave and one in the second wave). These nine putative transmission clusters accounted for 25.0% (N=41/164) of the infected HCWs. Similar to other studies (90, 94), we classified all infections occurring among HCWs in the same or linked departments that occurred within 3-10 days as probable HCW-HCW transmission. Thus, of the 41 infection episodes included in the putative transmission clusters, only 11 were likely HCW-HCW transmissions with the remaining 30 infection episodes showing evidence of transmission from a common source such as a patient, as the infection episodes were all detected within 24-48 hours of each other. Thus 93.3% (N=153/164) of the HCWs infected with SARS-CoV-2 during this study were likely infected from patients or to a lesser extent from the community. This is likely because HCWs have high exposure to multiple infected individuals with very severe disease. Further, limited personal protective equipment (PPE) and training in the use of the PPE was available during the initial stages of the pandemic (95, 96).

Our findings were contradictory to other studies which found that in 15-79% of infected HCWs the source of infection was from contact with an infected colleague (97-101). For example, there was a cluster of intra-hospital HCW-HCW transmission reported in Singapore (102) where they found that SARS-CoV-2 transmission from an asymptomatic HCW resulted in the exposure of an additional 68 HCWs (103). Similar findings were reported in Madrid, Spain in which five HCWs contracted SARS-CoV-2 after having lunch with an asymptomatic HCW and subsequently went on to infect more than 25 patients in the hospital (104).

Our contradictory findings could be linked to our study design whereby the HCWs included in the study were tested weekly for SARS-CoV-2 infections, regardless of symptom, further, if a HCW developed any sign of illness or had a close contact with another positive HCW or family member then they were encouraged to come for SARS-CoV-2 testing immediately. This likely lead to the rapid identification of asymptomatic HCWs or HCWs in the very early stages of the infection cycle thus likely preventing further HCW-HCW infections. A similarly designed study at the NHS Hospital Trust in England investigated within-hospital transmission and found that by integrating routine SARS-CoV-2 testing and genomic surveillance, HCW-HCW transmission was reduced by 18.7% from the first to the second epidemic wave (105).

In our study, the second and third infection waves saw a very limited number of infected HCWs detected likely due to acquired natural immunity from previous viral infections (106) and vaccine-induced immunity as a result of vaccine rollout in February 2021 (107-109). Regardless, it must be noted that the peak of the second wave of SARS-CoV-2 infections occurred over the December period when many of the HCWs were on holiday and thus were less likely to present for screening at our facility. Further, during the second and third waves of COVID-19 infections study fatigue was noted with many HCW not wanting to do routine screening for infection and many HCWs dropping out of the study.

Regardless, four putative transmission clusters were still identified in the third wave including a large transmission cluster (cluster 9) comprised of seven HCWs from four different departments (ICU, General Paediatrics and Neonates, Internal Medicine and Obstetrics and Gynaecology) with infections spanning over more than 6 weeks. However, even though the specimens included in this transmission cluster were isogenic, typically during an outbreak, many individuals are infected with viruses that have a high degree of genetic similarity based on the dominant circulating variant present (89, 105), as seen in this study. Thus, other factors also need to be looked at to determine whether HCW-HCW transmission was plausible including whether the infections occurred within a timeframe corresponding to the distribution and incubation period of SARS-CoV-2 and linked hospital location of suspected related individuals (105). Linked departments were defined by the likelihood of patients being transferred from one department to other: for example, Obstetrics and

Gynaecology to General Paediatrics and Neonates or Internal Medicine to ICU, and for our study we defined a 0–10 day transmission period. However some studies reveal that this timeframe is highly SARS-CoV-2 variant specific with some strains extending up to 14 days (110) while others have a reduced COVID-19 transmission period (111). Thus, the predefined transmission period of the study could have allowed for the incorrect inclusion or exclusion of potential transmission clusters. Consequently, putative cluster 9 was likely not a true transmission cluster as either the HCW departments were not linked or the transmission periods between HCWs infections from linked departments were too long. Similarly, cluster 5 had HCW sequences that were isogenically similar, but unlikely a true transmission cluster as the HCW in this cluster were from departments with little to no cross over namely Wits-VIDA and Internal Medicine and the HCWs tested positive in excess of a month of one another. HCWs from Wits-VIDA were responsible for the processing and testing of samples from other departments thus, the likelihood of a Wits-VIDA HCW contracting the virus from a sample from any department is high. Conversely, transmission from Wits-VIDA to HCWs from other departments is unlikely as Wits-VIDA staff predominantly were not in direct contact with the HCWs from the other departments. Thus, we proposed that the shared genomic profiles of HCWs in this cluster were due to community importation of the virus rather than a linked infection episode which is further supported by these HCW sequences clustering with many contextual sequences. However, when a second and third infection timeline was considered in cluster 5, evidence for HCW-HCW transmission was observed in three HCWs that all belonged to the Internal Medicine department and had infection episodes occurring within 3-10 days of each other.

Similarly, clusters 1, 2, 4 and 8 were likely true transmission clusters because the HCWs in these clusters belonged to the same department (clusters 1, 2 and 4) or linked departments (cluster 8) and the HCW tested positive within 3-10 days of each other. In cluster 8 we concluded a linked infection episode that involved two HCWs that contracted the virus from the same infected patient that was transferred from Internal Medicine to ICU. Hospital patients are typically transferred to ICU wards if their medical status deteriorates, and they require intensive treatment and close monitoring. Thus, the likelihood of a patient infecting a HCW from both the ICU and Internal Medicine department was considered plausible.

Our study had several limitations in that not all hospital staff were enrolled in the study; therefore, additional clusters or HCW-HCW transmission events could have been missed. Additionally, viral sequences from positive patient cases and contact tracing between HCW and patients in the same department was not investigated therefore, while we postulated that the HCWs were infected by their patients, we cannot prove it. Further, while we classified all infections occurring among HCWs in the same of linked departments that occurred within 3-10 days as probable HCW-HCW transmission, we cannot exclude that HCWs could have been infected from the same patient in a ward especially as immunocompromised individuals have been shown to have prolonged infection cycles (112, 113). In future work, the evidence behind transmission clusters in healthcare facilities can be strengthened with the inclusion of genomic surveillance of positive patients. Additionally, interviews outlining social behaviours and shared transportation could be included for the purpose of inferring links between HCWs that work in different departments.

In conclusion, similarly to the rest of the world, a large proportion of the HCWs were infected with SARS-CoV-2 during the first wave of infection; however, unlike many other studies (98, 100, 102), these infections were more likely a result of infection from their patients rather than HCW-HCW transmissions. Thus, without the intense routine surveillance of the HCWs included in the study and the rapid identification of asymptomatic infections, the SARS-CoV-2 infection burden in the HCWs at CHBAH could have been substantially worse. Our study highlights the importance of continual genomic surveillance and molecular characterisation of communicable diseases as well as routine surveillance to implement effective infection control practices and reduce HCW morbidity and mortality in future outbreaks.

7. References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
2. Peiris JS, Yuen KY, Osterhaus AD, Stohr K. The severe acute respiratory syndrome. N Engl J Med. 2003;349(25):2431-41.
3. Madhi SA, Kwatra G, Myers JE, Jassat W, Dhar N, Mukendi CK, et al. Population Immunity and Covid-19 Severity with Omicron Variant in South Africa. N Engl J Med. 2022;386(14):1314-26.
4. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020;382(18):1708-20.
5. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3):181-92.
6. Jassat W, Cohen C, Tempia S, Masha M, Goldstein S, Kufa T, et al. Risk factors for COVID-19-related in-hospital mortality in a high HIV and tuberculosis prevalence setting in South Africa: a cohort study. Lancet HIV. 2021;8(9):e554-e67.
7. Monroe AK, Polyak CS, Castel AD, Esber AL, Byrne ME, Maswai J, et al. Clinical similarities and differences between two large HIV cohorts in the United States and Africa. PLoS One. 2022;17(4):e0262204.
8. van de Water BJ, Fulcher I, Cilliers S, Meyer N, Wilson M, Young C, et al. Association of HIV infection and antiretroviral therapy with the occurrence of an unfavorable TB treatment outcome in a rural district hospital in Eastern Cape, South Africa: A retrospective cohort study. PLoS One. 2022;17(4):e0266082.
9. Chakaya J, Khan M, Ntoumi F, Aklillu E, Fatima R, Mwaba P, et al. Global Tuberculosis Report 2020 - Reflections on the Global TB burden, treatment and prevention efforts. Int J Infect Dis. 2021;113 Suppl 1:S7-S12.
10. Rossi GA, Sacco O, Mancino E, Cristiani L, Midulla F. Differences and similarities between SARS-CoV and SARS-CoV-2: spike receptor-binding domain recognition and host cell infection with support of cellular serine proteases. Infection. 2020;48(5):665-9.
11. Woo PC, Huang Y, Lau SK, Yuen KY. Coronavirus genomics and bioinformatics analysis. Viruses. 2010;2(8):1804-20.
12. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. Nat Rev Microbiol. 2009;7(6):439-50.
13. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507-13.
14. Brussow H. The Novel Coronavirus - A Snapshot of Current Knowledge. Microb Biotechnol. 2020;13(3):607-12.
15. Belouzard S, Chu VC, Whittaker GR. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc Natl Acad Sci U S A. 2009;106(14):5871-6.
16. Heald-Sargent T, Gallagher T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. Viruses. 2012;4(4):557-80.
17. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. Int J Biol Sci. 2020;16(10):1686-97.
18. Liu J, Xie W, Wang Y, Xiong Y, Chen S, Han J, et al. A comparative overview of COVID-19, MERS and SARS: Review article. Int J Surg. 2020;81:1-8.
19. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020;5(4):562-9.
20. Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, Dijkman R, et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature. 2013;495(7440):251-4.

21. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497-506.
22. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. bioRxiv. 2020.
23. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020;182(4):812-27 e19.
24. A B. What Is "R-naught"? Gauging Contagious Infections: Healthline; 2020 [cited 2023 10/05]. Available from: https://www.healthline.com/health/r-naught-reproduction-number.
25. Aronson JK BJ, Mahtani KR. "When will it be over?": An introduction to viral reproductionnumbers, R0 and Re. Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences,University of Oxford.

26. Yadav AK, Kumar S, Singh G, Kansara NK. Demystifying R Naught: Understanding What Does it Hide? Indian J Community Med. 2021;46(1):7-14.
27. Elsaid M, Nasef MA, Huy NT. R(0) of COVID-19 and its impact on vaccination coverage: compared with previous outbreaks. Hum Vaccin Immunother. 2021;17(11):3850-4.
28. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395(10224):565-74.
29. Rella SA, Kulikova YA, Dermitzakis ET, Kondrashov FA. Rates of SARS-CoV-2 transmission and vaccination impact the fate of vaccine-resistant strains. Sci Rep. 2021;11(1):15729.
30. When do people infected with SARS-CoV-2 infect others? [Scientific brief]. WHO; 2020 [Available from: https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.
31. Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV-2. Science. 2020;368(6498):1422-4.
32. Contreras S, Dehning J, Loidolt M, Zierenberg J, Spitzner FP, Urrea-Quintero JH, et al. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. Nat Commun. 2021;12(1):378.
33. Park AK, Kim IH, Kim J, Kim JM, Kim HM, Lee CY, et al. Genomic Surveillance of SARS-CoV-2: Distribution of Clades in the Republic of Korea in 2020. Osong Public Health Res Perspect. 2021;12(1):37-43.
34. Lauring AS, Hodcroft EB. Genetic Variants of SARS-CoV-2-What Do They Mean? JAMA. 2021;325(6):529-31.
35. Wang S, Xu X, Wei C, Li S, Zhao J, Zheng Y, et al. Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. J Med Virol. 2022;94(1):310-7.
36. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403-7.
37. consortiumcontact@cogconsortium.uk C-GU. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe. 2020;1(3):e99-e100.
38. Organization WH. SARS-CoV-2 variants of concern and variants of

interest 2021 [cited 2022 2 February]. Available from: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/.
39. Prevention CfDCa. SARS-CoV-2 variant classifications and

definitions 2021 [cited 2022 2 February]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html
40. Tao K, Tzou PL, Nouhin J, Gupta RK, de Oliveira T, Kosakovsky Pond SL, et al. The biological and clinical significance of emerging SARS-CoV-2 variants. Nat Rev Genet. 2021;22(12):757-73.

41. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science. 2021;372(6538).
42. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature. 2021;593(7858):266-9.
43. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature. 2021;592(7854):438-43.
44. Imai M, Halfmann PJ, Yamayoshi S, Iwatsuki-Horimoto K, Chiba S, Watanabe T, et al. Characterization of a new SARS-CoV-2 variant that emerged in Brazil. Proc Natl Acad Sci U S A. 2021;118(27).
45. Khateeb J, Li Y, Zhang H. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. Crit Care. 2021;25(1):244.
46. Winger A, Caspari T. The Spike of Concern-The Novel Variants of SARS-CoV-2. Viruses. 2021;13(6).
47. Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira I, et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. Nature. 2021;599(7883):114-9.
48. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. J Med Virol. 2021.
49. Zuckerman NS, Fleishon S, Bucris E, Bar-Ilan D, Linial M, Bar-Or I, et al. A Unique SARS-CoV-2 Spike Protein P681H Variant Detected in Israel. Vaccines (Basel). 2021;9(6).
50. He X, Hong W, Pan X, Lu G, Wei X. SARS-CoV-2 Omicron variant: Characteristics and prevention. MedComm. 2021.
51. Mohapatra RK, Kandi V, Verma S, Dhama K. Challenges of the Omicron (B.1.1.529) Variant and Its Lineages: A Global Perspective. Chembiochem. 2022:e202200059.
52. SARS-CoV-2 Omicron variant: Characteristics and prevention. Wiley Online Library.
53. SARS-CoV-2 Variant Classifications and Definitions: Centers for Disease Control and Prevention; 2022 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html#anchor_1632154493691.
54. Phan T, Boes S, McCullough M, Gribschaw J, Marsh JW, Harrison LH, et al. First detection of SARS-CoV-2 Omicron BA.4 variant in Western Pennsylvania, United States. J Med Virol. 2022.
55. OMICRON LINEAGES BA.4 AND BA.5 FAQ: National Institute of Communicable Diseases; 2022 [Available from: https://www.nicd.ac.za/omicron-lineages-ba-4-and-ba-5-faq/.
56. Poudel S, Ishak A, Perez-Fernandez J, Garcia E, Leon-Figueroa DA, Romani L, et al. Highly mutated SARS-CoV-2 Omicron variant sparks significant concern among global experts - What is known so far? Travel Med Infect Dis. 2022;45:102234.
57. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat Med. 2021;27(3):440-6.
58. 2021 [Available from: https://nextstrain.org/ncov/africa?dmax=2021-03-03&f_country=South%20Africa.
59. Zuckerman NS, Pando R, Bucris E, Drori Y, Lustig Y, Erster O, et al. Comprehensive Analyses of SARS-CoV-2 Transmission in a Public Health Virology Laboratory. Viruses. 2020;12(8).
60. Parczewski M, Ciechanowicz A. Molecular epidemiology of SARS-CoV-2: a review of current data on genetic variability of the virus. Pol Arch Intern Med. 2020;131(1):63-9.
61. Fernandez A. Structural Impact of Mutation D614G in SARS-CoV-2 Spike Protein: Enhanced Infectivity and Therapeutic Opportunity. ACS Med Chem Lett. 2020;11(9):1667-70.
62. Ladhani SN, Chow JY, Janarthanan R, Fok J, Crawley-Boevey E, Vusirikala A, et al. Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020. EClinicalMedicine. 2020;26:100533.
63. Wise J. Covid-19: The E484K mutation and the risks it poses. BMJ. 2021;372:n359.

64. Collier DA, De Marco A, Ferreira IATM, Meng B, Datir R, Walls AC, et al. SARS-CoV-2 B.1.1.7 escape from mRNA vaccine-elicited neutralizing antibodies. medRxiv. 2021:2021.01.19.21249840.
65. Collier DA, De Marco A, Ferreira I, Meng B, Datir RP, Walls AC, et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. Nature. 2021;593(7857):136-41.
66. Daily New Cases in South Africa: Worldometer; 2022 [Available from: https://www.worldometers.info/coronavirus/country/south-africa/.
67. Scheepers C, Everatt J, Amoako DG, Tegally H, Wibmer CK, Mnguni A, et al. Emergence and phenotypic characterization of the global SARS-CoV-2 C.1.2 lineage. Nat Commun. 2022;13(1):1976.
68. Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, et al. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat Commun. 2021;12(1):4196.
69. Olmos C, Campana G, Monreal V, Pidal P, Sanchez N, Airola C, et al. SARS-CoV-2 infection in asymptomatic healthcare workers at a clinic in Chile. PLoS One. 2021;16(1):e0245913.
70. Xiao J, Fang M, Chen Q, He B. SARS, MERS and COVID-19 among healthcare workers: A narrative review. J Infect Public Health. 2020;13(6):843-8.
71. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. N Engl J Med. 2021;384(19):1866-8.
72. Chowell G, Abdirizak F, Lee S, Lee J, Jung E, Nishiura H, et al. Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study. BMC Med. 2015;13:210.
73. Evans S, Agnew E, Vynnycky E, Robotham J. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. medRxiv. 2020:2020.05.12.20095562.
74. Evans S, Agnew E, Vynnycky E, Stimson J, Bhattacharya A, Rooney C, et al. The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. Philos Trans R Soc Lond B Biol Sci. 2021;376(1829):20200268.
75. Grad YH, Lipsitch M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. Genome Biol. 2014;15(11):538.
76. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. Lancet Infect Dis. 2020;20(11):1263-72.
77. Nunes MC, Baillie VL, Kwatra G, Bhikha S, Verwey C, Menezes C, et al. SARS-CoV-2 infection among healthcare workers in South Africa: a longitudinal cohort study. Clin Infect Dis. 2021.
78. Madhi SA, Baillie V, Cutland CL, Voysey M, Koen AL, Fairlie L, et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. N Engl J Med. 2021;384(20):1885-98.
79. Madhi SA, Kwatra G, Richardson SI, Koen AL, Baillie V, Cutland CL, et al. Durability of ChAdOx1 nCoV-19 (AZD1222) vaccine and hybrid humoral immunity against variants including omicron BA.1 and BA.4 6 months after vaccination (COV005): a post-hoc analysis of a randomised, phase 1b-2a trial. Lancet Infect Dis. 2022.
80. Nunes MC, Madhi SA. COVID-19 vaccines in pregnancy. Trends Mol Med. 2022;28(8):662-80.
81. Nunes MC, Baillie VL, Kwatra G, Bhikha S, Verwey C, Menezes C, et al. SARS-CoV-2 infection among healthcare workers in South Africa: a longitudinal cohort study. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2021.
82. Lee JS, Goldstein JM, Moon JL, Herzegh O, Bagarozzi DA, Jr., Oberste MS, et al. Analysis of the initial lot of the CDC 2019-Novel Coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel. PLoS One. 2021;16(12):e0260487.

83. Cleemput S, Dumon W, Fonseca V, Abdool Karim W, Giovanetti M, Alcantara LC, et al. Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes. Bioinformatics. 2020;36(11):3552-5.
84. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
85. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, et al. Genome Detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics. 2019;35(5):871-3.
86. O'Toole A, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7(2):veab064.
87. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-W6.
88. San JE, Ngcapu S, Kanzi AM, Tegally H, Fonseca V, Giandhari J, et al. Transmission dynamics of SARS-CoV-2 within-host diversity in two major hospital outbreaks in South Africa. Virus Evol. 2021;7(1):veab041.
89. Turcinovic J, Schaeffer B, Taylor BP, Bouton TC, Odom-Mabey AR, Weber SE, et al. Understanding Early Pandemic Severe Acute Respiratory Syndrome Coronavirus 2 Transmission in a Medical Center by Incorporating Public Sequencing Databases to Mitigate Bias. J Infect Dis. 2022;226(10):1704-11.
90. Johansson MA, Quandelacy TM, Kada S, Prasad PV, Steele M, Brooks JT, et al. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. JAMA Netw Open. 2021;4(1):e2035057.
91. Walsh KA, Spillane S, Comber L, Cardwell K, Harrington P, Connell J, et al. The duration of infectiousness of individuals infected with SARS-CoV-2. J Infect. 2020;81(6):847-56.
92. Edouard Mathieu HR, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Joe Hasell, Bobbie Macdonald, Saloni Dattani, Diana Beltekian, Esteban Ortiz-Ospina, Max Roser Coronavirus Pandemic (COVID-19): OurWorldInData.org; 2020 [updated 10/11/2022. 2020:[Available from: https://ourworldindata.org/coronavirus.
93. OVERVIEW OF CHRIS HANI BARAGWANATH ACADEMIC HOSPITAL: Parliament of the Republic of South Africa; 2022 [cited 2022 27 March]. Available from: https://www.parliament.gov.za/storage/app/media/Pages/2022/3-march/22-03-2022_National_Council_of_Provinces_Provincial_Week/Gauteng/Chris_Hani_Baragwanath_Academic_Hospital_Site_Profile.pdf.
94. Jung J, Kim JY, Park H, Park S, Lim JS, Lim SY, et al. Transmission and Infectious SARS-CoV-2 Shedding Kinetics in Vaccinated and Unvaccinated Individuals. JAMA Netw Open. 2022;5(5):e2213606.
95. Meinus C, Singer R, Nandi B, Jagot O, Becker-Ziaja B, Karo B, et al. SARS-CoV-2 prevalence and immunity: a hospital-based study from Malawi. Int J Infect Dis. 2022;116:157-65.
96. Semaan A, Banke-Thomas A, Amongin D, Babah O, Dioubate N, Kikula A, et al. 'We are not going to shut down, because we cannot postpone pregnancy': a mixed-methods study of the provision of maternal healthcare in six referral maternity wards in four sub-Saharan African countries during the COVID-19 pandemic. BMJ Global Health. 2022;7(2):e008063.
97. Al Maskari Z, Al Blushi A, Khamis F, Al Tai A, Al Salmi I, Al Harthi H, et al. Characteristics of healthcare workers infected with COVID-19: A cross-sectional observational study. Int J Infect Dis. 2021;102:32-6.
98. Gordon CL, Trubiano JA, Holmes NE, Chua KYL, Feldman J, Young G, et al. Staff to staff transmission as a driver of healthcare worker infections with COVID-19. Infect Dis Health. 2021;26(4):276-83.
99. Maltezou HC, Dedoukou X, Tseroni M, Tsonou P, Raftopoulos V, Papadima K, et al. SARS-CoV-2 Infection in Healthcare Personnel With High-risk Occupational Exposure:

Evaluation of 7-Day Exclusion From Work Policy. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2020;71(12):3182-7.
100. Mdzinwa N, Voigt M, Janse van Rensburg C, Paruk F. SARS-CoV-2 infection prevalence in healthcare workers and administrative and support staff: The first-wave experience at three academic hospitals in the Tshwane district of Gauteng Province, South Africa. S Afr Med J. 2021;111(11):1092-7.
101. Piapan L, De Michieli P, Ronchese F, Rui F, Mauro M, Peresson M, et al. COVID-19 outbreak in healthcare workers in hospitals in Trieste, North-east Italy. J Hosp Infect. 2020;106(3):626-8.
102. Wee LE, Sim JXY, Conceicao EP, Aung MK, Tan JY, Venkatachalam I. Containment of COVID-19 among ancillary healthcare workers: an integral component of infection control. J Hosp Infect. 2020;106(2):392-6.
103. Wee LE, Fan EMP, Heng R, Ang SY, Chiang JL, Tan TT, et al. Construction of a container isolation ward: A rapidly scalable modular approach to expand isolation capacity during the coronavirus disease 2019 (COVID-19) pandemic. Infect Control Hosp Epidemiol. 2021;42(9):1162-4.
104. Suarez-Garcia I, Martinez de Aramayona Lopez MJ, Saez Vicente A, Lobo Abascal P. SARS-CoV-2 infection among healthcare workers in a hospital in Madrid, Spain. J Hosp Infect. 2020;106(2):357-63.
105. Lindsey BB, Villabona-Arenas CJ, Campbell F, Keeley AJ, Parker MD, Shah DR, et al. Characterising within-hospitalSARS-CoV-2 transmission events using epidemiological and viral genomic data across two pandemic waves. Nat Commun. 2022;13(1):671.
106. Baraniuk C. How long does covid-19 immunity last? BMJ. 2021;373:n1605.
107. Bekker LG, Garrett N, Goga A, Fairall L, Reddy T, Yende-Zuma N, et al. Effectiveness of the Ad26.COV2.S vaccine in health-care workers in South Africa (the Sisonke study): results from a single-arm, open-label, phase 3B, implementation study. Lancet. 2022;399(10330):1141-53.
108. Nordstrom P, Ballin M, Nordstrom A. Risk of SARS-CoV-2 reinfection and COVID-19 hospitalisation in individuals with natural and hybrid immunity: a retrospective, total population cohort study in Sweden. Lancet Infect Dis. 2022;22(6):781-90.
109. Team C-F. Past SARS-CoV-2 infection protection against re-infection: a systematic review and meta-analysis. Lancet. 2023;401(10379):833-42.
110. Huang PY, Wu TS, Cheng CW, Chen CJ, Huang CG, Tsao KC, et al. A hospital cluster of COVID-19 associated with a SARS-CoV-2 superspreading event. J Microbiol Immunol Infect. 2022;55(3):436-44.
111. Wu Y, Kang L, Guo Z, Liu J, Liu M, Liang W. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. JAMA Netw Open. 2022;5(8):e2228008.
112. Alshukairi AN, Tolah AM, Dada A, Al-Tawfiq JA, Almagharbi RS, Saeedi MF, et al. Test-based de-isolation in COVID-19 immunocompromised patients: Cycle threshold value versus SARS-CoV-2 viral culture. Int J Infect Dis. 2021;108:112-5.
113. Thornton CS, Huntley K, Berenger BM, Bristow M, Evans DH, Fonseca K, et al. Prolonged SARS-CoV-2 infection following rituximab treatment: clinical course and response to therapeutic interventions correlated with quantitative viral cultures and cycle threshold values. Antimicrob Resist Infect Control. 2022;11(1):28.

Appendix A. HREC Certificate (200405)

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Human Research Ethics Committee: (Medical)
FWA Registered No IRB 00001223

SECRETARIAT: Suite 189, Private Bag x2600, Houghton 2041, South Africa Tel: +27-11-274 9200 Fax: +27-11-274 9281

17 April 2020

EMAILED & COURIERED

Prof MC Nunes

Respiratory & Meningeal Pathogens Research Un
11th Floor, West Wing, Nurses Residence
Chris Hani Baragwanath Hospital
P O Box 90753, Bertsham
2013

Email: nunesm@rmpru.co.za

Dear Prof Nunes,

PROTOCOL: COVID-HCW STUDY - SURVEILLANCE AMONG HEALTHCARE WORKERS FOR SARS-CORONAVIRUS-2 INFECTION (COVID-HCW STUDY)

ETHICS REFERENCE NO: 200405

RE : FINAL ETHICS APPROVAL

This is to certify that the above-mentioned trial has been approved by the University of the Witwatersrand, Human Research Ethics Committee (HREC), and the Protocol/Expert Reviewer. Date of Meeting where trial was reviewed: [redacted]

The University of the Witwatersrand, Human Research Ethics Committee Approval Granted for the above mentioned study is valid for five years. Where required by Sponsor to have approval on a more frequent basis it remains the responsibility of the Sponsor and Investigator to apply for continuing review and approval, or for the duration of the Trial.

1. It is the responsibility of the Sponsor and Principal Investigator to ensure, where required, that relevant approvals are in place and compliance with the following is adhered to before a trial may begin:

- If trial is being conducted in Provincial Health facilities: Approval from the Hospital CEO / Clinic Manager / District Research Committee (whichever is applicable) be obtained.
- The study is submitted onto The National Health Research Database (NHRD).
- The relevant approvals are uploaded onto the NHRD system: Ethics Approval, SAHPRA Approval, Hospital CEO / Clinic Manager / District Research Committee Approval.

* A copy of the SAHPRA Approval and/or SAHPRA Notification letter must be submitted to the Ethics Secretariat Office for record purposes (IF SAHPRA APPROVAL / NOTIFICATION IS APPLICABLE).

* The study is conducted according to the protocol submitted to the University of the Witwatersrand, Human Research Ethics Committee. Any amendments to the protocol must first be submitted to the Human Research Ethics Committee for approval.

* During the study, the University of the Witwatersrand, Human Research Ethics Committee is informed immediately of :

- Any Unexpected Serious Adverse Events or Unexpected Adverse Drug Reactions, which, in the Investigator and/or the Sponsor's opinion are suspected to be related to the study drug. (Refer to POL-IEC–001 and SOP-IEC–005, Item 3.4).
- Any data received during the trial which, may cast doubt on the validity of the continuation of the study.

Appendix B. HREC Certificate (M211042)

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

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

TO: Ms J du Plessis; Profs SA Madhi and M Nunes
School of Pathology
Wits VIDA - Vaccines and Infectious Diseases
Analytics Research Unit
Medical School
University

E-mail: Jeanine.duPlessis@wits-vida.org

CC: Supervisor: Dr V-L Baillie
<Vicky.Baillie@wits-vida.org>
and <HREC-Medical Research Office@wits.ac.za>

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

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

DATE: 2021/10/20

REF: R14/49

PROTOCOL NO: **M211042** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)

PROJECT TITLE: *Molecular epidemiologyof SARS-CoV-2 in healthcare workers at Chris Hani Baragwanath Academic Hospital*

Please find attached the Clearance 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 Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R49 Ms J du Plessis; Profs SA Madhi and M Nunes

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M211042

NAME: (Principal Investigator)	Ms J du Plessis; Profs SA Madhi and M Nunes
DEPARTMENT:	School of Pathology Wits VIDA - Vaccines and Infectious Diseases Analytics Research Unit Medical School University
PROJECT TITLE:	*Molecular epidemiologyof SARS-CoV-2 in healthcare workers at Chris Hani Baragwanath Academic Hospital*
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under WHC 200405
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Dr V-L Baillie
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2021/10/20

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in 2021/10/01 and therefore reports and re-certification will be due in the month of **2021/10/01** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator — Date: 2021/11/01

Appendix C. Turnitin report

Wits Vaccines & Infectious Diseases Analytics
Research Unit (formerly RMPRU)
11th Floor, Central West Wing
Nurses Residence
Chris Hani Baragwanath Academic Hospital
Tel +27 11 983 4283

WITS VIDA

UNIVERSITY OF THE WITWATERSRAND
VACCINES AND INFECTIOUS DISEASES ANALYTICS RESEARCH

Research Office, Faculty of Health Sciences
University of the Witwatersrand
06 February 2023

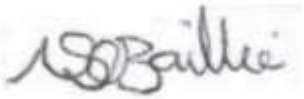
To whom it may concern

Re: Acknowledgement of the Turnitin report and permission for Jeanine du Plessis to submit her Dissertation for examination

This letter serves to inform you that as the Supervisor of Jeanine du Plessis (807629), I have reviewedthe Turnitin report and acknowledge that the similarity index is higher than what is required by the Faculty. However, the phrases that have been highlighted by Turnitin are common scientific terms that are used within the field. The student did not commit plagiarism therefore I give her my full support to submit her Dissertation for examination.

I trust that you will not hesitate to accept her Dissertation for examination.

Yours sincerely



Dr Vicky-Lynne Baillie, Ph.D
Senior Scientist

vicky.baillie@wits-vida.org
Mobile: +27845680299

MSc dissertation_807629_Jeanine du Plessis-1.docx

ORIGINALITY REPORT

20%	14%	17%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	covid19dataportal.es Internet Source	1%
2	www.frontiersin.org Internet Source	1%
3	bcuhb.nhs.wales Internet Source	1%
4	egonw.github.io Internet Source	1%
5	www.medrxiv.org Internet Source	1%
6	www.mdpi.com Internet Source	1%
7	www.spglobal.com Internet Source	<1%
8	Firdose Lambey Nakwa, Reenu Thomas, Alison van Kwawegen, Nandi Ntuli et al. "An outbreak of infection due to severe acute respiratory corona virus-2 in a neonatal unit	<1%